



Acoustic and articulatory dynamics
in second language speech
production: Japanese speakers'
production of English liquids

Takayuki Nagamine, BA, MA

Department of Linguistics and English Language

Lancaster University

A thesis submitted for the degree of

Doctor of Philosophy

November, 2024

**Acoustic and articulatory dynamics in second language speech
production: Japanese speakers' production of English liquids**

Takayuki Nagamine, BA, MA.

Department of Linguistics and English Language, Lancaster University

A thesis submitted for the degree of *Doctor of Philosophy*. November, 2024.

Abstract

This PhD thesis considers how second language (L2) learners use dynamic, time-varying phonetic cues to produce consonants in L2 speech, with the case of first-language (L1) Japanese speakers' production of L2 English liquids as a testing ground. Previous research has shown that L1 Japanese speakers have substantial difficulty in producing L2 English liquids because of the L1 influence. The articulatory mechanism that causes the difficulty, however, remains unclear due to the lack of articulatory data and consideration of dynamic properties, which could be predicted to be an area of difficulty given existing articulatory descriptions of Japanese and English liquids. Acoustic and articulatory data were collected from a total of 55 participants, including 41 L1 Japanese speakers and 14 L1 North American English speakers. Midsagittal tongue movement was recorded using ultrasound tongue imaging while speakers read aloud Japanese and English liquid consonants appearing word-initially, word-medially and word-finally.

The data were analysed both acoustically and articulatorily, resulting in five empirical studies included in the thesis. Acoustic and articulatory analyses of word-initial liquid-vowel sequences in L1 and L2 English suggest a greater variability in tongue dorsum height across different vowel contexts than L1 English speakers, which is not readily observable in static analysis. Studies investigating L1 Japanese speakers' production of L2 English allophony commonly show a mismatch between acoustics and articulation, suggesting that L1 Japanese speakers may utilise different sets of articulatory strategies to achieve target-like acoustic output. Taken all

these findings together, this PhD thesis proposes that dynamics involved in speech production could be a source of L2 speech production difficulty. It demonstrates that the combination of the dynamic and articulatory aspects involved in L2 speech production could further advance our understanding of the specific obstacles L2 learners encounter in the course of L2 speech learning.

Acknowledgements

I am a mediocre person without any special skill set, but one thing that I am very proud of is the people surrounding me. I have had the privilege of getting to know so many wonderful people during my PhD journey, all of whom have helped me stand where I am now. It is likely that I am missing out some people, but I will try my best to acknowledge all those who have been important to me.

I could never be more grateful for having the best supervisors in the world, Claire Nance and Sam Kirkham, who have guided, inspired, and supported me throughout my time at Lancaster University. Thank you so much for always providing crystal-clear feedback on my often redundant writing, for listening to me overthink tiny little issues, for celebrating key milestones together, and, most importantly, for being my role models. And thank you, Justin Lo, for chairing my viva, and to the thesis examiners, Alexei Kochetov and Danielle Turton, for your careful examination of my work and the exciting discussion during the viva. I am very glad to have had you all on board.

It has been such a pleasure to work with wonderful colleagues at the Lancaster Phonetics Lab. It was great to be at Interspeech 2022 in Korea together with George and Christin. Thank you to my lab mates: Andrea, Bahar, Emily, Lois, Luke, Maya, Pam, Robert, Sarah, and Seren, for lovely chats about phonetics, Praat, R, food, and cats—sometimes over cake, coffee, or beer. A big thank you also to my PhD colleagues, both in and outside of LAEL: Camilo, Chris, Eleanor, Ellen, Evelin, Hikaru, Javi, Katerina, Kelly, Kevin, Maka, Tas, Valentin, and Yuze. It has been so much fun hanging out with you all, and I look forward to seeing you again somewhere in the world!

As the thesis title suggests, Japan has been the key scene in making this thesis happen. Thank you, Kazuya Saito, Yasuhiro Fujiwara, and the two “T”s (Anthony Robins and Tony Ryan) for all your support during my PhD application and beyond. Thanks to J-SLARF student members, past and present: Shungo Suzuki, Aki Tsunemoto, Takumi Uchihara, Masaki Eguchi, Ryo Maie, Yuka Naito and Atsushi

Miura, especially for helping me maintain a connection with research during the COVID lockdown. Thank you also to Naosuke Amano for your friendship and for initiating the online reading group during the pandemic. My sincere gratitude also goes to Ian Wilson, Kikuo Maekawa, and Yukiko Nota, especially for their guidance on articulatory phonetics using ultrasound and EMA during the very early stages of my PhD when I had to stay in Japan, as well as to Shoko Oiwa for introducing me to the portable ultrasound system.

I am proud of the scale of the ultrasound data collection in my PhD research, and it was made possible thanks to massive help from Noriko Nakanishi, Yuri Nishio, and Bronwen Evans, who facilitated ethics applications, room bookings, and participant recruitment both in Japan and the UK. A massive thanks to all research participants in both countries, especially for the lovely chats at the end of each recording session, which gave me so much inspiration! I am also indebted to Alan Wrench for his continuous support in using ultrasound tongue imaging and the AAA software. Thank you, Alan, for responding so quickly to my crisis emails!

I feel that I am standing where I never thought I could be, and this is all thanks to the amazing friends, colleagues, and co-researchers in the phonetics community that I have had the privilege of getting to know, both in and out of the UK. Working with Patrycja Strycharczuk has definitely helped me shape my phonetician identity through the TARDIS project and Ultrafest X in Manchester (on my birthday!). Thank you also to Gwen Brekelmans for all the support in developing the perceptual experiment; Anton Malmi for the lovely chats during our pub “hopping” in Lancaster; Rasmus Puggaard-Rode for letting me visit LMU Munich on my very spontaneous trip; Alice Léger for the exciting discussion about L2/EFL acquisition of English liquids, ultrasound, and MRI; Ander Egurtzegi and Iñigo Urrestarazu-Porta for all the exciting discussions on statistics; and Maho Morimoto for sharing the excitement of liquid consonant articulation. Together with all the phoneticians I’ve met, you have all helped me cultivate a sense of belonging to the phonetics community, which has definitely played a big role in getting me through the difficult moments during

my PhD.

Towards the end of my PhD, I moved to London, and I am grateful to the people who warmly welcomed me into the new environment. Thank you to my PIs, Patti and Chris; my post-doc colleagues, Victor, Harriet, and Eri; and my fellow PhD colleagues, Ella, Hannah, Jonas, Magda, Rongru, Ruohan, Xiayun, and Ziyun, for creating a safe and fun working environment during the final sprint of my PhD. I look forward to continuing to work with all of you in the coming years!

My PhD would never have happened without the financial support of two funding bodies: the Japan Student Services Organization (JASSO) and the Murata Science Foundation. In handling the JASSO scholarship, I am grateful to Takagi-san and everyone else in the Centre for International Exchange at Aichi University of Education, who mediated all correspondence and helped with the paperwork from the very beginning of the funding application.

I lived in Germany during most of my first year, and I got to know my lovely neighbor friends, Daniele-san and Chisa-san—thank you for making me feel at home in Esslingen! Thanks also to James, whom I met at the brewery and who helped me move within Lancaster. Beer tasted so much better while chatting with you across the bar!

I had hoped I could show this completed thesis in person to my grandfather, the late Nobuki Nagamine, who always cared about my safety in the UK. Thank you to my mom, Yukari Nagamine, and all my family members for their constant support and encouragement—I've finished all the schools!

And last but not least, thank you to my wife, Natsumi Nagamine, for your patience, kindness, dedication, and unwavering support. I am so glad to have shared this PhD journey with you, which has overlapped with our entire married life. I am very much looking forward to navigating this new chapter of our life together!

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work. The material has not been submitted, either in whole or in part, for a degree at this, or any other university. This thesis does not exceed the maximum permitted word length of 80,000 words including footnotes, but excluding the appendices and bibliography. A rough estimate of the word count is: 70,132.

Takayuki Nagamine

Publications

This is an Alternative Format thesis, whose body consists of published and publishable papers as listed below:

Chapter 4 Nagamine, T. (revised & resubmitted). Quantifying between-speaker variation in ultrasound tongue imaging data. *Journal of Phonetic Society of Japan*.

Chapter 5 Nagamine, T. (2023). Dynamic tongue movements in L1 Japanese and L2 English liquids. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 2442–2446). Guarant International. <https://guarant.cz/icphs2023/198.pdf>.

Chapter 6 Nagamine, T. (2022). Acquisition of allophonic variation in second language speech: An acoustic and articulatory study of English laterals by Japanese speakers. *Interspeech 2022*, 644–648. <https://doi.org/10.21437/Interspeech.2022-11020>

Chapter 7 Nagamine, T. (2024b). Formant dynamics in second language speech: Japanese speakers' production of English liquids. *The Journal of the Acoustical Society of America*, 155(1), 479–495. <https://doi.org/10.1121/10.0024351>

Chapter 8 Nagamine, T. (In revision). Learning to resist: Japanese speakers' production of liquid-vowel coarticulation in L2 English. *Submitted to the Journal of Phonetics*.

Chapter 9 Nagamine, T. (in prep.). L1 Japanese speakers use a single articulatory strategy to produce onset-coda allophony in L2 English liquids. *To be submitted to Language and Speech*.

Contents

1 Introduction	1
2 Research background	7
2.1 Perception-based theoretical frameworks in L2 speech learning	8
2.1.1 Speech Learning Model (SLM)	8
2.1.2 Perceptual Assimilation Model of Second Language Learning (PAM-L2)	10
2.1.3 Second Language Linguistic Perception Model (L2LP)	11
2.1.4 Section summary	14
2.2 Acoustic and articulatory characteristics of liquid consonants in Japanese and English	15
2.2.1 Japanese taps/flaps [ɾ]	15
2.2.2 English liquids /l ɹ/	18
2.2.3 Section summary	21
2.3 The acquisition of L2 English liquids by L1 Japanese speakers	22
2.3.1 Perception	23
2.3.2 Production	24
2.3.3 Section summary	29
2.4 Spatiotemporal dynamics in second language speech production	30
2.4.1 Inherent dynamics within individual segments	30
2.4.2 Dynamics and coarticulation	32
2.4.3 Longer-term speech characteristics	35

2.4.4	Section summary	36
2.5	Chapter summary and research questions	37
2.5.1	Chapter summary	37
2.5.2	Research questions	38
3	General Methodology	40
3.1	Overview	40
3.2	Participants and Ethics	41
3.2.1	L1 Japanese speakers	41
3.2.2	L1 English speakers	44
3.3	Stimuli	45
3.3.1	Production experiment	45
3.3.2	Perception experiment	48
3.4	Procedure	50
4	Ultrasound tongue imaging	57
5	Pilot study 1: Dynamic tongue movements in L1 Japanese and L2 English liquids	72
6	Pilot study 2: Acquisition of allophonic variation in second language speech: An acoustic and articulatory study of English laterals by Japanese speakers	78
7	Study 1: Formant dynamics in second language speech: Japanese speakers' production of English liquids	85
8	Study 2: Articulatory dynamics in second language speech	103
9	Study 3: Intergestural timing of English liquids in L2 speech	154
10	Summary and conclusions	228
10.1	Summary of findings	228

10.2 Contribution of the thesis	233
10.2.1 Speech dynamics provides an important language-specific phonetic detail in L2 speech learning.	233
10.2.2 L2 speakers compensate L2 speech production with existing L1 articulatory strategies.	237
10.3 Limitations, future research and concluding remarks	240
Consolidated list of references	246
Appendix A Participant information (1)	281
Appendix B Participant information (2)	286
Appendix C Participant information (3)	291
Appendix D Information sheet	296
Appendix E Consent form	305
Appendix F Demographic questionnaire	309
Appendix G Word familiarity survey	314
Appendix H Lexical frequency for the perception experiment	317

List of Tables

3.1 English target words for the production task	46
3.2 Japanese target words in the production task	47
3.3 Stimuli for perception task.	50
A.1 Speaker’s demographic information (1/4)	282
A.2 Speaker’s demographic information (2/4)	283
A.3 Speaker’s demographic information (3/4)	284
A.4 Speaker’s demographic information (4/4)	285
B.1 Speaker’s language experience (1/4)	287
B.2 Speaker’s language experience (2/4)	288
B.3 Speaker’s language experience (3/4)	289
B.4 Speaker’s language experience (4/4)	290
C.1 Speaker’s occupation and experience (1/4)	292
C.2 Speaker’s occupation and experience (2/4)	293
C.3 Speaker’s occupation and experience (3/4)	294
C.4 Speaker’s occupation and experience (4/4)	295

List of Figures

3.1 Schematised experimental protocol 51

3.2 An experimental session with an L1 Japanese-speaking participant. . 53

Chapter 1

Introduction

Second language (L2) speech learning requires considerable effort, especially for language learners after puberty. L2 speech produced by adult L2 speakers typically exhibits a trace of *foreign accents*, which is a persistent, almost inevitable characteristic (Anderson-Hsieh et al., 1992; Flege, 1986; Sereno et al., 2016). Although foreign accents could result from various sources of difficulties, including non-target-like realisations of segments (i.e., vowels and consonants) and prosody (i.e., word stress, intonation), listeners' perception of foreign accentedness has been shown to be strongly correlated with segmental factors (Saito, 2021). One well-researched example in the acquisition of L2 segments is first-language (L1) Japanese speakers' production of English liquids in their L2 English speech (e.g., Aoyama et al., 2004; Bradlow, 2008; Flege et al., 2021; Flege et al., 1995; Goto, 1971; Miyawaki et al., 1975; Saito & Munro, 2014). L1 Japanese speakers have substantial difficulty in both perceiving and producing English liquids in a manner that L1 English speakers would do (henceforth 'target-like' manner), and it has been reported that they tend to substitute English liquids with Japanese /r/ in speech production (Aoyama et al., 2004; Guion, Flege, Akahane-Yamada, et al., 2000; Riney et al., 2000). With many empirical studies conducted over time, there is a good understanding of the mechanism by which L1 Japanese speakers classify English /l/ and /ɹ/ as instances of the Japanese /r/ category, with English /l/ being

phonetically more similar to and hence more confuseable with Japanese /r/ than English /ɹ/ is (e.g., Aoyama et al., 2004; Flege, 1995; Hattori & Iverson, 2009).

This PhD thesis aims to contribute to the existing body of L2 speech production research by introducing the *articulatory* and *dynamic* aspects. First, despite the rich amount of research conducted on L1 Japanese speakers' acquisition of L2 English liquid production, it seems that the articulatory aspects have rather been taken for granted and thus rarely been addressed. This may be because of the fact that L2 speech production research has overall been developed based on acoustic findings and that obtaining articulatory data had been challenging due to technological difficulties (Colantoni et al., 2015; Mennen et al., 2010). Previous research has often inferred articulatory properties based on acoustic findings in L1 Japanese speakers' production of L2 English liquids; this, however, only results in a rather abstract understanding of how 'accented' speech is generated. This is especially the case for English /ɹ/, in which previous research demonstrates that different tongue shapes result in similar acoustic outputs (Mielke et al., 2016; X. Zhou et al., 2008). Recent technical advances mean that articulatory methods have been made accessible by a wide range of researchers; the current research deploys *ultrasound tongue imaging*, one of the vocal-tract imaging methods that is now widely used in articulatory phonetics research. One aim of the PhD is, therefore, to demonstrate that there are still many things that remain to be understood, even in such a widely studied topic of L1 Japanese speakers' production of L2 English liquids.

More importantly, this PhD research takes a *dynamic* approach to investigating English liquid quality in L2 speech production. I argue that this is a crucial aspect of moving the research field forward, as previous research on L1 Japanese speakers' production of L2 English liquids has failed to consider articulatory findings that English /l/ and /ɹ/ are inherently dynamic segments (e.g., Campbell et al., 2010; Fowler, 2015; Kirkham et al., 2019; Krakow, 1999; Sproat & Fujimura, 1993). Dynamic analysis can be complex in terms both of computation and interpretation, but various methods have been made accessible to researchers recent years, including

Generalised Additive Mixed-effect Models (GAMMs; e.g., Sós-kuthy, [2017]; Wieling, [2018]; Wood, [2017]) and Functional Principal Component Analysis (FPCA; e.g., Cronenberg et al., [2020]; Gubian et al., [2015]; Ramsay et al., [2009]). While static analysis is suitable to obtain general properties of given segments, dynamic analysis has a greater potential in better understanding the fine phonetic details involved in the production of segments, including the inherent spectral changes and coarticulatory relationships with the neighbouring segments (Turton, [2023]; Williams & Escudero, [2014]). This PhD research aims to show that dynamic properties are indeed useful in understanding particular challenges that L2 learners have not just in articulation but also in acoustics, especially with segments involving dynamic coordination of articulatory gestures such as English /l/ and /ɾ/.

Taking these two overall themes into account, I first present a review of the previous literature to provide further research background in Chapter 2, followed by general methods in Chapter 3. In Chapter 4, I discuss ultrasound tongue imaging in greater detail, in which I outline a workflow of data collection and quantitative analysis while discussing key considerations for reliable population-level tongue shape comparisons. Chapter 4 is a tutorial paper of data collection and analysis of ultrasound tongue imaging data, which has undergone a review and been re-submitted to the *Journal of Phonetic Society of Japan*.

The remainder of the thesis consists of five empirical studies. The first two papers included in Chapters 5 and 6 present preliminary studies; in Chapter 5, I look at time-varying changes in midsagittal tongue shape in vowel-liquid-vowel sequences in English and in Japanese, compared between L1 English and L1 Japanese speakers. Although this study looks at both articulation and dynamics at the same time, a greater emphasis is on the dynamic aspect in which I demonstrate that L1 influence can be seen not just within the segment itself but also in the way how a given segment interacts with the neighbouring segments (i.e., a liquid consonant coarticulated with the neighbouring vowels). This study has been published in the *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*.

The study presented in Chapter 6 is another preliminary study, published in the *Proceedings of Interspeech2022*, in which I compare acoustics and articulation of L1 Japanese speakers' production of L2 lateral allophony. This study asks whether L1 Japanese speakers realise an expected, target-like difference between English lateral consonant /l/ embedded in two different prosodic positions; word-initially (as in *leap*) and finally (as in *peel*). This study addresses the discrepancy between acoustics and articulation and claims for the need of considering the dynamic nature of lateral production in English. Findings and methodological considerations from these studies feed subsequent empirical studies presented in Chapters 7 and 8, as well as Chapter 9.

Chapters 7 and 8 focus on the dynamic properties in L1 and L2 production of English liquids. Chapter 7 presents an acoustic study published from the *Journal of Acoustical Society of America*, in which time-varying acoustic signals in the word-initial liquid-vowel sequences in English (e.g., as in *lap* and *room*) are statistically modelled using Generalised Additive Mixed-effect Models (GAMMs) and compared between L1 Japanese and L1 English speakers. The liquid consonants include English /l/ and /ɹ/ and the vowel contexts include /æ/, /i/ and /u/. Inclusion of multiple vowel contexts allows me to compare the variability of formant trajectories across vowel contexts between groups. The findings suggest that dynamic analysis is useful in identifying specific challenges that L1 Japanese speakers may have in producing English liquids, including an unclear distinction between the liquid and vowel in the liquid-/u/ sequences.

Chapter 8 has a similar focus and design to Chapter 7 but presents an articulatory study, in which I compare between-group differences in time-varying changes of tongue shapes over time during the word-initial liquid-vowel sequences. The findings overall agree with the results in Chapter 7 that L1 Japanese speakers show a greater variability than L1 English speakers in articulation of liquid-vowel sequences across the vowel contexts. In addition, this study uses Principal Component Analysis (PCA) to identify a specific midsagittal lingual

dimension, tongue dorsum raising, that differs between L1 Japanese and L1 English speakers. The trajectories representing tongue dorsum movement are further analysed using functional Principal Component Analysis (FPCA) and Bayesian hierarchical regression, suggesting differences in the magnitude of coarticulation between the two groups of speakers. The focus on both articulation and dynamics in this study allows me to propose a specific mechanism involved in the L1 transfer in articulation by referring to differences in tongue dorsum movement between L1 Japanese and L2 English liquid consonants. This chapter is currently being revised after the initial round of review, with a major revision decision, to be re-submitted to the *Journal of Phonetics*.

Chapter 9 extends the pilot study presented in Chapter 6 and addresses the allophonic variation in two different prosodic positions (word-initial and word-final) of both liquid consonants in English, /l/ and /ɹ/, produced by L1 Japanese and L1 English speakers. Despite a somewhat reduced focus on dynamics (e.g., acoustic signals and tongue shape extracted statically at liquid midpoints), this study includes an analysis of the coordination pattern between the tongue tip and tongue dorsum gestures to address gestural timing in English liquid articulation. The findings show that L1 Japanese speakers do something similar to L1 English speakers in acoustics but different in articulation; whereas the F2–F1 profiles are comparable for both onset and coda liquids, the articulatory patterns, investigated through midsagittal tongue shapes and intergestual timing between coronal and dorsal gestures, differ between the speaker groups. Replicating findings from the pilot study, the study demonstrate a complex acoustic-articulatory relationship in L2 speech production, problematising a simplistic inference of articulatory properties based on the acoustic findings. Chapter 9 is a ready-for-submission version to *Language and Speech*.

With all these empirical studies presented here, I show that looking beyond the scope of individual segments, not only in acoustics but also in articulation, offers new insights into why certain sounds are more difficult than others in L2 speech

learning and why foreign accents are persistent in adult's L2 speech production (e.g., Archibald, 2021; Flege, 1986).

This thesis has been submitted as the Alternative Format (AF) thesis consisting of a collection of manuscripts and published papers. This decision enables me to build up a publication portfolio during the PhD research and learn ways to deal with reviewing processes alongside my academic supervisors, which is an important set of skills in academia. The thesis includes one manuscript that has been accepted with minor revision and has now been resubmitted (Chapter 4), two published proceedings papers from international conferences (Chapters 5 and 6), one published journal article (Chapter 7), one manuscript with a major revision decision and is undergoing revision (Chapter 8), and one manuscript that is to be submitted for a journal (Chapter 9). All these chapters result from my original research, in which I am responsible for conceptualisation, methodology, investigation, resources, software, formal analysis, data curation, writing original manuscripts, reviewing/editing the manuscripts, project administration and funding acquisition (according to the CRediT author statement system).

Chapter 2

Research background

This chapter lays out the research context for this PhD research project, focussing on explaining how the two themes of the thesis, *articulation* and *dynamics*, have emerged. The topic of L1 Japanese speakers' acquisition of English liquids /l ɹ/ has been extensively discussed within the frameworks of L2 speech learning theories focussing on segments, including the Speech Learning Model (SLM), the Perceptual Assimilation Model for L2 learning (PAM-L2), and, to a lesser extent, the Second Language Linguistic Perception model (L2LP). At the same time, L2 speech production research has investigated ways to account for production-based constraints that may cause foreign accents. This includes 'articulatory settings', a hypothesised language-specific manoeuvre of speech articulators through a long stretch of utterance. Alongside articulatory settings, a growing body of recent research claims the importance of dynamic properties both within and beyond individual segments, including time-varying acoustic properties in liquid consonants and vowels, as well as coarticulation, interactions between individual segments. The Bilingual Coarticulatory Model (BCM; Beristain, 2022) has been developed specifically to fill in the gap in the previous literature. In this chapter, I demonstrate that previous research has attempted to obtain a better understanding of L2 speech production from various perspectives, and I argue that these realms of research point to the importance of looking beyond the scope of individual segments to better

understand L2 segmental acquisition, especially in hypothesising the effects of the learners' L1 articulation on the articulation in their L2.

2.1 Perception-based theoretical frameworks in L2 speech learning

The mechanism of L2 speech learning has been modelled mainly from the viewpoint of perceptual learning, with prominent frameworks including the Speech Learning Model (SLM; Flege, 1995) and its revised version (SLM-r; Flege & Bohn, 2021), the Perceptual Assimilation Model for L2 Learning (PAM-L2; Best & Tyler, 2007) and the Second Language Linguistic Perception Model (L2LP; Escudero, 2005; van Leussen & Escudero, 2015). These models commonly posit that L2 learners employ a similar learning mechanism that they use during L1 acquisition and that L2 speech learning takes place based on the existing L1 system. Their postulates diverge, however, in several aspects, including the scope of hypothesis with PAM-L2 and L2LP mainly in speech perception whereas SLM in both speech production and speech perception (Chang, 2019; Nagle & Baese-Berk, 2022). The theories also have different postulates in the smallest unit of perception: acoustic details (SLM, L2LP) or articulatory gestures (PAM-L2).

2.1.1 Speech Learning Model (SLM)

The most influential framework in L2 speech learning research is Speech Learning Model, originally proposed in Flege (1995) and revised later in Flege and Bohn (2021). The model assumes relatively advanced L2 learners who learn the target language in a naturalistic, immersion setting (Chang, 2019; Tyler, 2019). The model has been developed to argue against the notion of the Critical Period Hypothesis (Lenneberg, 1967), claiming that naturalistic language acquisition ceases at a certain age, approximately 13 years of age, due to a neurological maturation (Flege & Bohn,

2021). Instead, SLM claims that the language learning mechanism is intact over one's life span; although the earlier onset of learning L2 may be advantageous, even adult learners who start learning an L2 later in their life can acquire the L2 (e.g., for the acquisition of VOT contrast; Flege, 1991).

SLM explicitly hypothesises that the success of L2 speech learning depends on the formation of *phonetic categories*, resulting from the perception of the acoustic-phonetic differences between the L2 segment and its counterpart in L1. The phonetic categories for L1 and L2 coexist in a common L1-L2 phonetic space in their long-term memory, with the influence being bidirectional (Flege, 1995). The formation of phonetic categories is created for individual segments, rather than phonemic contrasts (Flege, 1995).

The core mechanism regarding the formation of phonetic categories in SLM is *equivalence classification* of position-sensitive allophones (Flege, 1995, p. 239). According to SLM, L2 learners subjectively classify nonnative phones based on how close their acoustic profile is to that of the equivalent L1 phonetic category. An L2 phone may be classified as *similar* when it “is realized in an acoustically different manner than an easily identifiable counterpart in L1” (Flege, 1987, pp. 58-59) while as *new* when it “does not have a counterpart in L1, and may therefore not be judged as being the realization of an L1 category” (Flege, 1987, pp. 59). Formation of new phonetic categories is facilitated by a greater difference between an L2 sound and its counterpart in the L1, as L2 learners are more likely to be able to detect the difference between those phones. Otherwise, an L2 segment would be equated or merged into the existing L1 category when the two sounds are perceived to be similar (Flege, 2007). Flege (1987) demonstrated that, for L1 English learners of French, a new phonetic category can be formed more easily in learning a ‘new’ sound (e.g., French /y/) than a ‘similar’ sound (e.g., French /t/).

SLM is the only model among the L2 speech learning models that makes predictions about speech production. The SLM framework states that adult L2 learners become more sensitive to position-sensitive allophonic variations in the

target language as their L2 experience increases, leading to formation of phonetic categories based on phonetic input. The establishment of mental phonetic categories brings about accurate production in L2 speech. SLM posits that the phonetic categories specify realisation rules, including “the timing, amplitude, and duration of muscle contractions that position the speech articulators in space and time” (Flege, 1992, p. 165).

2.1.2 Perceptual Assimilation Model of Second Language Learning (PAM-L2)

Another perception-based model of L2 speech learning is the Perceptual Assimilation Model for L2 Learning (PAM-L2; Best & Tyler, 2007). The foundation of PAM-L2 is the earlier Perceptual Assimilation Model (Best, 1995), which attempted to account for the perception of nonnative sounds of “functional monolinguals” (Best & Tyler, 2007, p. 16), those who are not actively learning any additional language and have little knowledge in the sound systems of a different language. PAM-L2, on the other hand, assumes L2 learners in a naturalistic, immersion setting “who are in the process of *actively learning* an L2 to achieve functional, communicative goals” (Best & Tyler, 2007, p. 16, emphasis original).

Similarly to SLM, PAM-L2 predicts that L1 and L2 phonology exists in common phonological space. PAM-L2 diverges from SLM, however, in postulating that the listeners make use of not only phonetic but also phonological information, compared to SLM which argues for the perception of acoustic-phonetic details. PAM was founded on a basis of the direct-realist approach hypothesising that a listener would perceive “the distal articulatory events that produced the speech signal” (Best & Tyler, 2007, p. 22). The smallest unit of speech in PAM-L2 is *articulatory gestures*, following Articulatory Phonology (e.g., Browman & Goldstein, 1986); according to PAM-L2, L2 learners classify the incoming L2 sound contrasts relative to the L1 phonological category based on articulatory gestures that they perceive directly in the interlocutor’s vocal tract (Best, 1995). L1 and L2 phones are perceived to

belong to the same category when “they are recognized (correctly or incorrectly) as involving functionally the same gestural constellation” (Best & Tyler, 2007, p. 26,). Best and Tyler uses this assumption to argue that L1 English-L2 French speakers would perceive French /r/, normally realised as a voiceless uvular fricative, as being equivalent to English /r/ and assimilate it into the same phonological category despite clear differences in their acoustic properties (Best & Tyler, 2007).

The core mechanism that PAM-L2 assumes is *perceptual assimilation*, i.e., assimilating L2 sound contrasts into the articulatorily closest L1 phonological category. PAM-L2 postulates several ways whereby L2 learners would assimilate nonnative sound contrasts into the L1 phonological categories, yielding perceptual difficulties. When two sounds in L2 are perceived as exemplars of two distinct phonological categories in the learner’s L1, *Two-Category (TC) assimilation* is predicted to occur. TC involves the least difficulty for the L2 learner to distinguish the L2 contrast because they can use the discrimination ability that exists in L1. Alternatively, when two nonnative sounds are perceived as equally good or poor instances of the phonological category in their L1 in *Single-Category (SC) assimilation*, the learners have greater difficulty in discriminating the two sounds than in the TC scenario. When the two sounds can be perceived as a good and a poor exemplars respectively of the equivalent L1 phonological category (*Category-Goodness* difference; CG), the difficulty is predicted to be intermediate. Finally, one of the nonnative phoneme contrasts may not be categorised in any of the L1 phonemes (i.e., *Uncategorised-Categorised assimilation*), or neither of them may not be perceived as instances in any of the L1 phonological categories (i.e., *Uncategorised-Uncategorised assimilation*).

2.1.3 Second Language Linguistic Perception Model (L2LP)

Second Language Linguistic Perception Model (Escudero, 2005; van Leussen & Escudero, 2015, L2LP,) is an explicit computational model providing a comprehensive description of the acquisition process of L2 speech perception from

being a naïve beginner to an advanced L2 user (Escudero, 2005; van Leussen & Escudero, 2015). Although the L2LP model shares the view with the other models that L2 speech learning progresses based on the learner's L1, it specifically assumes that L2 learners copy and reuse the L1 perceptual mapping patterns at the initial state of L2 learning (i.e., the *Full Copying* hypothesis). This means that L2LP assumes separate L1 and L2 perception grammars and L2 learners initially copy their L1 perception grammar, which is defined in the model as the initial state of L2 speech learning (Colantoni et al., 2015; Yazawa et al., 2020).

The model distinguishes speech perception into two broad levels: *perception* and *recognition*, while postulating four levels of representations: acoustic, phonetic, phonemic and lexical forms (van Leussen & Escudero, 2015). It explicitly differentiates *pre-lexical* perception at the lower phonetic level, involving auditory mapping of L2 contrasts onto the L1 grammar, and *lexical* recognition, involving a meaning-driven, high-level mapping. By postulating these two different levels, L2LP models the holistic path of speech comprehension, from the perception of the physical entity of speech signals to the perceptual mapping at the levels of abstract representation (Escudero, 2005).

The L2LP model hypothesises that L2 learners face two types of learning tasks that result from the mismatch between their L1 optimal grammar and the target L2 grammar (Escudero, 2005). The first is a perceptual task, involving a reconstruction of the auditory mapping pattern between incoming acoustic signals and the phonemic categories in the learner's L2 grammar, involving either redistributing the optimal L1 categories or splitting an existing L1 category into two. At early stages in L2 speech learning, L2 learners equate L2 sounds to the most resembling L1 sound in perception based on the acoustic information. The second task is a representational task, in which the phonemic categories formed in the previous perceptual task are mapped to the lexical representations; beyond the initial stage, L2 grammar develops based on the input distribution (i.e. *distributional learning*), so that the L2 learner would become an optimal perceiver of L2 while

the L1 grammar remains optimal to L1 speech perception (Escudero, 2005). The difficulty involved in L2 speech learning depends on the number of learning tasks an L2 learner is faced with.

The development of L2 grammar follows three possible scenarios similar to PAM-L2: NEW, SUBSET and SIMILAR scenarios, with varying degrees of difficulty depending on the number of learning tasks involved. First, a NEW scenario is analogous to the Single-Category assimilation in PAM-L2. In this scenario, the learners must create a new L2 category or split the existing single L1 category, and L2 learners typically perceive multiple L2 sound categories as instances of a single L1 category; for instance, L1 Spanish speakers would perceive English tense-lax vowel contrast /i/ - /ɪ/ as instances of one L1 category of Spanish /i/. The NEW scenario is predicted to be the most challenging scenario for L2 learners given that it involves not only creating new categories but also integrating them into the existing dimensions (Escudero, 2005). Second, a SUBSET scenario is a case of “multiple category assimilation” (van Leussen & Escudero, 2015, p. 2), in which a single L2 sound is perceived as an instance of multiple L1 categories. This corresponds to the Uncategorised or Categorised-Uncategorised assimilation in PAM-L2. Finally, a SIMILAR scenario is when the number of categories is the same between L1 and L2, similar to the Two-Category assimilation in PAM-L2. This scenario imposes relatively smaller difficulty on L2 learners as they would only need to adjust the boundaries of the L1 categories to match those of L2. The prediction on the level of difficulty is similar to PAM-L2, but it disagrees with SLM which postulates that similar sounds are harder to trigger the creation of phonetic categories (Flege, 1995).

Another important predictions of the L2LP model is that the degree of perceptual difficulty may vary depending on the *language mode* (Grosjean, 2008). Language mode refers to “the state of activation of the bilingual’s languages and language processing mechanisms at any given point in time” (Grosjean, 2008, p. 36). L2LP’s *language mode activation hypothesis* (Escudero, 2005, p. 120) predicts that bilingual speakers’ perceptual pattern would be different depending on which of the two

perception grammars, L1, L2 or both, are being activated, and this depends on the linguistic environment in which the perception experiment takes place (van Leussen & Escudero, 2015). L2LP hypothesises that “intermediate L1-L2 sound perception is a consequence of the gradient and parallel activation of the learner’s two perception grammars during online perception” (Escudero, 2005, p. 120). Indeed, Yazawa et al. (2020) found that the L1 Japanese-speaking participants showed different cue weighting patterns in perception depending on the language mode. In their experiment, they manipulated the language mode by changing the language of instructions (Japanese vs English) to investigate L1 Japanese-speaking participants’ perception of /i:/ and /ɪ/ in American English. The findings demonstrated the language mode hypothesis; whereas they relied more on the spectral information in the English mode, they used the duration cue in the Japanese mode.

2.1.4 Section summary

Although they differ in the exact architecture, the prevalent theoretical frameworks in L2 speech learning: the Speech Learning Model (SLM), the Perceptual Assimilation Model for L2 learning (PAM-L2), and the Second Language Linguistic Perception model (L2LP) commonly posit that L2 speech learning takes place based on the learner’s L1. Also, whether it is in acoustics (SLM, L2LP) or articulation (PAM-L2), it is commonly argued that detecting phonetic details between L1 and L2 categories facilitates L2 speech learning. In the section below, we will turn to the specific context of the current PhD research: L1 Japanese speakers’ production of English liquids. I will first review existing descriptions of the liquid consonants in Japanese and in English, which is a starting point in better understanding how the L1 and L2 categories may interact with each other.

2.2 Acoustic and articulatory characteristics of liquid consonants in Japanese and English

This section outlines key articulatory characteristics for liquid consonants in Japanese and English. The theoretical frameworks of L2 speech learning agree that L2 segments are acquired in relation to the existing L1 categories, which suggests that the acquisition of L2 articulation could also be based on the learner's articulation for the equivalent sounds in their L1. Indeed, previous research on L2 speech production hypothesises the effects of the "L1 articulatory routine" on the L2 articulation (e.g., Colantoni et al., 2021). In this section, I first provide a comparison of the articulatory properties between Japanese and English liquid consonants. The comparison demonstrates the importance of looking beyond the scope of individual segments to better understand the articulation of liquid consonants in Japanese and English, through which the *articulation* and *dynamic* aspects of the thesis mainly emerge.

2.2.1 Japanese taps/flaps [ɾ]

The perception-based theoretical frameworks of L2 speech learning commonly hypothesise an interaction between L1 categories and L2 sounds. Given this, it is possible that L1 Japanese speakers' articulation of English liquids is influenced by that of the Japanese liquid. Japanese has one liquid consonant /r/, canonically realised as alveolar tap or flap [ɾ] (Bradlow, 2008; Davidson, 2011; Riney et al., 2000). Acoustically, alveolar taps/flaps can be characterised with its brief closure duration. In American English, the intervocalic flap occurs as an allophone of /t/ and /d/, and its duration ranges between 10 and 50 ms (Fukaya & Byrd, 2005; Rimac & Smith, 1984; Zue & Laferriere, 1979), which is usually shorter than alveolar stops [t] and [d] whose closure duration could be over 100 ms (Zue & Laferriere, 1979). More recently, Morimoto (2020) demonstrates that the mean constriction duration for alveolar taps/flaps in Japanese is 32.29 ms whereas the duration for stops /t/ and

/d/ is 70.03 ms and 49.78 ms respectively. In addition to the short closure duration, alveolar taps/flaps may involve an incomplete closure or show lack of release burst, making it difficult to clearly define the taps/flaps duration in the acoustic data (Rimac & Smith, [1984]; Warner & Tucker, [2011]; Zue & Laferriere, [1979]).

Articulation of alveolar taps/flaps has been studied using various techniques, including electropalatography (EPG; e.g., Kochetov, [2018]; Recasens, [1991]; Recasens & Espinosa, [2007]), electromagnetic articulography (EMA; e.g., Morimoto, [2020]), real-time magnetic resonance imaging (rtMRI; e.g., Maekawa, [2023]) and ultrasound (e.g., Proctor, [2011]; Recasens & Rodríguez, [2016], [2017]; Yamane et al., [2015]). The short closure duration seen in acoustics for alveolar taps and flaps is made by a stop-like ballistic movement of the tongue tip (Catford, [1988]; Derrick, [2011]). Although taps and flaps are not always distinguished clearly (e.g., Lindau, [1985]), alveolar taps involve the tongue tip directly moving towards the alveolar ridge, whereas alveolar flaps exhibit the tongue tip tangentially achieving the lingual contact against the alveolar ridge in passing either upwards or downwards (Catford, [1988]; Derrick, [2011]; Ladefoged & Maddieson, [1996]). Recent articulatory studies of Japanese /r/ found evidence in favour of regarding Japanese /r/ as taps with occasional flap realisations (Maekawa, [2019], [2023]).

Beyond the canonical alveolar tap/flap realisation, Japanese /r/ shows allophonic variations which seems to depend on the phonetic context. Japanese /r/ can be realised similarly to a voiced stop [d] when it occurs utterance-initially and post-nasally (Vance, [1987], [2008]). (Arai, [2013]) suggests based on the acoustic data that Japanese /r/ may exhibit stop-like realisations, including a retroflex stop [d] word-initially and a /g/-like plosive in children's speech. This non-tap realisations of Japanese /r/ Kawahara and Matsui ([2017]) also demonstrate using EPG that Japanese /r/ may involve a smaller degree of lateral lowering given the absence of lateral contact against the palate for a singleton /r/ in the /a/ context and for a geminate /rr/. Finally, Japanese /r/ can also be realised as an alveolar trill [r] but it has been suggested that this is based on an idiosyncrasy instead of contextual

effects (Vance, 2008).

One of the major characteristics of alveolar taps/flaps [ɾ] is that the articulation is sensitive to vowel context and thus shows a greater degree of vocalic coarticulation. This is reported across languages including American English (Derrick & Gick, 2011), Catalan (Recasens, 1991; Recasens & Rodríguez, 2016, 2017) and Japanese (Maekawa, 2023; Nakamura, 2001; Recasens, 1991; Sudo et al., 1982; Yamane et al., 2015). In American English, where alveolar taps/flaps [ɾ] can occur as an allophone of unstressed intervocalic /t d/ (e.g., *better*), articulatory data show at least four subphonemic variants that are conditioned largely by the neighbouring environments: alveolar tap, down-flap, up-flap, and postalveolar flap (Derrick & Gick, 2011). In Catalan, the degree of lingual contact to the palate varies according to the tongue height and backness of the neighbouring vowels, such that it is largest when [ɾ] is flanked by high vowels [i] compared to [a] or [u] (Recasens, 1991). Similarly, variability in the midsagittal tongue body movement is greater for alveolar taps [ɾ] than for alveolar taps [r] in Catalan (Recasens & Rodríguez, 2016).

In Japanese, Maekawa (2023) demonstrates using real-time MRI that precise alveolar place of articulation for Japanese [ɾ] changes according to adjacent vowels, arguing that the global tongue body movement for [ɾ] is determined largely by neighbouring vowels. Comparing nonpalatalised and palatalised taps in Japanese, Yamane et al. (2015) also finds that nonpalatalised taps are subject to a greater degree of vocalic coarticulation in tongue dorsum, arguing for lack of gestural specifications for tongue dorsum. Taken together, these studies support the view of Recasens (1991, p. 279) that “the positioning of the tongue body does not involve much articulatory control”.

Previous research has discussed how ‘active’ the dorsal gesture is in alveolar taps/flaps [ɾ]. One view is that the degree of active involvement of tongue dorsum in a consonant results from the requirement of the manner feature of a given segment; studies suggest, for instance, that alveolar taps/flaps do not involve active tongue dorsum control, as opposed to alveolar trills [r], palatalised consonants including [ɲ]

or alveolar fricatives such as [s] or [ʃ] (Recasens, 1991; Recasens & Rodríguez, 2016). Under this view, a stronger vocalic coarticulation for alveolar taps [ɾ] in Catalan results from lingual requirements to achieve the tongue tip movement, imposing less demand on tongue dorsum movement than trills [r] do. Similar findings are found in Japanese, in which nonpalatalised taps are subject to a greater degree of vocalic coarticulation in tongue dorsum than palatalised taps, arguing for lack of gestural specifications for tongue dorsum for plain taps (Yamane et al., 2015).

When alveolar taps are compared against coronal obstruents, however, the alveolar taps seem to involve more controlled tongue dorsum. In Spanish, the degree of vocalic coarticulation across vowel contexts has been shown to be smaller for alveolar taps [ɾ] than for alveolar fricatives [ð], which could suggest that the alveolar taps or flaps [ɾ] (and the liquid consonants broadly) could be characterised with an active dorsal ‘stabilisation’ gesture (Proctor, 2011). Similarly, alveolar taps/flaps [ɾ] in Japanese show a more constant and stable tongue retraction than voiceless alveolar stop [t] throughout vowel-consonant-vowel intervals in Japanese, pointing to a commonality with dark laterals (Morimoto, 2020). These findings could suggest that alveolar taps/flaps may involve active control of tongue dorsum gesture, and on a broader scale, it might be possible that all liquid consonants could be characterised by the presence of coronal and dorsal gestures (Proctor, 2011; Proctor et al., 2019). Although it is well beyond the scope of the current PhD thesis to discuss the possibility of gestural characterisation of liquids as a phonological class, what seems common in these contrasting views is that alveolar taps/flaps exhibit a strong vocalic coarticulation compared to other members of liquid consonants. This especially contrasts with stronger resistance (i.e., smaller susceptibility) to vocalic coarticulation in English liquids /l/ and /ɹ/, as reviewed in the next subsection.

2.2.2 English liquids /l ɹ/

In contrast to alveolar taps or flaps [ɾ], previous research suggests that English /l/ and /ɹ/ requires speaker’s active control of multiple lingual and labial gestures

(Campbell et al., 2010; Delattre & Freeman, 1968; Gick, 1999; Proctor et al., 2019; Sproat & Fujimura, 1993). In this regard, English liquids could be referred to as composite segments involving coronal gestures considered as consonantal (or C-gesture) and dorsal gestures as vocalic (or V-gesture) (Fowler, 2015; Gick, 1999; Proctor, 2021; Sproat & Fujimura, 1993). The gestural coordination patterns are conditioned primarily by syllabic positions for English /l/ and, to some extent, /ɹ/, potentially explaining the well-documented allophonic variations of laterals including ‘clear’ and ‘dark’ variants (Recasens, 2012; Sproat & Fujimura, 1993).

English /l/ is an alveolar lateral approximant, whose articulation involves occlusions made in the alveolar or dental region, accompanied by the airflow around one or both sides of the tongue (Ladefoged & Maddieson, 1996). In articulation of English /l/, coronal and dorsal gestures are temporally and spatially coordinated, in which syllabic positions influence differences in the coordination pattern. This positional effect explains the widely-attested allophonic variations of ‘clear’ and ‘dark’ /l/s (Carter & Local, 2007; Recasens, 2012). Pre-vocalic /l/s, for instance, can be ‘clearer’, characterised by a greater degree of coronal constriction that occurs earlier than or synchronously to the tongue body retraction/lowering, compared to the post-vocalic ‘darker’ counterpart in which tongue body retraction/lowering usually precedes the coronal constriction (Narayanan et al., 1997; Proctor et al., 2019; Recasens, 2012; Recasens & Espinosa, 2005).

Acoustically, the clear lateral is typically associated with higher F2 frequencies and lower F1, whereas the dark laterals with low F2 and higher F1 (Ladefoged & Maddieson, 1996), in which the F2 lowering for dark /l/ is associated with retraction/raising of the posterior tongue body in English /l/ (Narayanan et al., 1997). Recasens (2012) classified lateral allophony across languages into ‘extrinsic’ and ‘intrinsic’ allophones; in the case of the former, syllable-initial and final laterals would involve two distinct articulatory targets, whereas the latter ascribes to the lateral allophony in which initial and final laterals result from the syllable positional effects (Recasens, 2012). This clear-dark distinction is the two ends of a continuum,

and the existence of ‘intermediate’ variations of laterals has been reported in North American English (Lee-Kim et al., 2013; Mackenzie et al., 2018) and British English (Kirkham et al., 2020; Turton, 2017).

Articulation of English /ɹ/ typically involves three constrictions in the vocal tract: at the lips, palate, and pharynx (Espy-Wilson et al., 2000). English /ɹ/ can be classified into two types according to the tongue shape: the tongue-tip-up ‘retroflex’ and tongue-tip-down ‘bunched’ variants, that constitute the two ends of a continuum consisting of various intermediate realisations (Delattre & Freeman, 1968; King & Ferragne, 2020; X. Zhou et al., 2008). Variability in tongue shape, however, is usually not readily perceivable auditorily as far as the lower three formants are concerned, as they all achieve lower F3 (Delattre & Freeman, 1968). Acoustic differences indicating tongue shape variability are suggested to lie in higher formants such as F4 and F5 (Guenther et al., 1999; X. Zhou et al., 2008). Furthermore, the degree to which palatal and pharyngeal constrictions contribute to F3 lowering seems to be speaker specific (Harper et al., 2020). This illustrates a complex acoustic-articulatory relationship in English /ɹ/, and it is indeed quite challenging to identify articulatory property that directly influences the acoustic output of low F3 (Hashi et al., 2003). It could thus be argued that speakers use optimal articulatory strategies to achieve a common acoustic target of low F3 for English /ɹ/ (Guenther et al., 1999; Mielke et al., 2016).

Similarly to English /l/, it is argued in some previous studies that English /ɹ/ may show a certain coordination pattern of articulatory gestures depending on the syllabic position. Campbell et al. (2010) finds that word-initial /ɹ/ in Canadian English exhibits a front-to-back sequence in gestural coordination; the bilabial gesture precedes the tongue body gesture followed by the tongue root gesture. The final /ɹ/ tokens, on the other hand, involves the tongue root gesture being initiated before the labial gesture, and the tongue body gesture is activated after these two gestures. The initial /ɹ/ shows a somewhat reduced tongue root gesture but a greater magnitude of tongue body and labial gestures compared to final /ɹ/.

Whereas previous research has looked into gestural timing of English liquids focussing either on laterals or rhotics separately, Proctor et al. (2019) compares gestural timings of both English /l/ and /ɹ/ across onset and coda positions based on MRI images of four American English speakers producing monosyllabic words (e.g., *leap*, *reap*, *peel*, *beer*). Although their results largely agree with the previous research in that both English /l/ and /ɹ/ show spatiotemporal coordination patterns interacting with the syllabic position, one major finding that somewhat differs from the previous research is that they did not find evidence of pharyngeal constrictions with a tongue root gesture of /ɹ/. This leads them to argue that English /ɹ/ is instead characterised as having coronal and dorsal gestures as well as the labial gesture, allowing for framing English liquids using the same set of articulatory gestures (Proctor et al., 2019).

In addition to the gestural coordination patterns within the liquid segments as described so far, both English /l/ and /ɹ/ show certain degrees of coarticulatory influence from the neighbouring vowels. The degree of coarticulation is inversely correlated with the degree of constraints imposed on tongue body, such that segments whose tongue body is highly constrained should show a lesser degree of vocalic coarticulation (Recasens, 2012). English /ɹ/ shows a greater resistance to coarticulation than English /l/ does, given different degrees of variance in tongue shape displacement across different vowel contexts (Proctor et al., 2019). In addition, clear /l/s exhibit a similar degree of coarticulatory resistance as alveolar taps/flaps in Catalan (Recasens & Rodríguez, 2016).

2.2.3 Section summary

In this section, both articulatory and acoustic characteristics of the liquid consonants in Japanese and English are reviewed. Japanese has one liquid consonant /r/, which is canonically realised as an alveolar tap or flap [ɾ]. It involves a ballistic movement of the tongue tip making contact with the alveolar ridge, with the closure duration usually shorter than non-flap consonants like alveolar stops /t d/. Whereas alveolar

taps/flaps seem to be produced solely with the tongue tip, English liquids involve a global movement of the tongue, especially in the tongue tip and tongue dorsum/root regions. A clear difference between Japanese and English liquid consonants can be seen in the coarticulatory patterns between the liquid and vowels. While English liquids show resistance to vocalic coarticulation, to a greater degree for English /ɹ/ than for English /l/, Japanese /r/ is shown to be susceptible to vocalic coarticulation in which tongue body movement is largely dominated by the neighbouring vowels. In addition, it is challenging to infer articulation of English /ɹ/ based solely on acoustics. These considerations constitute the basis of the *articulatory* and *dynamic* approaches to the L2 production of English liquids, and in the section below, I will review how these aspects have (not) been addressed in the previous research.

2.3 The acquisition of L2 English liquids by L1 Japanese speakers

L1 Japanese speakers' acquisition of English liquids has been studied in previous research extensively, both in terms of perception (e.g., Aoyama et al., 2004; Best & Strange, 1992; Hattori & Iverson, 2011; Lively et al., 1993; Miyawaki et al., 1975; Sheldon & Strange, 1982; Shinohara & Iverson, 2018) and production (e.g., Aoyama et al., 2019; Aoyama et al., 2023; Flege et al., 1995; Saito & Munro, 2014; Saito & van Poeteren, 2018). Despite the rich amount of previous research, there are two areas in which further evidence would merit; first, L1 Japanese speakers' articulation of English has almost always been inferred from acoustic findings, which do not always provide the most accurate picture regarding what is happening in the learner's vocal tract. Second, the research reviewed here suggests that comparisons of the liquid articulation between Japanese (L1) and English (L2) necessitate dynamic information. In this section, I first discuss some key findings in the research on L1 Japanese speakers' production of L2 English liquids. I then review a handful articulatory studies that exist in this research context, pointing the need for further

research.

2.3.1 Perception

As reviewed earlier, the perception-based models of L2 speech learning share the broad view that L2 learners acquire L2 segments in relation to the equivalent L1 segments (Best & Tyler, 2007; Escudero, 2005; Flege & Bohn, 2021). In the current research context, it is well-known that L1 Japanese speakers have substantial difficulty in acquiring the English liquid contrast. An obvious account for this difficulty is the cross-linguistic difference in phoneme inventory; while English has two liquid phonemes /l/ and /ɹ/, Japanese has only one liquid phoneme /r/. Because of this mismatch in the number of liquid categories, English /l/ and /ɹ/ are not mapped well onto the Japanese liquid category, and L1 Japanese speakers may therefore perceive English /l/ and /ɹ/ as an instance of Japanese /r/ (Bradlow, 2008). Previous studies also demonstrate that L1 Japanese speakers confuse English liquids with labial-velar approximant /w/ or a high back vowel /u/ (Best & Strange, 1992; Guion, Flege, Akahane-Yamada, et al., 2000; Mochizuki, 1981), although previous empirical research has shown a rather strong influence of Japanese /r/ (Aoyama et al., 2004; Guion, Flege, Akahane-Yamada, et al., 2000; Takagi, 1993).

The influence of Japanese /r/, however, is not uniform between the two English liquids. This is best captured by the SLM's concept of *perceptual similarity*; according to the SLM, the success of L2 speech learning depends on formation of phonetic categories, which is facilitated when L2 learners detect acoustic differences between a given L2 phone and the closest equivalent L1 category (Flege, 1995; Flege & Bohn, 2021). Previous research shows that the perceived similarity of English /l/ and /ɹ/ varies against Japanese /r/, such that L1 Japanese speakers perceive English /l/ as being more similar to Japanese /r/ than is English /ɹ/ (Aoyama et al., 2004; Aoyama et al., 2008; Guion, Flege, Liu, et al., 2000). For example, L1 Japanese speakers perform better in discrimination between English /ɹ/ and Japanese /r/ than between English /l/ and Japanese /r/, suggesting a greater degree of dissimilarity

between English /ɹ/ and Japanese /r/ (Guion, Flege, Liu, et al., 2000). Because of this, L1 Japanese speakers, especially children, improve perceptual accuracy of English /ɹ/ to a greater extent than that of English /l/ due to a greater perceptual dissimilarity (Aoyama et al., 2004).

Under the PAM account, on the other hand, it is not very clear what assimilation patterns L1 Japanese speakers would exhibit in perceiving the English /l ɹ/ contrast. It could be via the Single-Category assimilation (Best & Strange, 1992), in which English /l/ and /ɹ/ are uniformly perceived as poor exemplars of Japanese /r/. This scenario agrees with the broader notion that discrimination of English liquids involves considerable difficulty for L1 Japanese speakers (Hattori & Iverson, 2009). A closer investigation of L1 Japanese speakers' best exemplars for English /l/ and /ɹ/, however, suggests Category-Goodness or Categorized-Uncategorized assimilations (Hattori & Iverson, 2009); they found that English /l/ and /ɹ/ differ in the category goodness within the Japanese /r/ category, with English /ɹ/ being perceived as more dissimilar from Japanese /r/ than English /l/. These findings rather agree with the *perceived similarity* account in SLM (e.g., Aoyama et al., 2004). Overall, these pieces of evidence seem to suggest that it is reasonable to understand that L1 Japanese speakers perceive English /l/ and /ɹ/ in relation to the Japanese /r/ category, in which English /ɹ/ is easier for them to discern from Japanese /r/ than English /l/ due to perceptual dissimilarity. The exact nature of L1-L2 mapping, however, requires further research.

2.3.2 Production

The SLM assumes that perception and production are linked, based on the observations that many production errors have perceptual basis, suggesting that the effects of perceptual learning could also be mirrored in speech production (Flege, 1995; Nagle & Baese-Berk, 2022). While it is not clear whether perception should strictly precede production in a uni-directional manner, the model still views perception as central to L2 speech learning given the role of perceived

similarity in category formation (Flege et al., 2021). SLM observes that articulatory commands are specified in the perceptual representation of a given L2 sound, including “the amplitude and duration of muscular contractions that position the speech articulators in space and time” (Flege, 1992, p. 165). Identifying L2 sounds as instances of a given L1 category, L2 learners redeploy and refine articulatory strategies for the L1 category to produce the L2 sounds (Flege et al., 1986). This assumption suggests that L2 learners who have better perceptual accuracy might be able to articulate L2 sounds more accurately than those who have less accurate perceptual realisation, resulting in a greater accuracy in their L2 speech production.

Previous perception training studies suggest that the asymmetry in L1 Japanese speakers’ perception of English /l/ and /ɹ/ could also be observed in production. L1 Japanese speakers who received intensive perceptual training through high variability phonetic training (HVPT) paradigm improved both perception and production, but the magnitude of improvement in production is larger for English /ɹ/ than for English /l/ (Bradlow et al., 1997; Shinohara & Iverson, 2018). In Bradlow et al. (1997), identification training of English /l/ and /ɹ/ has resulted in overall improvement in L1 Japanese speakers’ production of English liquids before and after the training, assessed by the proportion of tokens correctly identified by L1 English-speaking listeners. The production accuracy was overall better for English /ɹ/ than for English /l/. Similarly, a combination of identification and discrimination training of English /l/ and /ɹ/ for 41 adult L1 Japanese speakers resulted in a greater degree of F3 lowering for English /ɹ/ compared to F3 raising for English /l/ (Shinohara & Iverson, 2018). While it is unclear to how exactly perception and production are linked with each other, these studies still suggest that the SLM’s postulations regarding the asymmetric tendency in perceptual difficulties could also be extended in production.

In terms of evaluating L1 Japanese speakers’ production, the majority of previous research employs perceptual evaluation by L1 English-speaking listeners (e.g., Aoyama et al., 2004; Bradlow et al., 1997; Riney et al., 2000). Instrumental

studies exist, but they almost exclusively focus on acoustic properties (Aoyama et al., 2019; Aoyama et al., 2023; Flege et al., 1995; Saito & Munro, 2014; Saito & van Poeteren, 2018), while there is a handful of studies looking into articulatory data (Masaki et al., 1996; Moore et al., 2018; Zimmermann et al., 1984).

Previous acoustic studies investigate how L1 Japanese speakers produce English liquids are acoustically characterised in terms of the second (F2) and third formants (F3). English /l/ typically shows a relatively high F3 that is well separated from F2, whereas English /ɹ/ can be typically characterised by notably low F3 frequencies at approximately 1,500 – 2,600 Hz that are very close to F2 (Alwan et al., 1997; Delattre & Freeman, 1968; Espy-Wilson et al., 2000; Narayanan et al., 1997; Stevens, 2000). It was found that L1 Japanese speakers have difficulty in lowering the F3 frequencies for English /ɹ/ as low as that of L1 English speakers because F3 is not used in L1 Japanese to distinguish phonological contrasts, resulting in their different cue weighting patterns (Aoyama et al., 2019; Iverson et al., 2003; Saito & Munro, 2014; Saito & van Poeteren, 2018). As a consequence, L1 Japanese speakers distinguish English liquids along the F2 dimension instead of the more difficult F3; e.g., lower F2 for English /ɹ/ than for English /l/ (Aoyama et al., 2019). Although F2 is a less reliable cue in contrasting English /l/ and /ɹ/, it is shown to correlate with L1 English-speaking listener's judgements of perceived intelligibility, in which the lower F2 the better L1 English-speaking listeners identify the token as English /ɹ/ (Aoyama et al., 2023). Overall, these results suggest that L1 Japanese speakers make a contrast between English /l/ and /ɹ/ in a different manner from L1 English speakers using the F2 dimension that does contribute to the increased perceived intelligibility. Different acoustic profile also suggests that L1 Japanese speakers may employ different articulatory strategies from L1 English speakers (Aoyama et al., 2023; Saito & van Poeteren, 2018).

In contrast to ample research on perception and acoustics, however, articulatory properties of L1 Japanese speakers' production of English liquids remain poorly understood. Flege (1992) states that establishment of mental phonetic categories

brings about accurate production in L2 speech as the mental phonetic categories specify realisation rules, including “the timing, amplitude, and duration of muscle contractions that position the speech articulators in space and time” (p. 165). At the same time, SLM accounts for L2 speech learning almost exclusively from the acoustical viewpoints, and it remains unclear how L2 learners figure out articulatory parameters based on acoustics when articulatory characteristics are not immediately clear in acoustic signals, especially in the case of English /ɹ/ (Delattre & Freeman, 1968; Mielke et al., 2016; Tiede et al., 2004).

A handful of studies have investigated the articulatory properties of articulation of English /l/ and /ɹ/ by produced Japanese speakers (Masaki et al., 1996; Moore et al., 2018; Zimmermann et al., 1984). Zimmermann et al. (1984) utilised X-ray to image the vocal tract of two Japanese speakers and one American English speaker. The Japanese speakers were “inexperienced” and “experienced” learners of English, residing in the US for three years and seven years in the US respectively. The researchers imaged their vocal tract during utterances of the target words “lap”, “rap”, “lorry”, “laurel”, and “parallel” in a carrier sentence “It’s a ___.” The findings demonstrate that the inexperienced Japanese speakers showed little movement in the tongue posterior portion compared to the experienced speaker. They argued that a lack of tongue body retraction and the labial gestures (i.e., rounding and protrusion) characterises the L1 influence from Japanese. In addition, it was noted that the production by the less experienced L1 Japanese - L2 English learner was influenced by vowel context. Despite the small sample size, this study provides a clear articulatory difference between L1 and L2 English speakers, particularly regarding the differences in the tongue posterior movement, which might be a redeployment of the tongue stabilisation strategy in producing the taps/flaps in Japanese.

A subsequent study conducted by Masaki et al. (1996) using magnetic resonance imaging (MRI) identified four types of articulation from five L1 Japanese-speaking and five L1 English-speaking participants’ production of English /l/ and /ɹ/. The

production quality was also acoustically evaluated along the F3–F2 dimension and perceptually through identification tasks by L1 English-speaking listeners. In articulation, the five L1 English speakers showed similar patterns in the vocal tract configuration, involving the primary constriction in the palatal region with the presence of the sublingual cavity for /ɹ/ or an apico-alveolar contact for /l/. A secondary constriction was also observed in the posterior part of the tongue and the pharyngeal region for both sounds. The Japanese speakers’ articulatory strategy, on the other hand, was categorised into four categories (Types A - D). Type A speakers showed a similar vocal tract configuration with that of English speakers. Type B speakers were somewhat similar to the Type A speakers in the presence of the sublingual cavity and distinctive tongue shapes for English /l/ and /ɹ/ respectively. The magnitude of the contrast between English /l/ and /ɹ/, however, was small. Type C speakers did not show a sublingual cavity, and English /ɹ/ was understandably confused with English /l/ by L1 English-speaking listeners in the identification task. Finally, while Type D speakers did show the sublingual cavity in their production of English /l/ and /ɹ/, they used similar tongue shapes for both /l/ and /ɹ/, mainly using the tongue shape for English /ɹ/ when producing English /l/.

One of the interesting findings was the relationships between the articulatory configuration classifications and the general tendency of the acoustic output and listeners’ evaluation; the /l/ and /ɹ/ tokens produced by Type A speakers were the most native-like in terms of acoustics and identification accurately. The acoustic quality and identifiability decreased in the order of Types B, C, and D. From these findings, the authors argued that “the characteristics of lingual contact/constriction and the sublingual cavity observed in the midsagittal vocal tract can provide crucial information for evaluating the validity of Japanese strategies of /r/ and /l/ production” (Masaki et al., 1996 p. 1584).

Taking a similar approach in classifying tongue shape configurations, Moore et al. (2018) used electromagnetic articulography (EMA) to investigate four L1 Japanese

speakers' articulatory strategies for English /l/ and /ɹ/. The participants varied in their L2 proficiency, clustered into three groups: Advanced ($n = 1$), Upper Intermediate ($n = 1$), and Lower Intermediate ($n = 2$). As a control, two L1 English speakers from the US and Australia also took part in the study. They recorded articulatory and acoustic data of three tokens of initial singleton liquids (two /l/s and one /ɹ/) and six tokens of liquids in the initial cluster (three /l/s and three /ɹ/s). Based on the articulatory data, they identified seven tongue shapes: 'retroflex', 'bunch', 'cup', 'cup in retroflex', 'flat', 'reach', and 'hunch', where the choice of the articulatory strategies partly correlates with the speakers' English proficiency. A lower intermediate speaker, for instance, employed a single tongue shape to produce both English /l/ and /ɹ/, although this was considered to be a different strategy from Japanese /r/ judging from a lower F3 in acoustics. In contrast, the other lower intermediate speaker utilised a diverse, inconsistent articulatory strategies, with some overlaps between English /l/ and /ɹ/. Finally, the advanced speaker preferred one articulatory strategy for /ɹ/ and nearly consistent tongue shapes for /l/, with a clear differentiation along the F3 dimension in acoustics, suggesting that this speaker has developed two distinct categories for the English /l/ and /ɹ/ production. The classification of the tongue shapes was based on qualitative and subjective judgements, but the findings illustrate a possible developmental path in articulation in relation to the English proficiency of the L1 Japanese-L2 English learners.

2.3.3 Section summary

L1 Japanese speakers' acquisition of L2 English liquids has been investigated extensively in the previous literature in terms of perception, production and the link between them. It is generally understood that L1 Japanese speakers perceive and produce L2 English /l/ and /ɹ/ as an instance of the equivalent category in L1, Japanese /r/, and their perception seems to be influenced by phonetic characteristics of the stimuli, giving rise to a percept that is similar to /w/ or /u/. In production,

the L1 influence can be inferred by their substitution pattern of English /l/ and /ɹ/ with an alveolar tap/flap [ɾ]. The majority of the instrumental studies focus on acoustic properties, in which they suggest that L1 Japanese speakers rely more on F2 instead of F3 when making a contrast between English /l/ and /ɹ/.

However, the articulatory dimensions that L1 Japanese speakers struggle with in producing target-like English liquids are largely unknown, due partly to the lack of articulatory research. A few previous articulatory studies, nevertheless, suggest that L1 Japanese speakers employ a wide range of tongue shapes to produce English /l/ and /ɹ/, which becomes more consistent as their proficiency goes up. The tongue shape difference is, however, not necessarily observable in the acoustic data, similar to the previous claims for L1 English speakers' production of English /ɹ/. Finally, it is suggested that the influence of L1 articulation could be observed along the tongue posterior portion, which might reflect the difference between Japanese and English liquids in the way tongue body movement is coarticulated with the vowel context.

2.4 Spatiotemporal dynamics in second language speech production

2.4.1 Inherent dynamics within individual segments

Theoretical frameworks emphasise the importance of language-specific, phonetic details in L2 speech learning; SLM, for instance, postulates that “language-specific aspects of speech sounds are specified in long-term memory representations called *phonetic categories*” (Flege, 1995, p. 239). Best and Tyler (2007, p. 19) argues in PAM-L2 that “SL [second language] learners' perception of L2 contrasts varies systematically according to L1 phonotactic, allophonic, and coarticulatory patterning”. Within the L2LP framework, Escudero (2005, p. 261) argues that “the model's Optimal Perception hypothesis states that perception strongly depends on the specific production environment, so that the optimal way of perceiving the

sounds of a language depends on how such sounds are produced.” Despite differences in the number of postulates and hypotheses, these models seem to agree for the importance of language-specific phonetic details in L2 speech learning.

Our understanding of these language-specific phonetic details, however, is often confined in the size of ‘segments.’ The SLM is based on a lot of previous research investigating voice onset time (VOT) in stop consonants; e.g., Spanish-English bilinguals produce English stop consonants with an intermediate VOT value because of the influence of Spanish exhibiting short-lag VOT (Flege, 1991). The L2LP framework is built upon investigations of vowel production, in which the spectral characteristics of vowels are often extracted at vowel midpoint (Escudero, 2005).

Through the comparison of the articulatory characteristics of Japanese and English liquids so far, however, it has become apparent that the liquid consonants in Japanese and English differ not only in their general articulatory descriptions, but also in time-varying coarticulation as a function of neighbouring vowels, suggesting that dynamic information can be important in better understanding the acquisition of English liquid acoustics and articulation. While the acoustic (e.g., formant frequencies) and articulatory properties (e.g., tongue shape) of L1 Japanese speakers’ production of English liquids have been evaluated at single point in time, such as at the F3 minimum, the consonantal onset or the spectral release (e.g., Flege et al., 1995; Masaki et al., 1996; Moore et al., 2018; Saito & Munro, 2014), an increasing amount of evidence suggests that time-varying properties are also important characteristics and thus should be addressed in L2 speech research (Beristain, 2022; Espinal et al., 2020; Schwartz & Kaźmierski, 2020). The importance of dynamic properties could also be drawn from the articulatory descriptions of English liquids, in which English /l/ and /ɹ/ exhibit certain spatiotemporal coordination among articulatory gestures (Browman & Goldstein, 1995; Campbell et al., 2010; Gick, 1999; Sproat & Fujimura, 1993). It could be argued, therefore, that, while a static analysis would be sufficient to describe a general pattern of liquid consonants, understanding the complex nature of English liquid production would call for a

time-varying, dynamic analysis that would uncover further phonetic details (Turton, 2023).

One of the aims of this PhD thesis is to demonstrate that the dynamic information provides specific evidence uncovering which acoustic/articulatory characteristics hinder L1 Japanese speakers from producing L2 English /l/ and /ɹ/ in a target-like manner. The dynamic approach has the possibility of capturing the dynamic nature of English liquids and uncovering coarticulatory patterns between the liquid and the neighbouring segments, which could be another source of foreign accents (Beristain, 2022; Espinal et al., 2020; Reidy, 2016). Dynamic analysis has been widely used in recent speech production research to study inherent spectral changes in vowels (Watson & Harrington, 1999), liquids (Howson & Redford, 2021; Kirkham et al., 2019) and fricatives (Reidy, 2016; Wikse Barrow et al., 2022). In the context of L2 speech production, Schwartz and Kaźmierski (2020) found that L1 English and L1 Polish-speaking English learners differ in their spectral changes for /ɛ/ and /æ/, in which the F2–F1 trajectory for the L2 Polish learners of English is characterised with a later rise compared to that of L1 English speakers. Similarly, L1 Korean-speaking learners of English and L1 English speakers differed significantly in the time point where changes in the F3–F2 occurred for word-final American English liquids /ɹ/, /l/ and /ɹl/ (as in ‘Carl’) (Espinal et al., 2020). Overall, these findings suggest that the non-static nature and the coarticulatory patterns associated with English /l/ and /ɹ/ could impose an additional layer of difficulty for L1 Japanese speakers of English. The dynamic analysis will, therefore, help us understand time-varying acoustical properties that characterise the production of L2 English liquids which tend to be lost or averaged out inevitably in the static analysis.

2.4.2 Dynamics and coarticulation

Dynamic properties in L2 speech production could not only span individual segments but could also be shown over multiple segments, manifested through the process of coarticulation. Coarticulation can be defined broadly as “patterns of coordination,

between the articulatory gestures of neighbouring segments, which result in the vocal tract responding at any one time to commands for more than one segment” (Manuel, 1999, p. 179). Empirical evidence suggests that each language shows specific patterns of coarticulation between segments and thus need to be learnt during language acquisition (Beristain, 2022; Keating, 1985; Öhman, 1966).

Coarticulation can be understood in terms of *resistance* and *aggressiveness* (Proctor et al., 2019; Recasens & Rodríguez, 2016). Coarticulatory resistance is defined by Recasens and Espinosa (2009, p. 2288) as “a measure of its degree of articulatory variability as a function of phonetic context.” Coarticulatory resistance of a given segment correlates with the degree of articulatory constraints; the degree of articulatory constraint (DAC) model of coarticulation suggests that the major constraints would be imposed by the degree of tongue dorsum constraint (Recasens & Espinosa, 2009; Recasens et al., 1997). Coarticulatory aggressiveness positively correlates with coarticulatory resistance, such that segments with a higher degree of coarticulatory resistance involve a greater constraint on articulation (e.g., on tongue dorsum), which would then exert strong coarticulatory influence on the neighbouring segments (Recasens & Espinosa, 2009). For example, consonants that involve a greater demand on tongue dorsum movement, such as (alveo)palatal consonants /ʃ/, /ɲ/, as well as a high front vowel /i/, would exhibit a greater degree of resistance to coarticulation as well as exert a stronger coarticulatory influence on the neighbouring segments at the same time, as opposed to labial consonants such as /p/ that impose little constraint on tongue dorsum in articulation (Recasens & Espinosa, 2009).

Previous studies of consonants in Catalan show that alveolar taps [ɾ] involve minimal tongue dorsum movement in their articulation, corresponding to a smaller degree of coarticulatory resistance, which means that articulation of alveolar taps is highly susceptible to the vowel context (Recasens, 1991; Recasens & Rodríguez, 2016, 2017). A study in Japanese echoes this observation, arguing that the tongue shape for alveolar taps and flaps in Japanese could be determined largely by the phonetic context (Maekawa, 2023). English liquids, on the other hand, exhibit

a greater coarticulatory resistance given the active involvement of tongue dorsum in their articulations (Proctor et al., 2019; Recasens, 2012; Recasens & Espinosa, 2005). Previous research shows that English /ɹ/ is more resistant to coarticulation than English /l/ (Proctor et al., 2019). It has also been shown that dark /l/s exhibit a greater degree of coarticulatory resistance than clear /l/s given a greater degree of tongue retraction involved for dark /l/s (Recasens & Rodríguez, 2016). Although clear /l/s can be similar in the degree of coarticulatory resistance with alveolar taps [ɾ], laterals overall should still be more resistant to coarticulation than non-lateral consonants due to the laterality requirement and tongue retraction that is still involved in clear laterals (Recasens & Rodríguez, 2016).

Beristain (2022) has recently proposed the Bilingual Coarticulatory Model (BCM) based on his findings that L1 English and L1 Spanish speakers show different anticipatory coarticulation patterns. English and Spanish differ in the timing of nasalisation in consonant-vowel-nasal (CVN) and consonant-vowel-nasal-vowel (CVNV) sequences, in which Spanish exhibits a later onset of nasalisation with a lesser degree compared to English (Beristain, 2022). The results indicated that L1 English-speaking learners of Spanish transferred the timing pattern of nasalisation from L1 English to L2 Spanish as they exhibited earlier onsets for both CVN and CVNV sequences in their production of L2 Spanish (Beristain, 2022). Furthermore, linguistic proficiency was inversely correlated with the degree of L1 transfer, such that more advanced L1 English-speaking learners of Spanish showed a more native-like coarticulation pattern. Corroborated by these findings, the BCM proposes that L2 speakers could use their L1 coarticulatory patterns in their L2 speech production similarly to the case for articulation of segments and that L2 learners with a higher L2 proficiency would be more able to adjust their coarticulatory patterns than those who are less proficient (Beristain, 2022).

2.4.3 Longer-term speech characteristics

Even on a longer domain than coarticulation, previous research has suggested that each language might be characterised by *articulatory settings*. Articulatory settings refer to “the gross oral posture and mechanics” (Honikman, 1964 p. 73) that coordinates movements of speech organs so that “the facile accomplishment of natural utterance” can be made (Honikman, 1964, p. 73). Scholars use different terminology to distinguish a range of long term characteristics in speech, including phonetic settings (Laver, 1994; Mennen et al., 2010), voice quality settings (Esling & Wong, 1983), and the base of articulation (Recasens, 2011). Despite slight differences in scope among researchers, the consensus here is that each language has an optimised, habitual articulatory manoeuvre to make the transition from one segment to another smoothly and economically in a given language. Earlier impression-based documentations suggest that an overall settings of English can, for example, be characterised with a loose jaw, neutral lips activity and the tongue anchored to the roof laterally, as opposed to French in which the jaw is slightly open, lips are usually rounded and the tongue is anchored to the floor centrally (Honikman, 1964). Similarly, Someda (1966) provides a description of articulatory settings in Japanese, suggesting that the jaw is more open than English, lip movement is very little, and the tongue is anchored to the floor centrally similarly to French. A more recent instrumental study using real-time magnetic resonance imaging (rt-MRI) demonstrates that articulatory settings, measured based on the articulatory posture during grammatical pauses, “afford large changes with respect to speech tasks for relatively small changes in lower-level speech articulators” (Ramanarayanan et al., 2014, p. 7). Furthermore, some studies suggest that articulatory settings may differ at the level of regional dialects in Dutch (Wieling & Tiede, 2017) and in Catalan (Recasens, 2011).

One way to account for foreign accents in L2 speech production would be by assuming differences in settings; in the acquisition of L2 English, for instance, “the second language learner may impose the new phonemes of English on the

old background posture of a non-English, and perhaps inappropriate, voice quality setting” (Esling & Wong, 1983, p. 90). It is challenging, however, to test the existence of articulatory settings empirically given a possible confound of effects; it is not yet clear, for example, whether the settings are the accumulation of slight differences in phonetic implementations of phonemes in a given language or whether it is indeed the settings that define the phonetic details of the phoneme production (Mennen et al., 2010). Nevertheless, empirical studies have demonstrated that proficient L2 speakers may develop distinct interspeech postures (ISP) in French-English bilinguals (Wilson & Gick, 2014), L1 Polish-L2 English speakers (Święciński, 2013), and L1 Japanese-L2 English speakers (Wilson & Kanada, 2014). However, a clear difference between L1 and L2 articulatory settings is not always found; one French-English bilingual participant in Wilson and Gick (2014) developed an ‘intermediate’ articulatory settings that approximate neither of the two languages. Also, while Colantoni et al. (2021) found that their French-English bilingual participants seemed to show systematic changes in sub-phonemic place of articulation among coronal consonants, the shifting pattern did not correlate with their oral proficiency. They also found that sonorants (especially laterals /l/) showed a clearer difference in place of articulation between the participants’ French and English production.

2.4.4 Section summary

Both acoustics and articulation of a given segment can be characterised not only with a static property captured by single-point measurements but also with dynamic properties. As reviewed in the earlier sections, the possibility that Japanese and English liquids may differ in coarticulatory patterns suggests that the dynamic measurement would provide further insights into the specific challenges that L1 Japanese speakers face in acquiring target-like production of English /l/ and /ɹ/. This chapter demonstrates that the language-specific nature of long-term speech characteristics has been implicated in research on articulatory settings, and more

recent research explicitly argues for the role of coarticulation in L2 speech learning.

2.5 Chapter summary and research questions

2.5.1 Chapter summary

This chapter outlined key theoretical backgrounds and relevant empirical findings related to L1 Japanese speakers' acquisition of L2 English liquid production. The chapter first introduced three most prevalent theoretical frameworks in L2 speech learning that focus on explaining perceptual difficulty (PAM-L2, L2LP) and on the link between perception and production (SLM, SLM-r). While they diverge in several ways, they commonly posit that L2 speech learning takes place in relation to the learner's L1 and that L2 speech learning considers phonetic details in determining how L2 sounds are categorised. They also commonly focus on explaining the acquisition of individual segments.

Given the L1-L2 interaction, I have then compared acoustic and articulatory characteristics between Japanese and English liquid consonants to identify potential differences. A greater focus was placed on articulatory properties, where it was suggested that Japanese and English liquid consonants differ in the way they interact with neighbouring vowels, which could be one of the sources of difficulty for L1 Japanese speakers. Specifically, the articulation of alveolar taps/flaps [ɾ], a canonical articulation of Japanese /r/, are susceptible to vowel contexts, whereas English /l/ and /ɹ/ show greater resistance to vocalic coarticulation.

I have also argued that, despite articulatory differences emerging from the comparison, previous research on L1 Japanese speakers' acquisition of L2 English liquids has not fully addressed this, which could be due to the extensive focus on acoustics in previous research. Nevertheless, the few existing articulatory studies suggest that (1) L1 Japanese speakers' articulation is as variable as L1 English speakers' articulation and (2) articulation does not necessarily correlate with acoustic outputs. In addition, one study demonstrates that L1 Japanese speakers

who are less experienced in L2 English show little movement in the tongue posterior and a greater variability as a function of vowel contexts, compared to the more experienced L1 Japanese - L2 English speaker and the L1 English speaker. These considerations point to the need of dynamic analysis, which would uncover time-varying properties that could better characterise the liquid consonants (especially in English) and that could better capture the interactions between the liquid consonants and the neighbouring segments.

2.5.2 Research questions

Given this background, the running themes of this PhD thesis are *articulation* and *dynamics*, with the aim of investigating how dynamic information involved in the production of L2 segments could hinder L2 learners from producing L2 segments in a target-like manner. The testing case here is L1 Japanese speakers' acquisition of L2 English liquid production. Previous literature shows that the liquid consonants in Japanese and English may have different gestural coordination patterns, especially in terms of the degree of active involvement of tongue dorsum; while the tongue tip and dorsum gestures are actively involved in the production of English liquids, resulting in the position-dependent gestural coordination pattern and a certain degree of resistance to vocalic coarticulation, it is suggested that the degree of tongue dorsum involvement is small in the production of the Japanese liquid consonant. Such differences in gestural coordination could surface as different patterns of (1) liquid-vowel coarticulation and (2) position-dependent allophonic variation. This PhD thesis therefore investigates these two strands of research by combining acoustic and articulatory methods.

The overarching research question presented in this research is:

How do L1 Japanese speakers make use of the dynamic, time-varying phonetic cues in their production of L2 English liquids?

This question is broken down into three sub-questions, with a particular focus on

the dynamic, time-varying properties in the English liquid acoustics and articulation. Each of the following studies corresponds to an individual chapter presented in the later sections of the thesis:

- **Formant dynamics in L2 speech production (Chapter 7):** How do the time-varying spectral properties differ in L1 and L2 English liquids?
- **Articulatory dynamics in L2 speech production (Chapters 5 and 8):** How do articulatory dynamics differ between L1 and L2 English liquid productions?
- **Gestural coordination in L2 English liquid allophony (Chapters 6 and 9):** How do L2 English speakers signal onset-coda allophony in English liquids?

In addition to these empirical studies, Chapter 4 outlines the data collection and analysis procedures using ultrasound tongue imaging, an articulatory method I use to answer the research questions mentioned above.

Chapter 3

General Methodology

In this section, I outline the general methodology that I follow to collect the main dataset. Each experimental chapter uses a subset of this dataset to serve respective aims and purposes. One of the two pilot studies (Chapter 5) is based on a different data set, but this will not be presented in this chapter.

3.1 Overview

In the sections below, I first describe participants recruitment, ethics consideration and the experimental stimuli. I then discuss the data collection procedure involving (1) tongue movement data using ultrasound tongue imaging, (2) audio recording synchronised with the ultrasound data, (3) participants' perceptual accuracy, (4) a demographic survey, and (5) a questionnaire on the degree of participants' familiarity with the lexical items used in the experiment. Note that further details on ultrasound tongue imaging, a vocal tract imaging method used in this study, have been included separately in Chapter 4.

3.2 Participants and Ethics

The data collection took place between October and December 2022. A total of 55 participants took part in the recording. This consists of 41 L1 Japanese speakers and 14 L1 English speakers. The speakers were recruited by advertising the research project through the professors in Japan, by posting an advert on my social media page, and from a network of my friends. The participation was entirely on a voluntary basis, and it was made very clear at the recruitment phase that any decisions regarding participation in this research project would not result in any disadvantages in academic record or course credits, especially in the case for L1 Japanese-speaking participants who were all undergraduate students at the time of recording. Further details of the participants are included in Appendices [A](#) - [C](#).

Prior to participant recruitment and data collection, ethics approval was obtained in January 2022 from Lancaster University, which was effective for data collection sessions taking place in the UK. Following this, additional ethics approvals were obtained in July and in September 2022 from each of the two institutions in Japan where participant recruitment and recording took place. Written consent was obtained from all the participants. Copies of the information sheets and consent forms can be found in Appendices [D](#) and [E](#).

3.2.1 L1 Japanese speakers

The L1 Japanese-speaking participants were recruited on the basis of the following criteria:

1. Being an L1 speaker of Japanese
2. Aged 18 or above
3. Having completed the English curriculum from primary to high schools in Japan
4. Having stayed in an English-speaking country for less than a month

5. Agreeing to provide a score from an English proficiency test
6. No history of language, speech or hearing difficulties

In the course of participant recruitment, however, some criteria had to be relaxed to obtain a reasonable number of participants. First, I had to extend participant recruitment to those who had experience in staying in an English-speaking country for up to half a year as students were encouraged to participate in an overseas language training programme at both institutions where data collection took place. Also, it turned out that some students, especially those in the first year of their university study, were not able to provide any English proficiency score (Criterion 5) as they had never taken any English proficiency exams. Furthermore, the two institutions encourage students to take different English proficiency tests, meaning that it was impossible to obtain a common, single scale that could reliably evaluate all L1 Japanese-speaking participants' L2 English proficiency.

This recruitment procedure resulted in participation of 41 L1 Japanese speakers (25 female and 16 male). They were all undergraduate students enrolled at a university in the central or western part of Japan, with the mean age being 19.85 years ($SD = 1.04$). The majority of them came from where the universities are located: Hyogo ($n = 12$) and Aichi ($n = 13$), but others came from regions nearby as well as regions that were far from the universities. The mean overseas experience is 0.82 months (1 week = 0.25 months), equivalent to slightly over three weeks, with the standard deviation of 1.38. All of them completed the school curriculum of English up until high school until the age of 18, with a mean length of English study being 9.55 years ($SD = 2.33$). Because of the recruitment considerations mentioned above, nine participants had experience in staying in an English-speaking country for 1.5 months ($n = 4$), 4 months ($n = 3$), 4.25 months ($n = 1$), and 5 months ($n = 1$). Nevertheless, none of the participants had an overseas experience longer than approximately four months. The majority of them spoke only English as their foreign language, although some had experience learning Chinese ($n = 12$), Korean ($n = 3$), French ($n = 1$) and Swedish ($n = 1$). Finally, eleven participants had never

taken any university modules related to linguistics or phonetics, whereas 24 of them had. Three participants majored linguistics and three worked on their graduation thesis related to linguistics or phonetics at the time of recording.

In addition to the demographic survey, they were asked to indicate on the scale of one to seven; (1) how they would assess their English ability (1: ‘I do not speak English at all.’ ~ 7: ‘No problem in using English in daily life.’), (2) how much they were accustomed to using English (1: ‘I am not accustomed to it at all.’ ~ 7: ‘I’m fully accustomed to it.’), (3) how much they use English in general per week (1: ‘I do not use English at all.’ ~ 7: ‘I only use English every day.’), and (4) how much they use English to talk with other people per week (1: I do not speak English at all.’ ~ 7: ‘I only speak English with people.’). Through these scales, I aimed to index participants’ self-perceived proficiency and the amount of English use, as commonly asked in previous research (P. Li et al., 2006), in order to make sure that participants were homogeneous within each speaker group, especially given the absence of a common proficiency scale available for L1 Japanese-speaking participants.

Based on the rating (1), the L1 Japanese-speaking participants evaluated their English ability as intermediate ($M = 3.88$, $SD = 1.03$). They were more or less familiar with using English, with the rating (2) being 4.07 on average ($SD = 1.08$). They indicated that they use English often given the rating (3) being 3.83 ($SD = 1.07$), which could possibly include English classes at the university. The amount of speaking English, however, is slightly less given the rating (4) ($M = 2.76$, $SD = 1.58$). With all these considerations, the current population of L1 Japanese speakers could be considered to reflect typical Japanese learners of English as a foreign language who do not have an extensive experience of staying in an English-speaking country or who do not use English regularly in their daily life in Japan.

3.2.2 L1 English speakers

The L1 English-speaking participants were recruited in Lancaster and in London in the UK on the basis of the following criteria:

1. Being an L1 North American English speaker
2. Born and raised in the US or Canada
3. Aged 18 or above
4. No history of language, speech or hearing difficulties

The recruitment process resulted in a total of 14 participants (11 female and three male) aged 28.93 years on average ($SD = 6.08$). Nine of them were born and raised the US and five in Canada (including one who was born in Poland but raised in Canada). All of them grew up using English at least up until 13 years of age. In addition, two participants from Canada used Polish and French alongside English until 13 years of age respectively. One participant from the US had a parent from Taiwan, so she was raised both in the US and Taiwan. Six of them were postgraduate students whereas the rest worked in the UK at the time of the recording. They had various overseas experiences, with the residency in the UK ranging from 2 months to 10 years.

I recruited L1 North American English speakers (Criterion 1) in consideration of English language teaching in Japan, in which North American English is often used as the pedagogical model (Setter & Jenkins, 2005) and thus is an appropriate variety of English to be compared against L1 Japanese speakers' production. In addition, rhotic varieties need to be included in this research because I aimed to compare pre- and post-vocalic rhotics, as presented in Chapter 9.

The questionnaire ratings suggest that all L1 English speakers recruited in this study identify themselves as fluent L1 North American English speakers. They all rated seven for the rating for (1) the participant's self-evaluation of their English ability and (2) the degree of familiarity with the English use. They use English

almost always in which 11 participants gave a rating of seven for the rating (3), except for one who gave four and two who gave six ($M = 6.64$, $SD = 0.84$). Similarly, nine participants rated seven for the rating (4) concerning the amount of speaking English per week, whereas the rest gave a rating in the range between four to six ($M = 6.43$, $SD = 0.94$). It is likely that some of my participants use non-English languages to interact with their family members whose L1 may not be English. Some of the participants indicated that they spoke a range of languages other than English, including French, Portuguese, Spanish, Turkish, Arabic, Italian, German, Polish, Chinese, and Cantonese. Finally, seven participants had no experience studying linguistics or phonetics, whereas two have experience in taking a linguistics class. Two of them had majored in linguistics or phonetics before, and three of them had worked on a research project (e.g., an MA thesis) related to linguistics or phonetics.

3.3 Stimuli

3.3.1 Production experiment

Production data consists of ultrasound tongue images and audio recordings of the participants' word-list reading. The participants read 26 words that contain English liquids word-initially ($n = 22$), word-medially ($n = 2$), and word-finally ($n = 6$). The target words have been selected based on the previous research (Proctor et al., 2019). It is important that the liquids are surrounded by vowels whose realisations are maximally similar across the two speaker groups; since L1 Japanese - L2 English learners may produce vowels differently from L1 English speakers, only vowels that would be the least variable even in the Japanese learners' production were selected, resulting in the following three vowel conditions: /i:/, /æ/ or /u:/ (Makino, 2009; Takebayashi & Saito, 1998; Vance, 2008). The target words are shown in Table 3.1 and the participants repeat each of them in isolation up to five times. In addition to these target words, six filler words have been included with each vowel: *ham*, *heap*, *hoop*, *bam*, *beep*, *boom*, in which flanking consonants are chosen to avoid

Table 3.1: English target words for the production task

Vowel	Initial		Final		Medial	
	/l/	/ɹ/	/l/	/ɹ/	/l/	/ɹ/
/i:/	leap	reap	peel	peer		
	leaf	reef	feel	fear		
	leave	reeve	veal	veer	believe	bereave
/æ/	lap	rap				
	lamb	ram				
	lamp	ramp				
/u:/	lube	rube				
	loom	room				

coarticulatory effects on the vowel.

Adapting the word-list elicitation would increase the consistency among production compared to the carrier phrase as stress placement and vowel realisation patterns would be highly variable in L2 English speech, the observation arises from the pilot study in Chapter 5 where a carrier phrase was used.

In addition to the English words, 11 Japanese words be recorded so that we could compare them with the English /l/ and /ɹ/ given the typical substitution patterns. Similarly with the English tokens, the three vowel environments are selected: /i/, /a/ and /u/ based on the reported substitution patterns of the English vowels in the Japanese EFL learners' speech, making it easier for us to minimize the effect of the following vowels (Makino, 2009; Takebayashi & Saito, 1998; Vance, 2008).

As seen in Table 3.2 Japanese target words in the production task include two groups of words. The first is the lexical words that contain a word-initial liquid. This is to maintain the comparability between English and Japanese in that lexical items are used in both language conditions. On the other hand, previous research on Japanese liquids has commonly used mimetic or nonsense monosyllable

Table 3.2: Japanese target words in the production task

Vowel	Lexical words	Mimetics
/i/	リーフ /ri:hu/	ビリビリ /biribiri/ ピリピリ /piripiri/
/a/	ラフ /rahu/ ラム /ramu/	バラバラ /barabara/ パラパラ /parapara/
/u/	ループ /ru:pu/ ルーム /ru:mu/	ブルブル /buruburu/ プルプル /purupuru/

phrases in eliciting the intervocalic liquids (Katz et al., 2018; Kawahara & Matsui, 2017; Morimoto, 2020; Sudo et al., 1983; Yamane et al., 2015). The inclusion of the mimetic words, therefore, would also increase the comparability of the current results with the existing Japanese articulatory studies. The mimetic word-list was developed based on a previous study (Morimoto, 2020).

In choosing the lexical and mimetic items, the other consonant in each word is kept consistently to the bilabials /p b m ϕ / in order to minimise the intrusive effects of the consonants on the articulation of liquids. As such, the lexical words always have the liquid-vowel-bilabial sequence word initially and only /b/ or /p/ are flanked with the vowels in the mimetics. Similarly to the English word-list, six filler items have also been included: タフ /tahu/, チーフ /ti:hu/, バタバタ /batabata/, パタパタ /patapata/, プツプツ /putuputu/, and タム /tamu/. These words were chosen to (1) minimise the coarticulatory effects on the vowel and (2) allow for future investigation of the articulation of Japanese liquids against obstruents matching in the place of articulation, similar to the analysis design in Proctor (2011).

3.3.2 Perception experiment

Perceptual accuracy is commonly measured through identification and discrimination tasks (Colantoni et al., 2015; Strange & Shafer, 2008). In the identification task, participants are aurally presented with one token at a time and asked to select one of the two orthographic representations that match the token they hear (Bradlow et al., 1999; Bradlow et al., 1997; Ingvalson et al., 2012; Lively et al., 1993; Logan et al., 1991). In the discrimination task, participants hear multiple tokens serially and choose one that falls into a different category from the others (Ingvalson et al., 2012; Shinohara & Iverson, 2018).

In this research, the identification task has been used to measure the participant's perceptual accuracy. It measures how accurately learners perceive the individual sounds in the target language in comparison to corresponding to L1 categories (Colantoni et al., 2015), and it has been commonly used in studies which investigated the relationships between L2 speech perception and production (Bradlow et al., 1999; Bradlow et al., 1997; Saito & van Poeteren, 2018). L1 Japanese speakers' production accuracy has also been shown to correlate more with their identification accuracy than with their discrimination accuracy, or they at least have similar effects on production (Hattori & Iverson, 2011; Shinohara & Iverson, 2018; Yamada & Tohkura, 1992). Based on these findings, it was decided that identification tasks should be appropriate to index participants' perceptual accuracy of word-initial English liquids for the purpose of the current research.

The stimuli for the perception experiment were developed based on the word-list and recording in Brekelmans et al. (2022), available via the Open Science Framework (OSF) repository. A total of 24 minimal pairs (i.e., 48 words), produced by seven L1 Canadian English speakers, were included in this experiment. The use of multiple talkers would discourage the listeners from relying on speaker-specific, idiosyncratic cues and encourage them to identify the stimuli phonologically rather than phonetically.

The 24 minimal pairs that had a word-initial English /l/ and /ɹ/ contrast (see

Table 3.3) were presented to the participants based on the previous findings that word-initial singleton liquids are difficult for L1 Japanese speakers to accurately identify compared to the word-final singleton liquids (Bradlow et al., 1997; Gordon et al., 2001; Logan et al., 1991; Saito, 2011; Sheldon & Strange, 1982; Strange & Dittmann, 1984). At the same time, multiple vowel environments were included as liquids are identified by L1 Japanese listeners less accurately when preceding back and low vowels (Shimizu & Dantsuji, 1983).

Another consideration in constructing the stimuli is lexical frequency. L2 learners' perceptual accuracy is likely to be influenced by the lexical frequency and participants' subjective familiarity with the lexical items (Flege et al., 1996). Since the L1 Japanese-speaking participants' lexical knowledge in this study would be much smaller than that of L1 English speakers, it would be optimal to use the word frequency measure that better reflects the occurrence of the lexical items in the context of English Language Teaching in Japan.

For this reason, the lexical items were screened according to The New JACET List of 8000 Basic Words, a lexical frequency list developed by the Japan Association of College English Teachers (Daigaku Eigo Kyoiku Gakkai Kihongo Kaitei Tokubetsu linkai, 2016). The list selects the most frequent 8,000 words with reference to not only the existing corpora, British National Corpus and Corpus of Contemporary American English, but also the school and university entrance exams in Japan. In addition, another corpus, AmeE06 (Potts & Baker, 2012), was referred to in order to complement the JACET list whose coverage is relatively small. AmeE06 is a relatively up-to-date corpus, containing 1,017,879 tokens from 500 sources from texts published between 2004 and 2007, and the coverage has been limited to American English, produced by authors who had been born or continuously lived for the majority of their lives in the United States (Potts & Baker, 2012).

With these two corpora, two groups of the perceptual stimuli were created: familiar and unfamiliar sets. The L1 Japanese-speaking participants in this study should be able to recognise the words in the familiar group because they are

Table 3.3: Stimuli for perception task.

Vowel	Front		Back	
	Familiar	Unfamiliar	Familiar	Unfamiliar
High	rim/limb	<i>reek/leak</i>	room/loom	rude/ <i>lewd</i>
	read/lead	reach/ <i>leech</i>	root/loot	<i>ruse/lose</i>
Mid	raid/laid	<i>rake/lake</i>	road/load	robe/ <i>lobe</i>
	red/led	raise/ <i>laze</i>		
Low	right/light	<i>rife/life</i>		
	rack/lack	rice/ <i>lice</i>	rock/lock	<i>rot/lot</i>
	rag/lag	<i>rash/lash</i>	wrong/long	rob/ <i>lob</i>

Notes: Familiar = both minimal pairs should be familiar with the L1 Japanese participants. Unfamiliar = one of the minimal pair words may be unfamiliar to the L1 Japanese-speaking participants. Italicised words fall outside the JACET8000 list, deemed to be unfamiliar.

mostly covered in the JACET list. The other group, the unfamiliar group, includes minimal pairs in which at least one of the members in a minimal pair is outside the JACET list (the italic words in Table 3). Inclusion of the unfamiliar group would better differentiate the Japanese participants because experienced Japanese English learners would be able to recognise unfamiliar words better than inexperienced learners (Flege et al., 1996). Further details of the lexical frequency analysis is included in Appendix H

3.4 Procedure

The recording procedure was standardised across the participant populations as much as possible. Recording took place in a quiet classroom in the universities in Japan for L1 Japanese speakers and in a sound-attenuated booth in the universities

Steps	Details	Lang. of instruction
Briefing / Set-up	<ul style="list-style-type: none"> - Participant's arrival - Information sheet - Consent form - The ultrasound headset set-up 	Japanese
Production Task (1)	<ul style="list-style-type: none"> - Palate and bite plane trace - Japanese wordlist reading (5x) 	Japanese
Production Task (2)	<ul style="list-style-type: none"> - A brief conversation in English - Vocabulary check - English wordlist reading (5x) - Disassemble the ultrasound headset 	English
Perception Task	<ul style="list-style-type: none"> - Identification task 	English
Debriefing		

Figure 3.1: Schematised experimental protocol

in the UK for L1 North American English speakers. In the recordings of some of the L1 Japanese-speaking participants, however, there was a minor continuous noise from a fan because ventilation was mandated at the time of recording for a Covid-19 measure. A summary of the overall procedure is shown in Figure 3.1.

Upon arrival to the recording venue, participants were given the information sheet in order to familiarise themselves with the experiment. L1 Japanese-speaking participants were given two sets of information sheets, one in English and one in Japanese, to fulfill the ethics requirements at both Lancaster University and at the institutions where the data collection took place. The content of the information sheets was kept identical as much as possible, and they had a chance to ask any questions in Japanese. L1 English-speaking participants were given the information sheet in English only. Once they understood the experiment and agreed to take part, they signed two sheets of the consent forms, one to be retained by the researcher and the other by the participant. This also involved two sets of consent forms (four in total) for L1 Japanese-speaking participants (i.e., two in Japanese and two in English) whereas L1 English-speaking participants signed only on the English versions. At this stage, they completed the demographic questionnaire

and signed the receipt of the honorarium that they received as a compensation of their time and participation. They were paid either ¥2,000 or £15 in cash or vouchers, commensurate with the regulations at each institution. Copies of the information sheets, the consent forms and the demographic questionnaire can be found in Appendix [D](#), [E](#) and [F](#).

For the recording, midsagittal tongue images were recorded using the Telemed MicrUS system with a 2–4 MHz 20mm radius convex probe, connected via a Sound Devices USB-Pre2 audio interface to a laptop computer operating the Articulate Assistant Advanced software version 220.4.1 (Articulate Instruments, [2022](#)). Audio recording was made simultaneously using an Opus 55 MK ii condenser microphone attached to the ultrasound headset, pre-amplified and digitised at 44.1 kHz with 16-bit quantisation.

Once all the necessary paperwork was completed, I fitted the ultrasound headset to stabilise the probe. In this experiment, I used the UltraFit headset (Spreafico et al., [2018](#)) that was made out of light-weight plastic. Once the headset was fitted, I recorded the participant's bite plane by asking them to bite a thin plastic plate and push their tongue up against the plate, so that the flat surface and the tongue deformation can be recorded, which is used to align everyone's midsagittal tongue shape onto a common coordinate system (Scobbie et al., [2011](#)). At this stage, the palate shape was recorded while participants were swallowing water. As the details of the headset fitting and the bite plane rotation is described in Chapter 4, this will not be elaborated further here.

In addition to the tongue images, I recorded the side-profile view of the participants' lips, which would provide useful information as to the labial gesture especially for English /ɪ/ (e.g., King & Ferragne, [2020](#)). The lip movement was recorded with a small camera attached to the extension arm attached to the headset, and the camera angle was adjusted so that the lips were imaged parallel to the x -axis of the lip video wherever possible. Efforts were also made to ensure that background colour did not interfere with the lip camera images by e.g., choosing a plain white



Figure 3.2: An experimental session with an L1 Japanese-speaking participant.

wall. The lip data are not analysed for the purpose of this PhD thesis but the lip movement will be addressed in future research.

For the recording, participants sat in front of a laptop computer which displayed the recording stimuli on the Articulate Assistant Advanced software version 220.4.1 (Articulate Instruments, 2022). Prior to the recording, participants were given a list of all the stimuli words to ensure that they could read them. This is especially important for L1 Japanese speakers who were learners of English as a foreign language, in which it was expected that some of the words may be unfamiliar for them. While the researcher did not instruct them to pronounce words in certain ways, some of the L1 Japanese-speaking participants asked the researcher how to read words, in which case I demonstrated the pronunciation based on the North American English pronunciation. An example image illustrating the experiment set-up is shown in Figure 3.2.

During the recording, each target word was presented individually one by one automatically using the ‘continuous’ presentation mode on the AAA software. This

allows the researcher to make notes on the things he noticed during the experiment, such as mispronunciation, the status of probe placement and presence of noise, which would be useful when screening the recorded tokens for further analysis. Presentation of target words was delayed by 1,000 ms relative to the onset of ultrasound recording in order to allow a synchronisation pulse (generated by the BrightSyncUp system) to be recorded for lip-tongue-audio synchronisation and to prevent participants from making preparatory posture for the initial consonant. The latter consideration is especially important in the articulatory timing analysis that is presented in Chapter 9. This prevention measure is deemed to be successful as the articulatory analyses suggest participants took a ‘pre-speech’ posture before the onset of articulatory movement associated with the first segment (cf. Palo et al., 2014; Wilson & Kanada, 2014).

When recording bilingual speakers, previous research has emphasised the importance of controlling the language mode (Escudero, 2005; Grosjean, 2008; Yazawa et al., 2020). Theoretically, it would have been desirable to have separate experimenters with different first language backgrounds and to conduct the recording of different languages (Grosjean, 2008), but such a practice was less realistic in the context of the current research. Nevertheless, I controlled the L1 Japanese-speaking participants’ language modes by (1) separating recording sessions between English and Japanese and (2) conducting a brief conversation in English to facilitate their transition from the Japanese mode into the English mode.

As shown in Figure 3.1, recording of the Japanese and English words was conducted in separate sessions for L1 Japanese-speaking participants. They recorded Japanese words first, proceeding directly from the initial briefing session conducted also in Japanese, because this would make the overall experiment proceed more quickly than otherwise, as this would allow them to fully understand the experiment, ask questions, and signal discomfort in Japanese when fitting the headset. Up until the end of the Japanese word recording, I gave instruction entirely in Japanese. I then switched the language of instruction into English and asked the participants

five simple questions in English. The questions included (1) What's your name?, (2) Where do you live?, (3) What do you study?, (4) What do you like about the university?, and (5) Which country would you like to visit? These questions were decided so that even participants who were less proficient in L2 English could respond and that it would not take up too much time. The questions were presented both visually using paper where each question was printed and aurally by the researcher saying each question aloud. The majority of L1 Japanese-speaking participants majored in foreign languages at the time of recording and they seemed to enjoy engaging in this conversation. Finally, once completed, they recorded English words while the researcher still gave instruction to them in English.

Recording L1 English speakers did not require the consideration of the language mode given that it was conducted in an English-speaking environment in the UK. All instruction was, therefore, given entirely in English, and they only recorded English words. Overall, the recording for L1 Japanese speakers, involving both English and Japanese recording, took approximately between 45 and 60 minutes whereas up to 30 to 45 minutes for L1 English speakers, depending largely on the time taken for headset fitting.

Once the speech production task completed, I took the ultrasound headset off the participants and then invited them to the perceptual experiment. The perception experiment was conducted to obtain an index of English proficiency on the common scale across participants. The experiment was prepared and conducted using Gorilla (Anwyl-Irvine et al., 2019) on a Macbook Pro laptop computer, with audio prompts presented through the Audio-technica ATH-M20x headphone. As described earlier, this experiment involved an identification task, in which participants were asked to listen to an English word presented only once, and move the cursor using the laptop track pad to click one of the two buttons on the computer screen to choose what consonant appeared at the beginning of each word, with each button showing 'L' or 'R'. The participants began with a practice session in order for them to get familiarised with the procedure, in which they listened to nine words that did

not contain word-initial English liquids. They were given feedback in this practice session. After this, they proceeded to the main experimental session, where no feedback was given. The main task phase consisted of four blocks, in each of which there were 24 tokens. The order of presentation was randomised within each block across participants. The perception experiment took approximately 10 minutes for L1 Japanese speakers and five minutes for L1 English speakers.

Once both the production and perception experiments completed, the participants indicated the degree of familiarity with the lexical items used throughout the experiment in the scale of four: ‘Yes, I know it.’, ‘I think I know it.’, ‘I might know it.’, and ‘I don’t know it.’. The scale was adopted from a previous study (Thomson & Isaacs, 2009). Overall, the entire experimental session lasted up to approximately 90 minutes for L1 Japanese speakers and up to approximately 60 minutes for L1 English speakers.

Chapter 4

Ultrasound tongue imaging

This chapter outlines ultrasound tongue imaging, one of the vocal tract imaging techniques, that has been employed extensively in this PhD project. This chapter presents a manuscript that has been accepted for publication for a special issue in the *Journal of Phonetic Society of Japan*, in which I detail some of the methodological considerations for ultrasound data collection and analysis based on my own practices from the data collection sessions for this PhD research and the standard practice at the Lancaster University Phonetics Lab. It also introduces the key methodological considerations including the headset fitting and bite-plane rotation (cf. Scobbie et al., 2011), as well as a quantitative analysis workflow using the Principal Component Analysis (PCA).

Quantifying between-speaker variation in ultrasound tongue imaging data

XXXXXX XXXXX

超音波舌撮像データにおける話者間特性の定量化手法

SUMMARY: This article outlines a quantitative, between-group comparison of tongue shapes using *ultrasound tongue imaging*, one of the vocal tract imaging techniques widely used in articulatory phonetics research. This article first provides a brief overview of ultrasound tongue imaging, followed by a description of a cross-speaker normalisation method based on bite plane rotation. I then outline ultrasound data recording and analysis workflow with a case study illustrating data analysis using Principal Component Analysis (PCA). I demonstrate that the bite plane rotation, coupled with statistical normalisation methods, allows for a reliable tongue-shape comparison by establishing a common coordinate system across speakers.

Key words: articulatory phonetics, ultrasound tongue imaging, Articulate Assistant Advanced, bite-plane rotation, Principal Component Analysis (PCA)

1. Introduction

This article describes *ultrasound tongue imaging*, one of the vocal tract imaging techniques that is widely used in contemporary articulatory phonetics research (Kochetov, 2020). Ultrasound tongue imaging is a non-invasive, cost- and time-effective tool that allows for direct access to articulation. In this article, I aim to complement the existing ultrasound tongue imaging tutorials that provide general overviews (e.g., Gick et al., 2008; Wilson, 2014) by providing concrete data acquisition and analysis methods and illustrating them through a case study.

The focus of the current article is to demonstrate a quantitative, population-level comparison of multiple speakers' tongue shapes using ultrasound tongue imaging. In the section below, I start by providing the research context and how articulatory data could complement the existing findings. I then provide a discussion of issues and solutions to tongue shape comparison using ultrasound. I then explain a typical workflow of an ultrasound experiment based on the standard practice in our lab, focussing on the experiment preparation and data analysis. Finally, a case study illustrates a quantitative analysis of tongue shape using Principal Component Analysis (PCA).

The foundation of the quantitative analysis presented here is the combination of (1) bite plane rotation and (2) statistical within-speaker normalisation, allowing researchers to align multiple speakers' tongue shape onto a common coordinate system. PCA is a data dimensionality reduction technique that can be useful to capture articulatory dimensions that are salient in the data.

This article assumes ultrasound research using the Articulate Assistant Advanced (AAA) software (version 220.5.1; Articulate Instruments, 2022), which is one of the most widely used software for ultrasound research. Data processing, analysis and visualisation are done via R version 4.3.2 (R Core Team, 2023). Data and codes used in this article, as well as a list of useful resources, are publicly available in the online supplementary materials at <https://shorturl.at/hHJKS> (currently anonymised for peer review).

1.1 Research context

It is widely well-known that L1 Japanese speakers have difficulty producing English /l/ and /ɹ/ accurately. Previous studies show that L1 Japanese speakers' production of English /l/ and /ɹ/ is influenced by or even substituted with the Japanese liquid category /r/, canonically realised as alveolar tap or flap [ɾ] (e.g.,

* [45_Footnote] The affiliation(s) of author(s) (English and Japanese) (Final manuscript)

Riney et al., 2000). This may be because L1 Japanese speakers are less sensitive to the third formant (F3) frequencies, the important acoustic cue in the English liquid contrast (Iverson et al., 2003). Whereas English /ɹ/, in particular, is characterised by notably low F3 frequencies that provide a reliable acoustic cue to contrast with English /l/, L1 Japanese speakers instead rely on F2 to make a distinction between English /l/ and /ɹ/ in both perception (e.g., Iverson et al., 2003) and production (e.g., Saito & van Poeteren, 2018). The reliance on F2 in production suggests that L1 Japanese speakers redeploy articulatory strategies for Japanese /ɹ/ (e.g., front-back dimension) to produce English liquids (Saito & van Poeteren, 2018).

Despite a rich amount of previous research, it remains unclear *how* exactly L1 Japanese speakers are different from L1 English speakers in articulation. While acoustic events are the immediate product of adjustments in vocal tract configurations (Iskarous & Pouplier, 2022), articulation of English liquids cannot be easily inferred based solely on the acoustic signals. The two canonical configurations for English /ɹ/, ‘retroflex’ and ‘bunched’ tongue shapes, for instance, are not readily distinguishable by lower formants (Zhou et al., 2008). It may be the case that English speakers use speaker-specific articulatory strategies to achieve a common acoustic output of low F3 that characterises English /ɹ/, suggesting that it can be challenging to infer the exact articulatory properties for English /ɹ/ from the acoustic signals alone (Mielke et al., 2016).

Uncovering how L1 Japanese speakers differ from L1 English speakers in the articulation of English /l/ and /ɹ/ is of both theoretical and pedagogical interest, given the persistent nature of difficulty for L1 Japanese speakers in producing English /l/ and /ɹ/. I frame this article around the comparisons of tongue shapes for English /l/ and /ɹ/ between L1 Japanese and L1 English speakers. In the following section, I briefly outline the overview of ultrasound tongue imaging techniques, before explaining the methods in detail.

2. Ultrasound tongue imaging

2.1 Ultrasound research methods

Ultrasound tongue imaging provides a near-holistic midsagittal image of the tongue with high spatiotemporal resolutions. Typically, an ultrasound probe (or transducer) is placed underneath a participant’s chin, which contains a piezoelectric crystal emitting a high-frequency ultrasound (Stone, 2005). The ultrasound travels through the tongue and

reflects back once it reaches the air just above the tongue surface due to the change in density between the tongue and the air, which then is received by the probe (Stone, 2005). This results in a thick white curve just above the tongue surface as shown in Figure 1, enabling us to infer shapes, positions, and movements of the tongue. Note that ultrasound does not travel through hard structures like bones, including the mandible and the hyoid bone, and they appear as a black/dark shadow in the ultrasound tongue images (see the black shadow towards the right edge of the fan in Figure 1).

Ultrasound achieves quite a high sampling rate, which is suitable for capturing quick movements of the tongue. A faster frame rate is necessary to capture fast lingual movements in segments such as alveolar taps or flaps [ɾ]; although machines in previous research achieve approximately up to 30 fps (Derrick & Gick, 2011; Yamane et al., 2015), it is possible with more recent machines to achieve a much higher rate of ca. 80 fps or even greater (Kirkham et al., 2023; Kochetov, 2020; Nagamine, 2023). A recent co-registration study demonstrates a high degree of accuracy in the spatial and temporal alignment between ultrasound and electromagnetic articulography (EMA; Kirkham et al., 2023).

Furthermore, ultrasound imposes little discomfort for the speaker compared to other existing methods such as EMA and real-time magnetic resonance imaging (MRI), allowing for experiments with a wider range of participant populations. It is possible to obtain a clear image of the tongue by just holding the probe by hand and placing it underneath the participant’s chin without any stabilisation measure, which is useful in the field of speech therapy when restraints need to be minimised as much as possible or probe stabilisation is not practically feasible (Klein et al., 2013; Preston et al., 2017). Ease in visualising the tongue is also an advantage in pronunciation teaching classrooms, in which time and resources are often limited. An increased number of studies report the benefit of ultrasound visual feedback in L2 pronunciation teaching and learning (e.g., Antolík et al., 2019; Bryfonski, 2023).

Finally, the increased portability of the recent ultrasound machines enables us to conduct articulatory research in the field with a participant population that is not easily accessible through laboratory-based experiments. The ultrasound machine can be only slightly larger than a smartphone, such as the Telemed MicrUS system that operates with the Articulate Assistant Advanced software. A recent study

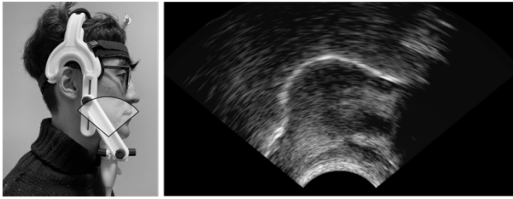


Figure 1 Left: A side-profile view of a participant wearing an ultrasound stabilisation headset with the ultrasound probe placed underneath the chin. The overlaid fan schematises the area scanned by ultrasound. Right: A midsagittal tongue image recorded with the Articulate Assistant Advanced software. The speaker is producing a tongue-tip-up retroflex /ɻ/. Tongue tip points to the right and the tongue posterior to the left. The mandible shadow is imaged as the dark black area towards the right in the fan, obscuring the tongue tip image.

demonstrates its utility in field work for ultrasound data collection as part of a social sciences festival at a local market in a small town in the UK (Nance et al., 2023).

To summarise, ultrasound is a non-invasive, relatively easy approach to studying lingual articulation compared to other existing methods. The high spatiotemporal resolution and increased portability are particularly suitable for a wide range of recording settings outside laboratories, including language classrooms and in the field. Despite the ease of data acquisition, however, data processing and analysis usually require an extensive amount of time, effort and consideration, as discussed below.

2.2 Probe stabilisation

The rich dimensionality of midsagittal tongue surface data obtained from ultrasound, as opposed to flesh-tracking methods such as EMA, introduces greater demands at the data processing phase. While a holistic midsagittal view of the tongue surface can visualise both global and local lingual movements, it is often quite challenging to partition different parts of the tongue and identify linguistically meaningful movements based solely on ultrasound images (Davidson, 2006; Stone, 2005). In addition, a lack of fixed anatomical structure in ultrasound images makes it challenging to infer the exact position of the tongue in the vocal tract (Gick et al., 2008; Stone, 2005). Individuals differ substantially from one to another in their vocal tract anatomy, including the length and width of the tongue as well as articulatory strategies to produce certain sounds, including English /ɻ/ (Slud et al., 2002). These overall suggest that it is not appropriate to directly compare tongue shapes obtained

from different subjects, making it difficult to perform a simple cross-speaker comparison.

Alternatively, a fruitful approach to articulatory investigation using ultrasound is to compare *within-speaker* articulatory strategies across multiple subjects, given the remarkably high degree of within-speaker consistency found in speech production (Johnson et al., 1993). As a consequence, previous ultrasound research attempts to quantify measures that capture within-speaker variability in midsagittal tongue movement. Strycharczuk and Scobbie (2017), for example, quantify the degree of GOOSE-fronting in British English by measuring the tongue position for /u:/ relative to the reference vowel /i:/ in each speaker. Similarly, Kirkham and Nance (2017) quantified the degree of tongue root advancement in each subject to investigate how Twi-English bilingual speakers realise the [ATR] and [TENSE] features in their vowel production. These derived measurements can be statistically normalised (e.g., using *z*-scores) within each speaker, facilitating cross-speaker comparison in subsequent statistical analyses.

In addition to the considerations above, reliable quantitative analysis of tongue contours using ultrasound is based on at least two implicit assumptions. First, at the data acquisition stage, it is crucial to ensure that the probe is stabilised so that it does not move substantially within one recording session (Gick et al., 2008). While tongue images can be easily obtained by a hand-held probe underneath the speaker's mandible, this inevitably imposes additional noise in data due to the movement of the probe itself, making it impossible for researchers to tease it apart from tongue movement (Derrick et al., 2018; Stone, 2005). Quantitative analysis attempts to minimise measurement errors by obtaining multiple observations and employing statistical tests, and it is therefore important to ensure that multiple observations from multiple speakers recorded with ultrasound can be reliably compared.

One way of minimising measurement errors in ultrasound recording is to stabilise the probe relative to the head wherever possible. Various stabilisation techniques have been proposed and used, including a stand on the table (Stone, 2005), a headrest on which the participant could rest their head (Derrick et al., 2018), and a wearable headset (Spreafico et al., 2018). The Haskins Optically Corrected Ultrasound System (HOCUS) combines ultrasound with optical tracking to correct the probe movement relative to the participant's head movement (Whalen et al., 2005). Recent stabilisation headsets are made of light plastic instead

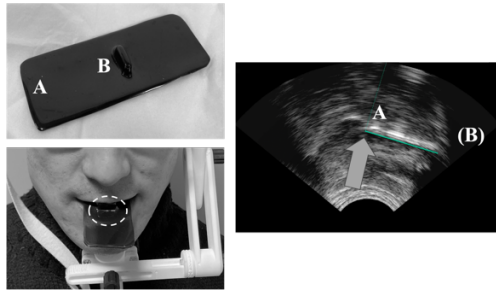


Figure 2 Top left: A bite plate made of biocompatible plastics, provided to Lancaster Phonetics Lab by courtesy of Dr Eleanor Lawson (University of Strathclyde). Bottom left: A speaker biting the bite plate by inserting the end ‘A’ in the mouth. The dashed circle indicates where the upper incisor makes contact against the barrier (i.e., ‘B’ on the bite plate). Right: Bite plane measurement superimposed on an ultrasound image. The arrow indicates the point at which the tongue shows deformation in shape (corresponding to ‘A’ in the top left image). Bite plane measurement traces the flat tongue surface observed anterior to the deformation point. Note that the label ‘B’ in the right image is in parentheses because it does not indicate the exact location of the incisors as it was not measured precisely here.

of more rigorous materials like metals (Articulate Instruments, 2008) and thus impose relatively small degrees of discomfort on the research participants who wear the headset while maintaining measurement accuracy (Pucher et al., 2020; Spreafico et al., 2018). While the need of probe stabilisation may cancel out the benefit of the ease in imaging tongue shape (see Section 2.1) and some headsets (e.g., Ultrafit) constrain the participant’s jaw movement to some degree, the discomfort that a plastic headset may impose on participants is fairly minimal considering that they are free to move their head. The example of a plastic headset is seen in the left picture in Figure 1.

2.3 Establishing a common coordinate system across speakers

Once probe stabilisation has been achieved, another consideration is to align tongue splines in a common coordinate system across speakers and repetitions relative to fixed, passive articulators (Scobbie et al., 2011, 2012). While this process is not necessary when the researcher’s interest lies solely in comparing tongue shapes irrespective of the tongue position, such as Dorsum Excursion Index (DEI; Zharkova, 2013), establishing a common coordinate system across

speakers’ tongue splines aids linguistically-meaningful interpretations of tongue movements regarding magnitude, orientation, location and relative timing (Scobbie et al., 2011; Westbury, 1994). The coordinate transform is incorporated in a common workflow in EMA, in which the locations of the lingual sensors can be normalised across speakers by the use of reference sensors attached to the speaker’s nasion, upper and lower incisors and both mastoids (Rebernik et al., 2021). Similarly, the holistic midsagittal view of the entire vocal tract captured by real-time MRI provides many possible structures that could be used as reference points, including the upper and lower incisors (e.g., Maekawa, 2023).

Ultrasound, on the other hand, does not usually provide relative positional information of the tongue in the vocal tract because it does not record as many anatomical landmarks that can serve as reference points as other methods such as EMA or MRI (Gick et al., 2008; Stone, 2005; Zharkova, 2013). In addition, the probe angle is usually determined in order to ensure clear visibility of the tongue surface, meaning that the horizontal and vertical axes from the ultrasound images bear no consistent linguistically-meaningful interpretations such as ‘frontness’ or ‘height’ of the tongue (Scobbie et al., 2012). Stone (2005) proposes the use of a dental cast by taking each speaker’s dental impression and using it as a reference to determine the probe position for multiple recording sessions for a single speaker. This method, however, relies heavily on the researcher’s qualitative assessment of the probe position, which may suffer from reduced precision in probe placement. Although it is also possible to rotate the tongue curves based on the horizontal axis defined by the tongue positions for the peripheral vowels, this may add extra difficulty in interpreting the resulting diagrams that may be inconsistent with findings based on EMA data (Scobbie et al., 2012).

A possible and practical solution to the lack of reference points in ultrasound is to establish a *post-hoc* common coordinate system to normalise tongue positions across speakers using the speaker’s bite plane, obtained through the use of a simple thin plastic plate (seen in the top left image in Figure 2). The plastic plate is 40 mm wide and approximately 60 mm long with a 2-mm thickness. The plate, made of biocompatible plastics by Dr Eleanor Lawson at University of Strathclyde, has a small ‘hump’ at approximately 45 mm from the end that barriers the upper front teeth, which maintains consistent across speakers the distance

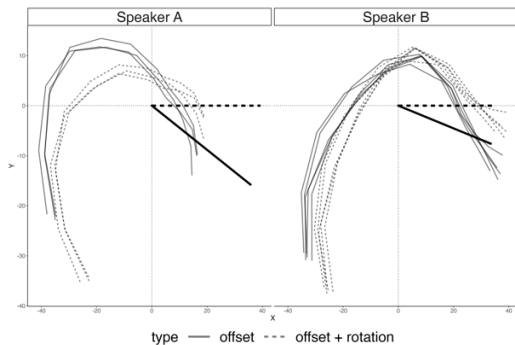


Figure 3 Example tongue splines after offsetting (solid line) and offsetting/rotating (dotted line). Thicker lines represent each speaker's bite plane.

between the upper front teeth and the end of the plate inside the mouth ('B' in the top left image in Figure 2). The speaker bites this plate by inserting the longer end of the plate into the mouth (i.e., 'A' in the top left image Figure 2) to the extent that the upper front teeth make contact with the barrier (illustrated in the bottom left image in Figure 2). When the participant pushes their tongue against the plate, the bite plane can then be imaged in ultrasound as a flat surface from the point where the tongue shape shows some deformation to the tongue anterior (see the image on the right in Figure 2).

The bite-plane normalisation involves 'offsetting' and 'rotating' the tongue splines against the bite plane once recording is completed and tongue splines are estimated/tracked. Bite plane offsetting and rotation is illustrated in Figure 3. Here, as an example, I compare the midsagittal tongue shapes of the English vowel /æ/ in the word *ham*, extracted at the vowel midpoint, produced multiple times by two speakers: one L1 Japanese speaker (Speaker A) and one L1 English speaker (Speaker B).

The tongue shapes and the bite planes in the solid line in Figure 3 show *offsetting* in which the origin of the coordinate system is aligned at zero (i.e., the point where the bite plane meets the tongue in the mouth, causing tongue deformation). At this point, the angle of the bite plane is still different between the two speakers, suggesting that it is not possible to perform reliable interpretations of tongue shape along conventional dimensions such as 'tongue height' or 'tongue retraction'. The dotted lines, on the other hand, show the tongue splines that are both *offset* and *rotated* against the bite plane, in which not only the coordinate origin is aligned but also the bite plane is rotated to be horizontal. This way, a common coordinate system can

be established for both speakers with shared x - and y -axes (Scobbie et al., 2011).

Note that AAA has various options for exporting the tongue splines, including exporting the x/y coordinates with/without offsetting and rotating as well as scaling the size of the tongue relative to the length of the measured bite plane being one. The origin of the tongue position offsetting/rotation also needs to be decided between *knot 0* and *knot 1*. *Knot 0* corresponds to the point where the dotted and solid lines meet in the L-shaped fiducial spline that can be defined on the AAA software (i.e., the right image in Figure 2). *Knot 1*, on the other hand, is the other end of the solid line. It is possible to define the point 'B' as the speaker's incisor location, although it was not measured precisely in Figure 2. A quick illustration of these exporting options is provided in the online supplementary materials.

Ultrasound offers researchers access to lingual articulation easily because of its non-invasiveness, high spatiotemporal resolution and portability. Processing midsagittal tongue images obtained with ultrasound, however, requires some methodological considerations. In addition to the need for head stabilisation during recording, it is suggested that a common coordinate system be established across speakers using bite plane measurement for each speaker as a way of cross-speaker normalisation. More information about the bite-plane rotation can be found in Scobbie et al. (2011) and Strycharczuk and Scobbie (2017). See also Scobbie et al. (2012) for more detailed discussions on the tongue curve rotation methods.

3. Example ultrasound recording and analysis workflow

As shown in the previous sections, a reliable cross-speaker comparison can be facilitated by incorporating probe stabilisation and bite-plane measurement in an ultrasound workflow. In this section, I provide further details on these procedures by showing an example workflow for ultrasound recording and providing detailed explanations for considerations that need to be made.

3.1 Data recording

This section mainly outlines the experiment preparation stage given its importance for a reliable data analysis. This includes headset fitting, probe position adjustment, recording parameter setting and bite plane measurement.

3.1.1 Fitting probe stabilisation headset

Once participants arrive at the recording venue and complete all necessary paperwork and briefing procedures, I fit the ultrasound headset on the participant's head in order to stabilise the ultrasound probe relative to head movement. I use UltraFit, a light-weight plastic ultrasound probe stabilisation headset (see Figure 1) that is commercially available from Articulate Instruments Ltd. (Spreafico et al., 2018). I make sure that the headset is fitted straight and tight enough for the probe to be stabilised but not too tight for participants to feel discomfort.

3.1.2 Adjusting probe position

Once the headset is successfully fitted, the position of the probe needs to be adjusted. I apply the ultrasound gel to the probe surface at this point so that the probe makes firm contact with the lower chin, as any pocket of air between the probe and the participant's chin can make the image quality poorer.

The probe position adjustment is carried out in the following order. First, I determine the probe position along the midsagittal plane by adjusting it from the participant's front. Second, looking at the probe from the side, I adjust the probe angle; in most cases, this is to make sure that the probe points straight up, which usually results in a good image quality provided a proper field-of-view setting (see Section 3.1.3 below). At the same time, I adjust the height of the probe so that the probe makes direct contact with the participant's chin, when the Articulate Assistant Advanced (AAA) software should start displaying the midsagittal tongue image.

Note that it is also necessary at this point to check from the participant's front again whether the probe still points straight up without any tilting, as a tilted probe

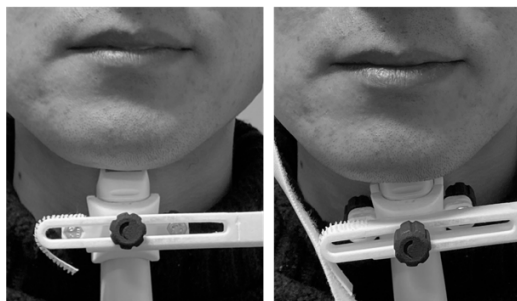


Figure 4 Comparisons of probe angle seen from front. While the probe stands straight up in the left picture, it is clearly tilted pointing off midsagittal in the picture to the right, caused by lifting the probe too much.

does not scan a midsagittal tongue shape appropriately, especially when the tongue goes far from the probe origin for e.g., a high front vowel /i/. When the probe is found to be tilted, the probe is likely being lifted up too much and thus applying too much pressure against the participant's chin, which is illustrated in Figure 4.

Although how clearly the tongue can be imaged varies from one participant to another, a rule of thumb in gauging the optimal probe angle is to check the position of the mandible and hyoid bone, both of which appear as black shadows in ultrasound images (Preston et al., 2017; see also the dark structures on both fan edges in Figure 2). When the tongue is recorded with the tongue tip pointing to the right, the mandible shadow is seen towards the right of the fan whereas the hyoid bone shadow is seen towards the left. It is suggested to orient the probe so that both shadows appear on each edge of the fan, or at least one of the shadows, as these structures are the few of the reference anatomical landmarks that can be imaged in ultrasound images (Preston et al., 2017).

3.1.3 Recording parameter setting

At this stage, recording parameters may also need to be adjusted on the AAA software. Crucial settings include *field of view* (FOV) and *depth*. A larger FOV results in a wider fan-shaped window capturing a wider area of midsagittal tongue surface. FOV is in a trade-off relationship with the framerate, where a larger FOV typically results in reduced framerate (Stone, 2005). AAA software achieves approximately 80 fps even with a 100% FOV setting (corresponding to a 101.2° FOV using a 20-mm radius convex probe in our Telemed MicrUS system), which should be adequate for capturing most of the lingual sounds including alveolar taps/flap [ɾ] (Recasens & Rodríguez, 2016).

The depth setting determines how far the ultrasound travels from the probe and how long the probe needs to wait to receive the reflected ultrasound (Stone, 2005). In short, this changes the size of the tongue for each participant displayed on the screen. I typically use the depth setting of 80 mm, which is optimal for most participants. For some participants, however, such as tall male speakers who have larger anatomical structures, I sometimes need to adjust the depth to 90 mm or even larger. Appropriate depth setting can be gauged by checking how well ultrasound captures the tongue shape for the high front vowel /i/.

3.1.4 Bite plane measurement

At this stage, almost everything is ready for the main

recording. Before proceeding, however, the participant's bite plane needs to be recorded. As described in the earlier section, I use a thin plastic plate similar to the one shown in Scobbie et al. (2011). I ask participants to bite the plastic bite plate by inserting the longer end into the mouth and placing the upper incisors at the small bump on the plate that acts like a stopper. Participants then push their tongues up against the flat surface under the plate, which visualises a flat occlusal plane with some deformation of the tongue (see the right image in Figure 2). Note that the palate shape can also be recorded at this stage by having the participant swallow water.

Overall, all these preparatory procedures should take up to 15 to 20 minutes depending on the researcher's experience and the anatomical characteristics of the participants. Note that it might not be possible to obtain the best quality images across all speakers, as it partly depends on factors related to participants such as the size of the lower jaw, allowing for a proper probe placement, as well as the presence of a beard.

3.2 Data processing

Ultrasound data obtained through the AAA software typically involves multiple short recordings of midsagittal tongue movement and synchronised audio recordings. While it might be possible to conduct subjective and qualitative judgements on tongue movement solely based on visual inspections of ultrasound images, it is often desirable to analyse the tongue data quantitatively through various data visualisation and/or statistical methods. For this, an ultrasound data analysis workflow usually begins with tongue spline estimation/tracking to allow for further quantitative analysis at the later stages.

3.2.1 Tongue spline tracing/estimation

Tongue shape analysis is usually conducted on tongue splines extracted from ultrasound images. Common software includes GetContours (Tiede, 2021) and EdgeTrack (Li et al., 2005). Most of the tongue spline extraction methods are implemented while relying on either the edge detection technique, in which pixel differences (i.e., differences in brightness) are used to estimate the tongue surface, or the whole image processing technique, in which whole ultrasound images are compressed into several numeric values via directly applying data dimensionality reduction techniques to ultrasound images (Wrench & Balch-Tomes, 2022). Edge detection based on pixel differences is, however, prone to substantial estimation

error due to the presence of noise or blurred images where tongue splines cannot be reliably determined (Wrench & Balch-Tomes, 2022). Similarly, it is not immediately easy to interpret the results of the whole image processing technique, especially in understanding what each of the dimensions represents in light of the original midsagittal tongue dimension.

One recent approach for faster and more reliable tongue spline estimation involves the *DeepLabCut* (DLC) toolkit (Mathis et al., 2018). DLC is a markerless pose estimation algorithm that has been used to estimate the position and movement of different body parts based on deep neural networks (Mathis et al., 2018). DLC can be implemented via the AAA software for ultrasound tongue surface estimation, and this achieves a faster and more accurate tongue surface estimation compared to the existing features in AAA or other software (Wrench & Balch-Tomes, 2022).

The current DLC models implemented in AAA estimate 11 key points along the tongue surface for each ultrasound frame, with two points capturing each of the key areas of the tongue surface including tongue root, tongue body, tongue dorsum, tongue blade and tongue tip as well as the epiglottic vallecula (i.e., a small depression between the root of the tongue and the epiglottis). Although some parts of the tongue (e.g., tongue tip and tongue root) can be difficult to see in ultrasound images when they are obstructed by hard structures, DLC *estimates* (rather than *tracks*) the position of these obscured parts based on the rest of the tongue available in the image based on pre-trained deep neural networks (Wrench & Balch-Tomes, 2022). The 11 sets of x/y coordinates are then exported for further data processing and statistical analysis. The bite plane information is expressed in the same manner, such that the origin and the end of the bite plane are expressed using the x/y coordinate, although the bite plane does not usually need to be exported for data analysis.

3.2.2 Bite plane normalisation

As shown in Section 2.2, the tongue splines need to be offset and rotated against the bite plane for each participant for cross-speaker tongue shape comparisons (e.g., Strycharczuk & Scobbie, 2017). In AAA, this can be implemented by defining the bite plane tracing as the 'fiducial' (i.e., reference) template, superimposing it to all ultrasound frames (so that both estimated tongue splines and bite plane tracing appear simultaneously for all ultrasound frames), and selecting 'offset and rotate' in the menu window for exporting the data.

4. Case study

In order to demonstrate further stages of a typical ultrasound data analysis workflow, I present a case study involving a brief ultrasound analysis in which I aim to answer the research question in the context of the current paper. The research question is ‘*How do L1 Japanese speakers differ from L1 English speakers in articulation for English /l/ and /ɹ/?*’. Codes and data to reproduce this analysis are publicly available in the OSF repository at <https://shorturl.at/hHJKS> (currently anonymised for peer review).

4.1 Participants

The data to be analysed here are obtained from 38 speakers, including 12 L1 North American English (10 female, 2 male), with a mean age of 29.7 years ($SD = 6.05$) and 26 L1 Japanese speakers (16 female, 10 male), with a mean age of 19.7 years ($SD = 1.05$). This is a subset of a larger corpus consisting of 56 speakers (42 L1 Japanese speakers and 14 L1 North American English speakers), chosen on the basis of clarity of ultrasound tongue image. L1 North American English speakers grew up either in the US or in Canada using English primarily up until the age of 13, all of whom identify as fluent L1 speakers of North American English. They resided in the UK at the time of the recording for their work and postgraduate study.

All L1 Japanese speakers were undergraduate students enrolled at universities located near the cities of Kobe and Nagoya in Japan at the time of recording. Their profile can be considered typical for learners of English as a foreign language in Japan who receive instruction primarily through the school curriculum. The mean length of their English study was 9.19 years ($SD = 2.02$). They did not have extensive experience in study abroad, with the mean length of overseas experience being 0.58 months ($SD = 1.08$). No participants reported any hearing or speaking impairments.

4.2 Procedure

The experiments took place between October and December 2022. Each session consisted of a speech production experiment involving ultrasound recording and a speech perception experiment. Audio and midsagittal tongue images were recorded simultaneously in a quiet room at universities in Japan for L1 Japanese speakers and in a sound-attenuated room in the UK for L1 North American English speakers. Audio signals were pre-amplified and

digitized at 44.1 kHz with 16-bit quantisation using a Sound Device USB-Pre2 audio interface. Ultrasound data were recorded using the Telemed MicrUS system with a 20-mm radius convex probe, synchronised with the audio signals and recorded on a laptop computer via the AAA software version 220.5.1 (Articulate Instruments, 2022). The recording parameters were consistent within-speaker but varied across speakers to obtain the most optimal tongue images within a range of probe frequency between 2-4 MHz, depth between 80-90 mm and field of view settings between 80-100% (91.6°-101.2° FOV), resulting in a framerate of ca. 80 frames per second.

Ethics approval was obtained from Lancaster University, Kobe Gakuin University and Meijo University. All participants were compensated for their time and participation with 2,000 Japanese Yen or 15 British Pound Sterling in the form of cash or vouchers commensurate with the regulations at each of the recording venues.

4.3 Data processing

The data to be analysed here are the tongue shapes for intervocalic English /l/ and /ɹ/ from a minimal pair ‘*believe*’ and ‘*bereave*’, produced between three to five times by the 38 speakers mentioned above. This results in a total of 297 tokens obtained from L1 North American English speakers ($n = 53$ for /l/; $n = 57$ for /ɹ/) and L1 Japanese speakers ($n = 94$ for /l/; $n = 93$ for /ɹ/). Segmentation was automatically conducted at the phone level via Montreal Forced Aligner version 2.0.6 (McAuliffe et al., 2017) and it was then adjusted wherever necessary using Praat (Boersma & Weenink, 2022). MFA performs overall well on tokens in which the liquid consonants were realised as approximants, but less so when English /l/ and /ɹ/ were substituted by an alveolar tap [ɾ] in some of L1 Japanese speakers’ data.

Tongue shapes were extracted at 11 equidistant points during the vowel~liquid~vowel intervals in these words (i.e., *believe* and *bereave*), such that the timepoint of 0% corresponds to the onset of the first vowel /ɪ/ and 100% to the offset of the second vowel /i/. Because visual inspections of spectrograms suggested that the English liquids occurred at approximately the 30% point during the interval, I extracted the tongue shape at the 30% timepoint as a proxy of tongue shapes for English /l/ and /ɹ/.

Following the data processing protocol above (see Section 3.2), I estimated tongue splines via DLC/AAA and tracked each speaker’s bite plane, which was then

superimposed onto all the ultrasound frames. I then exported the offset/rotated tongue splines within a defined interval. No hand correction was performed to the tongue splines.

4.4 Data analysis

In this case study, I describe a data analysis procedure using Principal Component Analysis (PCA). PCA is a data dimensionality reduction technique that can extract a small number of major abstract patterns based on correlations between data points in the raw articulatory data (Johnson, 2008; Mokhtari et al., 2007; Stone, 2005; Turton, 2017). PCA is particularly useful for ultrasound tongue imaging, in which systematically identifying main lingual variation has been a major challenge in data analysis (Davidson, 2006), and it has been used in previous ultrasound research to quantify articulatory characteristics of lateral allophony across dialects of British English (Turton, 2017), palatalised and non-palatalised consonants in Scottish Gaelic (Nance & Kirkham, 2022), and vowel-to-vowel coarticulations of consonants in German (Hoole & Pouplier, 2017).

PCA identifies orthogonal axes along which the greatest amount of variance can be found in the data, resulting in *eigenvectors* (principal components: PCs) expressing the direction of the variation, and *eigenvalues* indicating the amount of variance for each eigenvector (Hoole & Pouplier, 2017). These values are then used to compute *PC loadings* and *PC scores*, with the former expressing the relative weighting of each PC and the latter showing how much each PC can be associated with each token in the data set (Johnson, 2008).

Another advantage of PCA is that the identified PCs can be projected back to the original measurement unit; in this case, it is possible to show variations associated with each PC on the midsagittal tongue shape (Johnson, 2008). While the derived PCs are abstract, the reconstructed midsagittal tongue shapes facilitate linguistically meaningful interpretations of tongue movements on the front-back or high-low dimensions (Johnson, 2008; Stone, 2005).

The current study mostly follows the protocol for the PCA analysis in Nance and Kirkham (2022), whose analysis codes are also publicly available. Prior to the PCA analysis, the x/y coordinate values along the tongue splines are within-speaker z -normalised to facilitate cross-speaker comparisons. I then compute PC scores based on all tokens for all speakers using the *princomp* function in R. To interpret the PCs, I

reconstruct the variations expressed with each PC based on the PC loadings and the standard deviation scores for each data point against the mean tongue curve. Finally, I compare the PC scores between English /l/ and /ɫ/ across the two speaker groups. The codes for this analysis are available in the online supplementary materials.

4.5 Results

The PCA analysis identifies four principal components (PCs) that account for over 5% of variance in the data: PC1 (44.95%), PC2 (20.86%), PC3 (9.94%) and PC4 (8.34%), with a cumulative sum being 84.09%. The four PCs are retained here because the proportion of variance exceeds the threshold of 5%, following the recommendation by Baayen (2008). PC1 represents variation in tongue dorsum raising, whereas PC2 in tongue fronting. PC3 captures variation around the tongue posterior and tongue root, and PC4 captures a slight variation in tongue front. Due to the limitation of space, the current analysis focusses on the first two PCs only. Visualisation of reconstructed tongue shapes along each dimension is available in the online supplementary materials.

In order to interpret the dimension captured by the first two PCs in a linguistically meaningful way, tongue curves have been reconstructed showing the maximum and minimum tongue shapes represented by each PC in Figure 5. The thick curve in Figure 5 represents the mean tongue, with the variation in tongue shape expressed by adding and subtracting the standard deviation from the mean tongue curve, which are expressed using the plus (+) and minus (−) signs respectively. The left panel in Figure 5 suggests that PC1 captures tongue dorsum raising, in which lower PC1 values represent higher tongue dorsum. The right panel in Figure 5 shows the variation captured by PC2, corresponding to the degree of tongue fronting: Lower PC2 values indicate more retracted tongue positions.

Finally, the distributions of the PC scores are juxtaposed in Figure 6 for L1 English (dark grey) and L1 Japanese speakers (light grey) by segment (column) and PC (row). Overall, although the two speaker groups show similar mean PC scores, L1 Japanese speakers' distribution is wider than that of L1 English speakers, suggesting a more variable tongue shape for L1 Japanese speakers. Along the PC1 dimension, L1 English speakers show positive PC1 values ($M = 1.02$, $SD = 0.68$ for /l/; $M = 1.09$, $SD = 0.78$ for /ɫ/),

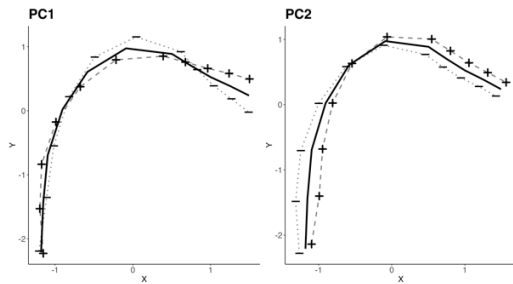


Figure 5 Reconstructed midsagittal tongue shape based on variation explained by PC1 (tongue dorsum lowering: left) and PC2 (tongue fronting: right). The thick curve represents the mean tongue shape, with the standard deviation added (+) and subtracted (-) from the mean to represent variation along each PC.

suggesting that their tongue dorsum is lowered for both /l/ and /ɹ/ (i.e., corresponding to the ‘+’ curve in the left panel in Figure 5). L1 Japanese speakers, on the other hand, show a slightly lower mean PC1 score with a wider distribution for English /l/ ($M = 0.52$, $SD = 0.92$) than L1 English speakers. For English /ɹ/, while the two speaker groups are similar in the mean PC1 value, L1 Japanese speakers show a slightly wider distribution judging from Figure 6 and the standard deviation ($M = 1.21$, $SD = 1.02$).

A similar tendency can be seen for PC2. For both English /l/ and /ɹ/, L1 English speakers show negative PC2 values ($M = -0.85$, $SD = 0.63$ for /l/; $M = -0.84$, $SD = 0.55$ for /ɹ/), indicating that their tongue shape is retracted along the tongue posterior and lower along the tongue anterior (i.e., the ‘-’ curve in the right panel in Figure 5). Although L1 Japanese speakers show negative mean PC2 values similar to L1 English speakers, the PC2 distributions are much wider, indicating a more variable tongue shape.

To summarise, the PCA analysis identifies four major components of the tongue shape for English /l/ and /ɹ/, with the first two PCs explaining approximately 66% of the variation in the data. PC1 corresponds to tongue dorsum raising whereas PC2 can be interpreted as tongue fronting. Visual inspection of the distributions of the PC scores suggests that L1 English speakers’ articulation is more consistent than L1 Japanese speakers, with an overall tendency of a lower and retracted tongue shape for L1 English speakers.

5. General Discussion and conclusion

The objective of this article is to illustrate a workflow of ultrasound tongue imaging analysis, guided by the

research context of L1 Japanese speakers’ production of English /l/ and /ɹ/. The research question asks what articulatory differences can be identified between L1 English and L1 Japanese speakers. I demonstrate that an appropriate data collection protocol combined with bite-plane rotation and Principal Component Analysis (PCA) facilitates a systematic quantitative articulatory analysis.

The main findings from the case study include that L1 Japanese speakers show greater variability in articulatory strategies for English /l/ and /ɹ/ in terms of (1) tongue dorsum height and (2) tongue retraction. L1 English speakers, on the other hand, show consistently lower tongue dorsum and retracted tongue shape. This finding agrees with the previous descriptions of English liquid articulation that commonly involves tongue retraction (Alwan et al., 1997; Narayanan et al., 1997; Stevens, 2000).

This paper shows that a combination of bite plane rotation and PCA is a powerful tool for quantitative analysis of midsagittal tongue shape using ultrasound. It is often challenging to systematically partition regions of the tongue surface due to the lack of anatomical landmarks in ultrasound images despite the rich dimensionality of the data (Davidson, 2006; Stone, 2005). Although it is possible to rely on metrics that do not depend on the tongue position for data analysis (e.g., Curvature Index, Stolar & Gick, 2013; Dorsum Excursion Index, Zharkova, 2013), it is often more meaningful to be able to explain articulatory variation in the data in phonetic dimensions such as ‘front-back’ or ‘high-low’ (Scobbie et al., 2011; Stone, 2005). The

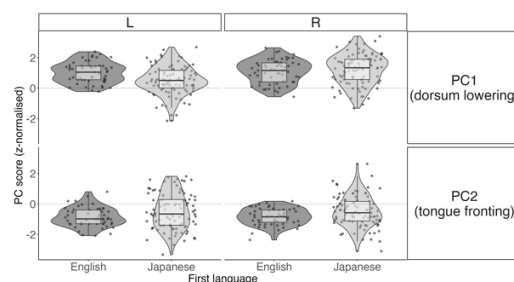


Figure 6 Violin plots and individual data points showing distributions of the PC scores for L1 English (dark grey) and L1 Japanese (light grey) speakers by segment (column) for each PC dimension (row). The x-axis represents the two speaker groups: L1 English (thick grey) and L1 Japanese (light grey) speakers. The y-axis indicates distributions of PC scores for each dimension. Zero corresponds to the mean tongue in Figure 5 and is indicated in a thin horizontal line.

PCA identifies the major variation in the data in a data-driven manner and the identified principal components can be used to reconstruct the tongue shape as illustrated in Figure 5.

The flexibility in subsequent statistical analysis is one of the strengths of PCA. Given that PCA summarises the dimension of data variability into numeric values (PC scores), as shown in Figure 6, further statistical analysis can be conducted using e.g., linear mixed-effect models (Nance & Kirkham, 2022) to formally test the effect of variables of interest. In the context of the current analysis, the PC scores along each dimension can be a dependent variable, predicted by fixed effects including the language group (two levels: L1 English vs L1 Japanese), segment (two levels: /l/ vs /l/), and the interaction between them. As suggested by an anonymous reviewer, each participant's proficiency effect could also be incorporated as a fixed effect. Note also that it is also possible to directly model differences in tongue contours using Generalised Additive Mixed-effects Models (Al-Tamimi & Palo, 2024; Strycharczuk et al., 2024; Wieling, 2018). While a previous study demonstrates that both GAMMs and PCA yield similar results for a dynamic analysis of tongue shape (Al-Tamimi & Palo, 2024), it could be argued that the flexibility in the choice of subsequent statistical analysis can be an advantage of the PCA approach.

The current analysis, however, has been conducted for illustrative purposes only and thus further methodological considerations are necessary for a more formal analysis. For example, the decisions as to where the tongue shape is extracted need to be justified further, such as at the maximal constriction during consonant production (e.g., Léger et al., 2023) or at the midpoint during a region of interest (e.g., Kirkham & Nance, 2017). Also, the within-speaker normalisation using *z*-score is one of many available normalisation methods, such as range normalisation. Recent developments in the DLC implementation of the AAA software also allow tongue shape to be normalised across speakers based on the distance between the short tendon and the tongue surface (Strycharczuk et al., 2024).

More importantly, while the current paper demonstrates the usefulness of the PCA analysis, it also illustrates some methodological challenges involved in data interpretation. PCA is purely a data-driven approach with no *a priori* physiological foundations, meaning that different data sets result in different PCs (Stone, 2005). This can be clearly shown by comparing the PCs identified in the current paper and in Nagamine (2023) that are similar data sets despite a few

methodological differences (e.g., dynamic vs static analysis, the number of participants and prompts). Whereas Nagamine (2023) identifies tongue retraction (PC1) and tongue height (PC2), which are the fundamental dimensions in describing tongue shape (e.g., Johnson, 2008), the current study does not offer such clear-cut decompositions of the tongue contours. Rather, PC2 in the current study may be similar to Turton's (2017) PC1 which corresponds to the clear-dark allophony of laterals in British English.

The entirely data-driven nature of PCA underscores the importance of appropriate data collection, including tongue image quality, probe stabilisation and bite plane rotation as described in this paper. However, with all these considered appropriately, ultrasound analysis using the bite plane rotation, coupled with appropriate statistical methods such as *z*-score normalisation and PCA, offers a promising avenue for a reliable between-speaker comparison (Stone, 2005). Ultrasound is one of the most accessible methods for articulatory research due to its low cost, non-invasiveness, portability and easier set-up procedure. This makes it particularly useful for use not just in a laboratory setting but also in language classrooms and fieldwork settings, offering insights into a wide range of research contexts by visualising the (once) invisible tongue movement.

Acknowledgements

The author thanks various people and funding bodies for supporting the research. This acknowledgements section is currently anonymised for peer-review.

References

- Al-Tamimi, J., & Palo, P. (2024). Retraction of the whole tongue induced by pharyngealisation in Levantine Arabic: A between-subject account using Static and Dynamic PCA and GAMMs. In I. Wilson, A. Mizoguchi, J. Perkins, J. Villegas, & N. Yamane (Eds.), *Ultrafest XI: Extended Abstracts* (pp. 79–83). University of Aizu.
- Alwan, A., Narayanan, S., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *The Journal of the Acoustical Society of America*, *101*(2), 1078–1089. <https://doi.org/10.1121/1.417972>
- Antolik, T. K., Pillot-Loiseau, C., & Kamiyama, T. (2019). The effectiveness of real-time ultrasound visual feedback on tongue movements in L2

- pronunciation training: Japanese learners' progress on the French vowel contrast /y/-/u/. *Journal of Second Language Pronunciation*, 5(1), 72–97. <https://doi.org/10.1075/jslp.16022.ant>
- Articulate Instruments. (2008). *Ultrasound stabilisation headset: Users manual revision 1.5*. Articulate Instruments.
- Articulate Instruments. (2022). *Articulate Assistant Advanced version 220* [Computer software]. Articulate Instruments.
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511801686>
- Boersma, P., & Weenink, D. (2022). *Praat: Doing Phonetics by Computer* (Version 6.2.09) [Computer software]. <https://www.fon.hum.uva.nl/praat/>
- Bryfonski, L. (2023). Is seeing believing?: The role of ultrasound tongue imaging and oral corrective feedback in L2 pronunciation development. *Journal of Second Language Pronunciation*, 9(1), 103–129. <https://doi.org/10.1075/jslp.22051.bry>
- Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America*, 120(1), 407–415. <https://doi.org/10.1121/1.2205133>
- Derrick, D., Carignan, C., Chen, W., Shujau, M., & Best, C. T. (2018). Three-dimensional printable ultrasound transducer stabilization system. *The Journal of the Acoustical Society of America*, 144(5), EL392–EL398. <https://doi.org/10.1121/1.5066350>
- Derrick, D., & Gick, B. (2011). Individual variation in English flaps and taps: A case of categorical phonetics. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique*, 56(3), 307–319. <https://doi.org/10.1353/cjl.2011.0024>
- Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Studies in Bilingualism* (Vol. 36, pp. 309–322). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.36.15gic>
- Hoole, P., & Pouplier, M. (2017). Öhman returns: New horizons in the collection and analysis of imaging data in speech production research. *Computer Speech & Language*, 45, 253–277. <https://doi.org/10.1016/j.csl.2017.03.002>
- Iskarous, K., & Pouplier, M. (2022). Advancements of phonetics in the 21st century: A critical appraisal of time and space in Articulatory Phonology. *Journal of Phonetics*, 95, 101195. <https://doi.org/10.1016/j.wocn.2022.101195>
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1), B47–B57. [https://doi.org/10.1016/S0010-0277\(02\)00198-1](https://doi.org/10.1016/S0010-0277(02)00198-1)
- Johnson, K. (2008). *Quantitative methods in linguistics*. Blackwell Pub.
- Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, 94(2), 701–714. <https://doi.org/10.1121/1.406887>
- Kirkham, S., & Nance, C. (2017). An acoustic-articulatory study of bilingual vowel production: Advanced tongue root vowels in Twi and tense/lax vowels in Ghanaian English. *Journal of Phonetics*, 62, 65–81. <https://doi.org/10.1016/j.wocn.2017.03.004>
- Kirkham, S., Strycharczuk, P., Gorman, E., Nagamine, T., & Wrench, A. (2023). Co-registration of simultaneous high-speed ultrasound and electromagnetic articulography for speech production research. In R. Skarnitzl & J. Volin (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 942–946). Guarant International.
- Klein, H. B., McAllister Byun, T., Davidson, L., & Grigos, M. I. (2013). A Multidimensional Investigation of Children's /r/ Productions: Perceptual, Ultrasound, and Acoustic Measures. *American Journal of Speech-Language Pathology*, 22(3), 540–553. [https://doi.org/10.1044/1058-0360\(2013/12-0137\)](https://doi.org/10.1044/1058-0360(2013/12-0137))
- Kochetov, A. (2020). Research methods in articulatory phonetics I: Introduction and studying oral gestures. *Language and Linguistics Compass*, 14(4), 1–29. <https://doi.org/10.1111/lnc3.12368>
- Léger, A., King, H., & Ferragne, E. (2023). Is rhoticity on the tip of your tongue? Tongue shapes for English /r/ in French learners with ultrasound. In R. Skarnitzl & J. Volin (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 2741–2745). Guarant International.
- Li, M., Kambhamettu, C., & Stone, M. (2005). Automatic contour tracking in ultrasound images. *Clinical Linguistics & Phonetics*, 19(6–7), 545–554. <https://doi.org/10.1080/02699200500113616>

- Maekawa, K. (2023). Articulatory characteristics of the Japanese /ɾ/: A real-time MRI study. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 992–996). Guarant International.
- Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), Article 9. <https://doi.org/10.1038/s41593-018-0209-y>
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech 2017*, 498–502. <https://doi.org/10.21437/Interspeech.2017-1386>
- Mielke, J., Baker, A., & Archangeli, D. (2016). Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /s/. *Language*, *92*(1), 101–140. <https://doi.org/10.1353/lan.2016.0019>
- Mokhtari, P., Kitamura, T., Takemoto, H., & Honda, K. (2007). Principal components of vocal-tract area functions and inversion of vowels by linear regression of cepstrum coefficients. *Journal of Phonetics*, *35*(1), 20–39. <https://doi.org/10.1016/j.wocn.2006.01.001>
- Nagamine, T. (2023). Dynamic tongue movements in L1 Japanese and L2 English liquids. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 2442–2446). Guarant International.
- Nance, C., Dewhurst, M., Fairclough, L., Forster, P., Kirkham, S., Nagamine, T., Turton, D., & Wang, D. (2023). Acoustic and articulatory characteristics of rhoticity in the North-West of England. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 3573–3577). Guarant International.
- Nance, C., & Kirkham, S. (2022). Phonetic typology and articulatory constraints: The realisation of secondary articulations in Scottish Gaelic rhotics. *Language*, 419–460.
- Narayanan, S. S., Alwan, A. A., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals. *The Journal of the Acoustical Society of America*, *101*(2), 1064–1077. <https://doi.org/10.1121/1.418030>
- Preston, J. L., McAllister Byun, T., Boyce, S. E., Hamilton, S., Tiede, M., Phillips, E., Rivera-Campos, A., & Whalen, D. H. (2017). Ultrasound Images of the Tongue: A Tutorial for Assessment and Remediation of Speech Sound Errors. *Journal of Visualized Experiments*, *119*, 55123. <https://doi.org/10.3791/55123>
- Pucher, M., Klingler, N., Luttenberger, J., & Spreafico, L. (2020). Accuracy, recording interference, and articulatory quality of headsets for ultrasound recordings. *Speech Communication*, *123*, 83–97. <https://doi.org/10.1016/j.specom.2020.07.001>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (Version 4.3.2) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rebernik, T., Jacobi, J., Jonkers, R., Noiray, A., & Wieling, M. (2021). A review of data collection practices using electromagnetic articulography. *Laboratory Phonology*, *12*(1), Article 1. <https://doi.org/10.5334/labphon.237>
- Recasens, D., & Rodríguez, C. (2016). A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound. *Journal of Phonetics*, *59*, 58–75. <https://doi.org/10.1016/j.wocn.2016.09.002>
- Riney, T. J., Takada, M., & Ota, M. (2000). Segmentals and Global Foreign Accent: The Japanese Flap in EFL. *TESOL Quarterly*, *34*(4), 711–737. <https://doi.org/10.2307/3587782>
- Saito, K., & van Poeteren, K. (2018). The perception–production link revisited: The case of Japanese learners’ English /s/ performance. *International Journal of Applied Linguistics*, *28*(1), 3–17. <https://doi.org/10.1111/ijal.12175>
- Scobbie, J., Lawson, E., Cowen, S., Cleland, J., & Wrench, A. (2011). A common co-ordinate system for mid-sagittal articulatory measurement. *QMU CASL Working Papers*, *20*, 1–4.
- Scobbie, J., Stuart-Smith, J., & Lawson, E. (2012). Back to front: A socially-stratified ultrasound tongue imaging study of Scottish English /u/. *Italian Journal of Linguistics/Rivista Di Linguistica*, *24*(1), 103–148.
- Slud, E., Stone, M., Smith, P. J., & Goldstein Jr., M. (2002). Principal Components Representation of the Two-Dimensional Coronal Tongue Surface. *Phonetica*, *59*(2–3), 108–133. <https://doi.org/10.1159/000066066>
- Spreafico, L., Pucher, M., & Matosova, A. (2018). UltraFit: A Speaker-friendly Headset for Ultrasound Recordings in Speech Science.

- Interspeech* 2018, 1517–1520.
<https://doi.org/10.21437/Interspeech.2018-995>
- Stevens, K. N. (2000). *Acoustic Phonetics*. The MIT Press.
- Stolar, S., & Gick, B. (2013). An Index for Quantifying Tongue Curvature. *Canadian Acoustics*, 41(1), Article 1.
- Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, 19(6–7), 455–501.
<https://doi.org/10.1080/02699200500113558>
- Strycharczuk, P., Lloyd, S., & Scobbie, J. M. (2024). Apparent time change in the articulation of onset rhotics in Southern British English. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 3602–3606).
- Strycharczuk, P., & Scobbie, J. M. (2017). Fronting of Southern British English high-back vowels in articulation and acoustics. *The Journal of the Acoustical Society of America*, 142(1), 322–331.
<https://doi.org/10.1121/1.4991010>
- Tiede, M. (2021). *GetContours version 3.5* [Computer software].
<https://github.com/mktiede/GetContours>
- Turton, D. (2017). Categorical or gradient? An ultrasound investigation of /l/-darkening and vocalization in varieties of English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1), 1–13.
<https://doi.org/10.5334/labphon.35>
- Westbury, J. R. (1994). On coordinate systems and the representation of articulatory movements. *The Journal of the Acoustical Society of America*, 95(4), 2271–2273.
<https://doi.org/10.1121/1.408638>
- Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S. (2005). The Haskins Optically Corrected Ultrasound System (HOCUS). *Journal of Speech, Language, and Hearing Research*, 48(3), 543–553. [https://doi.org/10.1044/1092-4388\(2005/037\)](https://doi.org/10.1044/1092-4388(2005/037))
- Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, 70, 86–116.
<https://doi.org/10.1016/j.wocn.2018.03.002>
- Wilson, I. (2014). Using ultrasound for teaching and researching articulation. *Acoustical Science and Technology*, 35(6), 285–289.
<https://doi.org/10.1250/ast.35.285>
- Wrench, A., & Balch-Tomes, J. (2022). Beyond the Edge: Markerless Pose Estimation of Speech Articulators from Ultrasound and Camera Images Using DeepLabCut. *Sensors*, 22(3), Article 3.
<https://doi.org/10.3390/s22031133>
- Yamane, N., Howson, P., & Po-Chun (Grace), W. (2015). An ultrasound examination of taps in Japanese. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 1–5). The International Phonetic Association.
- Zharkova, N. (2013). Using Ultrasound to Quantify Tongue Shape and Movement Characteristics. *The Cleft Palate-Craniofacial Journal*, 50(1), 76–81.
<https://doi.org/10.1597/11-196>
- Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/. *The Journal of the Acoustical Society of America*, 123(6), 4466–4481.
<https://doi.org/10.1121/1.2902168>

Chapter 5

Pilot study 1: Dynamic tongue movements in L1 Japanese and L2 English liquids

This chapter presents a pilot study in which I compare changes in the midsagittal tongue shape in the vowel-liquid-vowel sequences produced by L1 Japanese speakers and L1 English speakers. This study serves as a preliminary study for the dynamic analyses to be presented in Chapters 7 and 8, demonstrating through qualitative and quantitative analyses (using PCA) that L1 Japanese speakers may differ from L1 English speakers in the use of tongue dorsum when producing L2 English liquids. This study has been presented at the International Congress of Phonetic Sciences (ICPhS) 2023, held in Prague, Czech Republic, in August 2023 and published in the conference proceedings.

DYNAMIC TONGUE MOVEMENTS IN L1 JAPANESE AND L2 ENGLISH LIQUIDS

Takayuki Nagamine

Lancaster University
t.nagamine@lancaster.ac.uk

ABSTRACT

Dynamic tongue movements of intervocalic English and Japanese liquids are analysed based on ultrasound data obtained from 17 L1 Japanese and 12 L1 North American English speakers. Principal Component Analysis (PCA) identified three key articulatory properties, including variation in tongue retraction and height as well as tongue tip movement. The time-varying changes showed different magnitude and timing for tongue retraction for Japanese speakers compared to English speakers. Substitution of English liquids with Japanese /r/ was not clearly observed in the tongue movement pattern. The findings highlight the complexity involved in the articulation of English liquids in L2 speech and the usefulness of the finer-grained articulatory analysis in understanding the particular challenges L2 learners face in producing English liquids in L2 speech.

Keywords: L2 speech production, liquids, ultrasound, Principal Component Analysis

1. INTRODUCTION

The acquisition of English liquids /l/ and /ɹ/ presents particular challenges to L2 learners, notably to L1 Japanese speakers [1]. English /l/ and /ɹ/ can be considered gesturally complex sounds because multiple articulatory gestures need to be coordinated spatially and temporally [2, 3]. English /l/ requires two lingual gestures, tongue coronal and dorsal gestures, which are patterned differently depending on the syllable position. Similarly, English /ɹ/ requires coordination of labial, tongue anterior and posterior gestures, and the tongue shape shows substantial cross-speaker variation [4, 5].

While it is generally agreed that L2 learners' difficulty in L2 speech production is rooted in the perception of L2 sounds [1], adult L2 learners' established articulatory routines that are optimised for L1 production could also constrain the accurate production of L2 sounds, especially gesturally complex sounds such as English liquids [6]. L2

learners struggle to produce articulatory gestures that are absent in their L1 [7]. Previous articulatory studies suggest that Japanese speakers' tongue dorsum movement may be the key for them to produce L2 English liquids accurately. Less advanced L1 Japanese learners of English exhibit little movement in the tongue posterior compared to the more experienced learner or the L1 English speaker [8]. Also, the substitution of English liquids with Japanese /r/ was not clearly observed in articulatory data even for less advanced Japanese learners of English [9]. These studies illustrate a learning scenario of L1 Japanese learners of English; while they attempt to differentiate English liquids from Japanese /r/, they struggle to realise the dorsal gesture in producing English liquids.

Japanese has one liquid category /r/, canonically realised as alveolar taps or flaps [ɾ] [10]. In previous research, whether a dorsal gesture is actively involved in Japanese /r/ remains unclear. The tongue dorsum in plain taps [ɾ] shows a stronger coarticulatory effect with the vowels than the palatalised taps [ɾʲ] [11]. Compared to alveolar stops, however, the tongue dorsum in alveolar taps/flaps is more retracted and stabilised [12, 13]. It would, therefore, be useful to compare the tongue movement between English and Japanese liquids in discussing the articulatory L1 influence in L2 speech, especially with regard to the tongue dorsum gesture.

Identifying specific articulatory difficulties in L2 speech also has theoretical implications. The Speech Learning Model (SLM) [14] hypothesises that articulatory realisation rules for L2 sounds are specified in the phonological representation. The Perceptual Assimilation Model for L2 learning (PAM-L2) [15] posits that L2 learners perceive articulatory gestures directly, and they assimilate L2 sounds into the L1 phonological categories based on the gestural information. The case of Japanese speakers' acquisition of English liquids could be a good testing ground as to whether and what articulatory information is important for successful L2 speech learning.

Building on the previous research, the current

study aims to identify and compare key articulatory properties involved in English liquids produced by L1 English and L1 Japanese speakers. I particularly focus on the tongue dorsum movement for which I expect to observe the L1 influence carried over from Japanese /r/. Methodologically, the use of ultrasound tongue imaging would complement the findings of previous qualitative articulatory work.

2. METHOD

2.1. Participants

Data from 29 speakers are analysed in the study, including 17 L1 Japanese (eight females and nine males, $M_{age} = 19.76$ years, $SD_{age} = 0.97$) and 12 L1 North American English speakers (ten females and two males, $M_{age} = 29.08$ years, $SD_{age} = 6.30$).

The Japanese speakers represent the typical English-as-a-foreign-language (EFL) learner population in Japan. They are university students from two universities located in Western and Central Japan who studied English mostly through the school and the university curriculum. They had little experience of a long-term stay outside Japan, except for six who had stayed in an English-speaking country for up to four months.

The North American English speakers were recruited in London and Lancaster in the UK, eight of whom came from the US and four from Canada. Although one of the Canadian speakers was originally born in Poland, she moved to Canada early in her childhood and thus considered herself an L1 Canadian English speaker. No participants reported any history of speech or hearing impairments.

2.2. Materials

Intervocalic English and Japanese liquids are analysed, elicited with the words *believe* /bi'li:v/, *bereave* /bi'ri:v/, and *biribiri* /biribiri/ (ʔʔʔʔʔʔʔʔʔ). These words are a subset of a larger data collection session. The flanking vowel /i/ is chosen because the vowel quality is similar in Japanese and English and is appropriate for cross-linguistic comparisons [10]. *biribiri* is a Japanese mimetic word that describes the sound and/or the situation of paper being torn. The intervocalic environment most likely yields the canonical tap/flap realisations of Japanese /r/ as it is subject to allophonic variations in other environments [16]. The Japanese speakers were asked to read the Japanese *biribiri* in the LHHH accent so that the liquids in both languages appeared as an onset consonant in an accented syllable.

2.3. Data collection

Participants wore an ultrasound stabilisation headset to stabilise the ultrasound probe under their lower jaw [17]. At the beginning of the recording, they were asked to bite a plastic plate to measure their occlusal plane [18]. Then, the participants read aloud the target words in isolation, resulting in 271 tokens for analysis. The Japanese participants' language modes were controlled by changing the language of instructions and including an English conversation activity between the Japanese and English recording blocks.

Ultrasound data were obtained using a Telemed MicrUs system, with a 64-element probe of 20 mm radius, recorded with the Articulate Assistant Advanced (AAA) software version 220.4.1 [19]. Midsagittal tongue views were imaged with a fixed probe frequency between 2-4 MHz, 80 mm depth, 100% field of view and 64 scan lines, resulting in a framerate of ca. 80 per second. Simultaneous acoustic signals were collected with the signal pre-amplified and digitised using a USBPre2 audio interface, and then recorded onto a laptop computer at 44.1 kHz with 16-bit quantisation. Side-profile lip images and participants' perceptual identification accuracy data for English /l/ and /r/ were also collected but will not be presented here.

2.4. Data analysis

Tongue spline data were exported from AAA using the DeepLabCut (DLC) plug-in [20]. DLC estimates the tongue shape based on 11 key points along the tongue contours in the ultrasound videos based on the trained models [21].

The current study takes the dynamic measurements throughout the vowel-liquid-vowel (V_1LV_2) interval in the target words (i.e., *believe*, *bereave*, and *biribiri*) under the assumption that English liquids exhibit dynamic changes in the acoustic and articulatory realisations, and it is often difficult to specify a particular time point that best represents the liquid quality [22, 23]. Segmentation was carried out based on the acoustic signals using Montreal Forced Aligner [24], with the V_1 onset and the V_2 offset marked at the point where periodic cycles began or ended in waveforms and where the formant structures were clearly visible.

Principal Component Analysis (PCA) has been performed in order to summarise the tongue spline data into a manageable number of key articulatory dimensions that allow for cross-speaker comparisons [25]. Prior to running PCA, data from each speaker were normalised into z -scores to allow for

comparisons across speakers. Then, PCA was run based on all the x and y coordinates of the 11 points along the tongue surface at 11 equidistant time points during the V_1LV_2 interval produced by all speakers. PCA was performed using the *princomp* function in R [26]. The code and data for analysis are available online at <https://osf.io/29tac/>.

3. RESULTS

3.1. Identifying key articulatory properties using PCA

PCA identified three key dimensions that account for 88.33% of the data, with each dimension exceeding the 5% threshold suggested in the literature [27]: PC1: 58.03%, PC2: 22.65%, and PC3: 7.65%. In order to make the PCs interpretable, the loadings of each PC are plotted against the mean tongue shape using the standard deviation information [25]. Fig 1 shows key dimensions involved in the production of the V_1LV_2 sequence concerning the degrees of tongue retraction (PC1: left), tongue height (PC2: middle), and the variation in the tongue tip (PC3: right). Given the focus of the current study being the tongue posterior movement, I will focus on PC1 in the following sections.

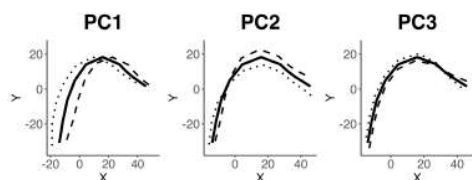


Figure 1: Variation captured in PCs 1 to 3. The thick line represents mean tongue for all of the data for all speakers, with the dimension of variation captured by that PC shown in the dashed and dotted lines.

3.2. Dynamic changes in tongue retraction

Fig 2 presents the time-varying changes of the PC1 values (i.e., tongue retraction) during the V_1LV_2 intervals aggregated for the L1 English and Japanese speakers, where larger PC values indicate more tongue fronting. The overall shape of the PC1 trajectories shows a somewhat similar front-back tongue movement for English liquids for both English and Japanese speakers. The variation associated with the degree of tongue retraction, however, is smaller for the Japanese speakers than the English speakers during the production of *believe* (blue). Regarding *bereave* (yellow), despite a similar

degree of tongue retraction between the two speaker populations, the Japanese speakers seem to show a different timing in which they achieve tongue retraction later during the interval than the English speakers do. Finally, changes in PC1 for Japanese /r/ in *biribiri* (grey) were different from either English /l/ or /ɹ/.

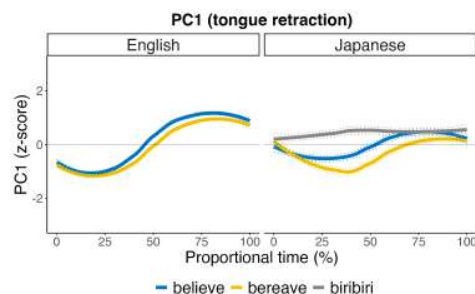


Figure 2: Time-varying changes of the tongue retraction (PC1).

In light of the articulatory patterns identified by PCA, three speakers have been selected to compare the articulatory differences in detail: a female speaker of Canadian English ('English A'), a male and a female EFL Japanese learner of English ('Japanese A' and 'Japanese B', respectively). The midsagittal tongue splines extracted at 11 equidistant points during the V_1LV_2 intervals are presented in Fig 3 and the time-varying changes in PC1 in Fig 4. The two Japanese speakers are chosen based on the auditory impression of the author.

Japanese A maintains an auditory three-way contrast among the three liquids, which is also obvious in the midsagittal tongue shapes. Changes in the PC1 values, however, suggest that the front-back tongue movement would be different from that of English A. Similarly, despite her clear substitution in the auditory analysis and the similarity between English and Japanese liquids in the tongue shapes, the PC1 movement for Japanese B's English /l/ and /ɹ/ is different from that for Japanese /r/.

4. DISCUSSION

The main findings of the current study include that English and Japanese speakers differed in the magnitude of tongue retraction for *believe* and timing for *bereave*. The results could be taken as evidence that the tongue posterior movement may impose difficulty on Japanese EFL learners in articulating English /l/ and /ɹ/ accurately.

While only one vowel environment has been investigated in this study, the current results could

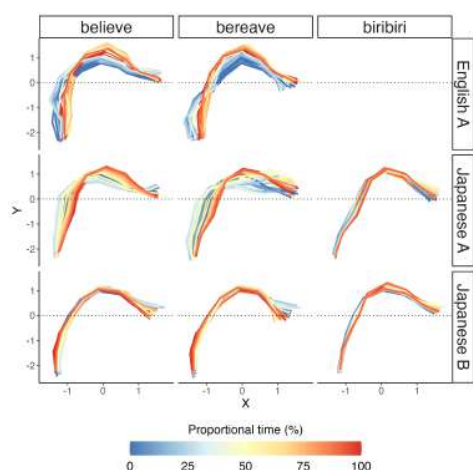


Figure 3: Midsagittal tongue shapes during the V_1LV_2 intervals for a Canadian English speaker (top) and two Japanese speakers (middle and bottom). Tongue tip to the right.

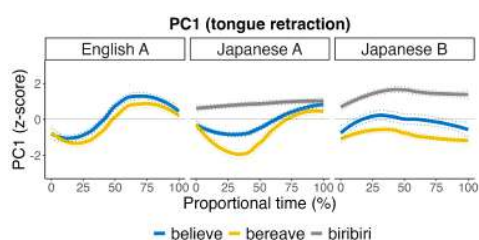


Figure 4: Time-varying changes of the tongue retraction (PC1) during the V_1LV_2 interval for a L1 English and two L1 Japanese speakers.

nevertheless characterise the front-back dimension of the tongue movement for Japanese /r/ compared to the English liquids. The changes in PC1 in Fig 2 suggest minor movements of the tongue posterior over the course of the V_1LV_2 sequence. This agrees with the previous claims that the tongue dorsum for the alveolar taps/flaps shows a substantial coarticulatory effect with the flanking vowels [11] and the dorsal ‘stabilisation’ strategy [12].

The time-varying trajectory of the front-back movement for *believe* in Fig 2 is ‘flatter’ for the Japanese speakers than for the English speakers. This may indicate the dorsal stabilisation carried over from Japanese /r/ to the English liquids. In addition, the trajectories for *bereave* (yellow in Fig 2) suggest different gestural timing patterns between Japanese and English speakers’ production, such that the Japanese speakers achieved tongue retraction later than the English speakers. Overall,

these results could provide evidence for the previous claim that L1 influence is observed in the tongue posterior movement in L1 Japanese speakers’ production of English liquids [8].

The individual data replicate the previous findings [9] that a less advanced EFL learner (i.e., Japanese B) does not substitute English liquids with Japanese /r/ completely given the clear differences in the trajectory height (see Fig 4). SLM might explain that she forms separate articulatory realisation rules for English /l/, /ɹ/ and Japanese /r/. Under PAM’s account, she might not yet be fully capable of using the dorsal gesture in learning the L1-L2 contrast of liquids. Given that dynamic information of segments might need to be part of phonological representation [28], the dynamic approach in this research is worth pursuing, especially for English liquids that show dynamic changes in articulation.

Finally, the interpretation of the liquid quality in the dynamic approach used in the current study may be subject to the articulatory realisations of vowels. This could be true of V_1 ; The preliminary acoustic analysis suggested that the second formant frequencies for V_1 seem to differ by 500 Hz between the two participant populations, meaning that the tongue retraction effect could not only be due to the articulatory realisations of English liquids but also the carry-over coarticulatory effects from V_1 .

5. CONCLUSIONS

The study analysed dynamic tongue movements in L2 English liquids produced by L1 Japanese speakers. Dynamic analyses of the principal components show a smaller tongue dorsum movement for English liquids produced by L1 Japanese speakers compared to L1 English speakers. Future research could analyse liquids in other vowel environments and the intergestural timing to better generalise this assumed dorsal coarticulatory effect. The participants’ perceptual identification accuracy data will also provide further theoretical insights into the nature of L2 speech learning.

6. ACKNOWLEDGEMENTS

Thank you to all the participants for their time and efforts. I thank Dr Claire Nance and Dr Sam Kirkham for their comments and support. Prof. Noriko Nakanishi, Prof. Yuri Nishio, and Dr Bronwen Evans helped me with data collection. The research is financially supported by Graduate Scholarship for Degree-Seeking Students by Japan Student Services Organization and the 2022 Research Grant by the Murata Science Foundation.

7. REFERENCES

- [1] J. E. Flege, K. Aoyama, and O.-S. Bohn, "The Revised Speech Learning Model (SLM-r) Applied," in *Second Language Speech Learning*, 1st ed., R. Wayland, Ed. Cambridge University Press, Feb. 2021, pp. 84–118.
- [2] D. Recasens, "A cross-language acoustic study of initial and final allophones of /l/," *Speech Communication*, vol. 54, no. 3, pp. 368–383, 2012.
- [3] M. Proctor, R. Walker, C. Smith, T. Szalay, L. Goldstein, and S. Narayanan, "Articulatory characterization of English liquid-final rimes," *Journal of Phonetics*, vol. 77, p. 100921, 2019.
- [4] J. Mielke, A. Baker, and D. Archangeli, "Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/," *Language*, vol. 92, no. 1, pp. 101–140, 2016.
- [5] H. King and E. Ferragne, "Loose lips and tongue tips: The central role of the /r/-typical labial gesture in Anglo-English," *Journal of Phonetics*, vol. 80, p. 100978, 2020.
- [6] B. Gick, P. Bacsfalvi, B. M. Bernhardt, S. Oh, S. Stolar, and I. Wilson, "A motor differentiation model for liquid substitutions in children's speech," in *Proceedings of the Meeting Acoustics*, vol. 1, Salt Lake City, Utah, 2008, p. 060003.
- [7] S. Harper, L. Goldstein, and S. S. Narayanan, "L2 Acquisition and Production of the English Rhotic Pharyngeal Gesture," in *Interspeech 2016*. ISCA, Sep. 2016, pp. 208–212.
- [8] G. N. Zimmermann, P. Price, and T. Ayusawa, "The production of English /r/ and /l/ by two Japanese speakers differing in experience with English," *Journal of Phonetics*, vol. 12, no. 3, pp. 187–193, 1984.
- [9] J. Moore, J. Shaw, S. Kawahara, and T. Arai, "Articulation strategies for English liquids used by Japanese speakers," *Acoustical Science and Technology*, vol. 39, no. 2, pp. 75–83, 2018.
- [10] T. J. Vance, *The Sounds of Japanese*. Cambridge University Press, 2008.
- [11] N. Yamane, P. Howson, and W. Po-Chun (Grace), "An ultrasound examination of taps in Japanese," in *Proceedings of the 18th International Congress of Phonetic Sciences*, Aug. 2015.
- [12] M. Proctor, "Towards a gestural characterization of liquids: Evidence from Spanish and Russian," *Laboratory Phonology*, vol. 2, no. 2, pp. 451–485, 2011.
- [13] M. Morimoto, "Geminated Liquids in Japanese: A Production Study," Ph.D. dissertation, University of California Santa Cruz, Mar. 2020.
- [14] J. E. Flege, "The intelligibility of English vowels spoken by British and Dutch talkers," in *Studies in Speech Pathology and Clinical Linguistics*, R. D. Kent, Ed. Amsterdam: John Benjamins Publishing Company, 1992, vol. 1, pp. 157–232.
- [15] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, O.-S. Bohn and M. J. Munro, Eds. Amsterdam: John Benjamins Publishing Company, 2007, pp. 13–34.
- [16] T. Arai, "On Why Japanese /r/ Sounds are Difficult for Children to Acquire," in *Interspeech 2013*, Lyon, France, 2013, pp. 2445–2449.
- [17] L. Spreafico, M. Pucher, and A. Matosova, "UltraFit: A Speaker-friendly Headset for Ultrasound Recordings in Speech Science," in *Interspeech 2018*. ISCA, Sep. 2018, pp. 1517–1520.
- [18] J. Scobbie, E. Lawson, S. Cowen, J. Cleland, and A. Wrench, "A common co-ordinate system for mid-sagittal articulatory measurement," *QMU CASL Working Papers*, vol. 20, pp. 1–4, 2011.
- [19] Articulate Instruments, "Articulate Assistant Advanced version 220," Articulate Instruments, Edinburgh, 2022.
- [20] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, "DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning," *Nature Neuroscience*, vol. 21, no. 9, pp. 1281–1289, 2018.
- [21] A. Wrench and J. Balch-Tomes, "Beyond the Edge: Markerless Pose Estimation of Speech Articulators from Ultrasound and Camera Images Using DeepLabCut," *Sensors*, vol. 22, no. 3, p. 1133, 2022.
- [22] J. Ying, J. A. Shaw, C. Kroos, and C. T. Best, "Relations Between Acoustic and Articulatory Measurements of /l/," in *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, Sydney, Dec. 2012, pp. 109–112.
- [23] S. Kirkham, C. Nance, B. Littlewood, K. Lightfoot, and E. Groarke, "Dialect variation in formant dynamics: The acoustics of lateral and vowel sequences in Manchester and Liverpool English," *The Journal of the Acoustical Society of America*, vol. 145, no. 2, pp. 784–794, 2019.
- [24] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in *Interspeech 2017*. ISCA, Aug. 2017, pp. 498–502.
- [25] C. Nance and S. Kirkham, "Phonetic typology and articulatory constraints: The realisation of secondary articulations in Scottish Gaelic rhotics," *Language*, pp. 419–460, 2022.
- [26] R Core Team, "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, 2022.
- [27] R. H. Baayen, *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press, 2008.
- [28] G. Schwartz and K. Kaźmierski, "Vowel dynamics in the acquisition of L2 English – an acoustic study of L1 Polish learners," *Language Acquisition*, vol. 27, no. 3, pp. 227–254, Jul. 2020.

Chapter 6

Pilot study 2: Acquisition of allophonic variation in second language speech: An acoustic and articulatory study of English laterals by Japanese speakers

This chapter presents the second pilot study investigating how L1 Japanese speakers produce L2 English lateral allophony. Chronologically, this is the first pilot study conducted ahead of the rest of the five studies between January and March 2022. The data were collected specifically for this study, so the study is based on a different data set, in which five L1 Japanese speakers who were deemed to be proficient L2 English speakers produced word-initial and word-final laterals in target words embedded in the carrier phrase. In this study, I demonstrate that proficient L1 Japanese-L2 English speakers realise the onset-coda lateral allophony in acoustics, but the midsagittal tongue shape data suggests that their articulation is influenced more by the vowel context than by the syllable position. This work constitutes a basis for Chapter 9 as well as studies outside the thesis (e.g., Colantoni et al.

(2023b)). This paper has been presented at Interspeech 2022, held in Incheon, South Korea in September 2022 and published in the conference proceedings.



Acquisition of allophonic variation in second language speech: An acoustic and articulatory study of English laterals by Japanese speakers

Takayuki Nagamine¹

¹Lancaster University

t.nagamine@lancaster.ac.uk

Abstract

Acquisition of positional allophonic variation is seen as the foundation of a successful L2 speech learning. However, previous research has mostly focused on the phonemic contrast between English /l/ and /r/, providing little evidence in the acquisition of positional allophones, such as those in English /l/.

The current study investigates the acoustics and articulation of allophonic variations in English laterals produced by Japanese speakers, focusing on the effects of syllabic positions and flanking vowels. Acoustic and articulatory data were obtained from five Japanese speakers in a simultaneous audio and high-speed ultrasound tongue imaging recording set-up while they read sentences containing syllable-initial and -final tokens of English /l/ in four different vowel contexts. Acoustic analysis was conducted on 500 tokens using linear-mixed effects modelling and the articulatory data were analysed using generalised additive mixed modelling.

Syllable position and vowel context had significant effects on acoustics, while midsagittal tongue shape was more influenced by vowel context, with fewer positional effects. The results demonstrate that differences in acoustics not always be mirrored exactly by midsagittal tongue shape, suggesting multidimensionality of articulation in second language speech.

Index Terms: articulation, acoustics, English laterals, second language speech, ultrasound tongue imaging

1. Introduction

In second language speech learning, it is well-established that the phonetics and phonology of a learner's first language (L1) interferes with acquisition of nonnative sounds in a native-like manner [1], [2]. The Speech Learning Model (SLM) posits that acquisition of position-sensitive allophones is seen as the fundamental mechanisms of second language (L2) speech learning [3]. However, while acquisition of the English liquid contrast between /l r/ by Japanese speakers has been well studied, there have been relatively fewer studies that investigate acquisition of allophonic variation in English /l/.

The two canonical allophonic variants of English /l/, typically referred to as *clear-L* and *dark-L*, are acoustically distinguished by F2 and the distance between F2 and F1 [4]. *Clear-L*, typically occurring in a syllable-initial position or before a consonant, shows higher F2 and greater F2-F1 distance than the darker, syllable-final variant [5]. The clear and dark /l/s exhibit different sequence and magnitude concerning the tongue tip (TT) and tongue dorsum (TD) gestures [6], [7]; Lowering of F2 for the *dark-L* and resultant smaller F2-F1 value could correspond to the retraction and raising of TD [6], [8].

Previous articulatory studies show that the TD gesture is the key in understanding allophonic variations in laterals. However, due to lack of articulatory data in L2 speech research, we do not know what articulatory properties correspond to lateral allophony, given that L2 speakers could plausibly use different strategies from L1 speakers. In the case of Japanese, previous research demonstrated that Japanese /r/ lacks the specific TD gestural target, based on the extent of variability across vowel contexts, which could also affect the way Japanese speakers articulate syllable-initial and final /l/ [9]. In addition, a great degree of individual variation can be expected in the articulatory domain; for instance, at least seven patterns of tongue postures were discovered in production of English /r/ by native Japanese learners of English with varying proficiency in English, which made it difficult to establish a clear relationship between acoustics and articulation (e.g., lower F3) [10].

The current study examines two predictions: (1) if the articulatory properties of nonnative sounds are influenced by those of the closest native category, given the variability in the TD gesture for Japanese /r/, the posterior tongue is also likely to be highly variable depending on the vowel context in English /l/; (2) changes in articulatory properties may not always be manifested in acoustics, making it difficult to generalise what articulatory strategies would contribute to the changes in F2 and F2-F1 in the case of English /l/. In the following section, I will report a study that investigates Japanese speakers' production of the allophonic variations in English laterals. Specifically, we compare the effects of: 1) syllable position and 2) vowel contexts on acoustics and articulation.

2. Methods

2.1. Participants

Five L1 speakers of Japanese (two female: "JP01F" and "JP04F", and three male: "JP02M", "JP03M", and "JP05M") participated in this study. They were aged between 23 and 30 years ($M = 24.6$ years) and enrolled in a university in the UK for a postgraduate study. All of them defined themselves as native speakers of Tokyo Japanese, and their English proficiency was considered to be high given their enrolment status at the UK university as well as their self-reported assessment. None of them reported speech and hearing impairment at the time of the recording.

2.2. Materials

The stimuli were developed to elicit /l/s occurring in syllable-initial and -final positions in a controlled manner. The list of target words is shown in Table 1. The material development conceptually followed previous studies [11], [12]. First, target words were determined such that words embedding both onset and coda tokens of /l/ share same sets of phonemes, differing

only in their sequences. For example, we used *leave* /li:v/ for a syllable-initial lateral whereas *veal* /vi:l/ for a syllable-final lateral, with flanking vowels being either /i/ or /a/.

Second, each token was flanked by another word before or after the target token depending on the syllable position of /l/s to control the vowel environment surrounding /l/. These words were always in /hVp/ structure, with the vowels either /i/ or /a/. The consonants /h/ and /p/ were chosen to minimise consonantal effects on tongue movement while enabling us maximally to discern the lateral tokens from the neighbouring sounds in the acoustic signal.

As a result, we developed a list of 32 phrases to elicit initial and final /l/s (16 tokens x 2 vowel conditions). These phrases were then embedded in a carrier phrase ‘(Someone) said “XY” to (someone’s) boss.’, in which XY corresponds to the two-word phrase that contains a target word. The participants recorded each sentence at least three times, yielding at least 96 tokens per speaker.

Table 1: List of stimuli and example phrases. (# indicates syllable boundary.)

Vowel	Initial	Final	Example phrase
high	leap	peal	heap leap (i#li), peal heap (il#i)
	lead	deal	heap lead (i#li), deal heap (il#i)
	lean	kneel	hap lean (a#li), kneel hap (il#a)
	leave	veal	hap leave (a#li), veal hap (il#a)
low	lap	pal	heap lap (i#la), pal heap (al#i)
	lag	gal	heap lag (i#la), gal heap (al#i)
	lab	bal	hap lab (a#la), bal hap (al#a)
	lack	Cal	hap lack (a#la), Cal hap (al#a)

2.3. Data collection

The recording took place at the Phonetics Lab at Lancaster University, UK. Ultrasound data were obtained using a Telemed MicrUs system, with a 64 element probe of 20 mm radius. Midsagittal tongue views were imaged with a 2MHz probe frequency, 80 mm depth, 74.5% field of view and 52 scan lines, resulting in a framerate of ca. 100 per second. Simultaneous acoustic signals were also collected with the signal pre-amplified and digitized using a TASCAM US 4x4 audio interface, and then recorded onto a laptop computer at 44.1 kHz with 16-bit quantisation. The research project has been approved by the Lancaster University ethics committee.

2.4. Data analysis

Audio recording and tongue spline data were exported from the Articulate Assistant Advanced software [13] for analysis. Segmentation was carried out by the author with Praat [14] onto the accompanying TextGrid files for each audio recording. The audio files were low-pass filtered to 11,025 Hz, down-sampled to 22,050 Hz, and scaled for its peak intensity of 0.8. Following previous research, the lateral portion was labelled based on a steady-state of F2, which therefore excluded formant transitions in and out of the flanking vowels [15]. Specifically, the lateral was identified based on the following multiple cues: 1) an abrupt change in F2, 2) decrease in resonance energy on spectrograph and 3) decrease in waveform amplitude [15], [16].

For the acoustic analysis, F1 to F3 were estimated and extracted at the 11 equal intervals throughout the lateral portion using Fast Track, an optimized formant analysis Praat plug-in [17]. Fast Track automatically adjusts optimal ceiling formant

values within a range of 4,700 and 7,550 Hz and returns a ‘winning’ analysis for each token based on 24 regression analyses. In this paper, I report the analysis of F2-F1 measure in 500 tokens (248 initial and 252 final). F2-F1 is often used in the previous research to give an approximate estimation of lateral quality [4], [6], [15], [16]; A higher F2-F1 corresponds to a clearer variant of English lateral. The F2-F1 values were then normalised by speaker so that it better captures within-speaker contrast in realisations of English /l/. Data were visualised through *ggplot* functions in the tidyverse suite [18].

For articulatory analysis, based on the TextGrid file that delimited the lateral portion on the acoustic signal, tongue splines were automatically fitted to the ultrasound tongue images through the best-fit batch processing function in the AAA software. In the current study, I focus on the tongue spline data at the lateral midpoint, given that the tongue shape at midpoint is shown to well-represent the articulation of /l/ in British English [7]. The analysis focused on the 497 tokens (252 initial and 245 final), as a result of some tokens having to be excluded due to errors in audio-ultrasound synchronization.

2.5. Statistical analysis

2.5.1. Acoustic data

The F2-F1 values were entered to Linear Mixed-Effect modelling (LME) to evaluate the effects syllabic position and vowel context using *lmer* package [18] on R [20]. For significance testing, the full model was tested against models that did not contain factors of interest through likelihood ratio testing [21]. This enables us to formally assess whether presence or absence of factors could improve the model fit. As a result, a maximal model was constructed for the acoustic data with z-scored F2-F1 values ($z.F2F1$) as the outcome variable and the syllabic position (i.e., initial vs final: ‘*position*’) and the vowel environment (i.e., /a a/, /a i/, /i a/, and /i i/ ‘*vowel*’) as the fixed effects. Also, the by-speaker random intercepts were included for *position* and *vowel*, whereas the random intercept for words was not included as it did not improve the model fit. Finally, the interaction between position and vowel context was included as it improved the model fit significantly. The final model specification was determined as: $z.F2F1 \sim position + vowel + (1 + position + vowel | speaker) + position:vowel$.

2.5.2. Articulatory data

For the articulatory analysis, tongue splines were fitted with generalised additive mixed models (GAMMs) using the *mgcv* and *itsadug* packages on R [22], [23]. The exported tongue spline data from the AAA software consisted of X and Y Cartesian coordinates based on the pixel intensity differences in the ultrasound tongue images. The X and Y values were then converted into polar coordinates using the *rticulate* package [24]. Note that data visualisation is based on unrotated tongue spline data, meaning that x-axis is not parallel to the occlusal plane. Therefore, I only focus on within-subject differences in the shape of tongue splines.

3. Results

3.1. Acoustic analysis

In Table 2, mean and standard deviation for F1, F2, and F2-F1 frequency values by position (initial vs final) are summarised across speakers. Overall, F1 was higher syllable-finally than syllable-initially, except for the speaker JP04F. F2 was

consistently lower for the syllable-final laterals than for the syllable-initial ones. Consequently, higher F2-F1 values were seen for syllable-initial /l/ than for syllable-final /l/. These results suggest that the Japanese speakers in this study mostly conform with the native-like allophonic variation pattern between syllable-initial and syllable-final /l/s acoustically.

Table 2: Mean and SD (in bracket) of F1, F2 and F2-F1 values (Hz) at the lateral midpoint.

Speaker	Position	F1	F2	F2-F1
JP01F	Initial	360.86 (68.56)	1850.17 (190.81)	1489.32 (226.87)
	Final	594.48 (86.25)	1134.23 (50.70)	539.75 (73.37)
JP02M	Initial	447.84 (95.75)	1290.76 (255.35)	842.92 (332.02)
	Final	504.60 (55.32)	882.45 (78.98)	377.85 (109.74)
JP03M	Initial	377.52 (48.30)	1717.50 (240.25)	1339.98 (267.79)
	Final	472.61 (85.98)	1685.55 (301.24)	1212.94 (362.94)
JP04F	Initial	471.87 (75.15)	1361.04 (193.07)	889.18 (217.92)
	Final	461.89 (97.85)	1059.48 (118.93)	597.60 (88.71)
JP05M	Initial	383.49 (57.25)	1394.87 (129.18)	1011.38 (143.97)
	Final	497.24 (58.98)	1161.98 (124.55)	664.74 (163.64)

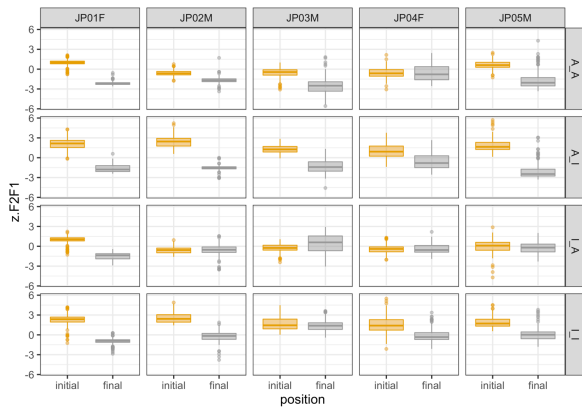


Figure 1 Comparisons of F2-F1 values (z-scored) for initial (in orange) and final (in grey) for different vowel positions across five speakers.

Figure 1 summarises the F2-F1 values extracted from the lateral midpoint by position (initial vs final) across four vowel environments across five speakers. Speakers JP01F, JP02M and JP05M made a contrast between initial and final /l/s in most of the vowel environments; While JP01F was consistent in making a clear-dark contrast in all vowel environments, such contrast was levelled in the I_A context for JP02M and JP05M. JP04F differentiated the two position-dependent variants in the A_I and I_A conditions but less so in the A_A and I_I contexts. JP03M differentiated the two allophones in the A_A and A_I contexts but to a lesser extent in the I_A and I_I environments. Finally, comparisons of several linear mixed-effect models

demonstrated that there were significant effects of *position* ($\chi^2(1) = 8.801, p = .003$), *vowel* ($\chi^2(3) = 10.727, p = .013$), and the interaction between *position* and *vowel* ($\chi^2(3) = 171.800, p < .001$).

In summary, the acoustic analysis based on the F2-F1 values suggests that Japanese speakers distinguish syllable-initial and syllable-final laterals, but the magnitude of this varies by vowel context. Particularly, visual inspections of the boxplots suggest that there is difference in the initial-final contrast depending on the preceding vowel (/i/ vs /a/). We will then turn to the articulatory analysis to examine whether similar trends can be observed in the articulatory data.

3.2. Articulatory analysis

The participants' tongue spline data are visualised by speaker in Figure 2. Based on the results from the acoustic analysis, the four vowel conditions have been collapsed into two categories for the preceding vowels /i/ vs /a/. GAMM models based on the polar coordinate were constructed using the *polar_gam* function in the *rticulate* package separately for each speaker [25] with the tongue height (Y) as the outcome variable and separate smoothing spline terms of the X coordinate by position and vowel. Residual autocorrelations were reduced by specifying AR1 models for each speaker as the amount of autocorrelation at lag 1 [26]. Significance testing was conducted through model comparisons using the *compareML* function between the full model ($Y \sim position + vowel + s(X, by = position) + s(X, by = vowel)$) and a model that did not contain a parametric term and a smooth term with the variable in by-parameter (Position model: $Y \sim position + s(X, by = position)$; Vowel-context model: $Y \sim vowel + s(X, by = vowel)$).

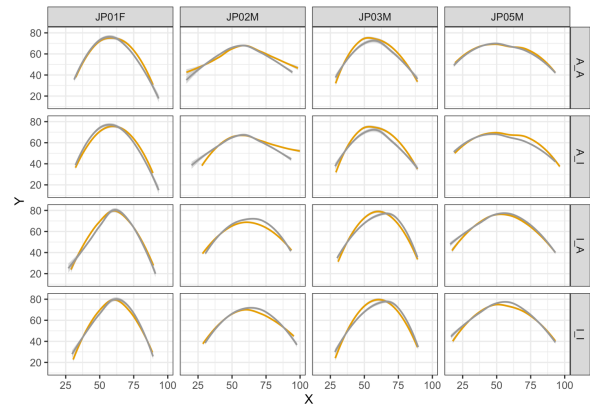


Figure 2 Comparisons of tongue splines for initial (in orange) and final (in grey) for different vowel positions across four speakers with tongue tip pointing rightward (based on the Cartesian coordinates). A speaker JP04F was excluded due to the poor imaging quality.

Overall, the full models significantly improved the model-fit against the position-only models (i.e., without terms associated with vowel context) across speakers (JP01F; $\chi^2(5.00) = 3589.510, p < .001$, JP02M; $\chi^2(5.00) = 500.325, p < .001$, JP03M; $\chi^2(5.00) = 522.015, p < .001$, JP05M; $\chi^2(5.00) = 231.851, p < .001$) suggesting that vowel context significantly improved the model fit. On the other hand, there were no significant differences between the full models and the vowel-context models (i.e., without terms associated with position),

with only a slight decrease in the AIC values, suggesting that the positional effect did not improve the model fit. In summary, the GAMM models show that the vowel context had a greater effect on the participants' tongue shapes for English /l/ than did syllable position.

In addition, the difference smooths plots show that the way the speakers differentiated the initial-final laterals differed from one another; Whereas JP05M differentiated them mainly at the middle of the tongue, JP03M did so on wider parts spanning tongue anterior and posterior when the vowel /i/ preceded the lateral /l/ (see Figure 3).

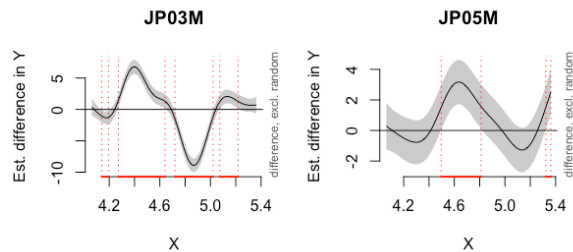


Figure 3 *Difference smooths between syllable-initial and -final /l/s that follow the vowel /i/ produced by JP03M (left) and JP05M (right).*

4. Discussion

The current study investigated the production of English laterals by Japanese speakers. The main finding of this study is that Japanese speakers contrast initial and final English /l/ in acoustics, but there is not strong support for this in midsagittal tongue shape. Instead, tongue shape appears to be more strongly affected by vowel context.

The acoustic analysis has shown that the participants were generally able to distinguish syllable-initial and -final /l/ in an expected native-like manner. The participants produced syllable-initial lateral with lower F1 and higher F2 as shown in Table 2, which generally agrees with documentation in the previous literature [5], [8]. While the interaction between *position* and *vowel* is not surprising, there were some between-speaker differences in the degree of vowel coarticulation. It is known that clear and dark /l/s differ in the degree of vowel coarticulation, in which dark /l/ is coarticulated with the neighbouring vowels to a lesser degree than clear /l/ [5]. This tendency was true of one of the speakers, JP01F, who produced syllable-initial /l/s with higher F2-F1 than syllable-final /l/ consistently. In contrast, the other two speakers (JP03M and JP04F) showed little contrast between the initial and final tokens across the four vowel contexts, with the case of the /l/ preceded by /i/ showing lack of the initial-final distinction compared to when /l/ followed /a/.

While the distinction between the initial-final contrast was seen relatively clearly in the acoustic results, such tendency was not observed in the articulatory results, suggesting that static midsagittal tongue shape may not capture the most articulatory salient dimension of this contrast [27]. In the articulatory domain, the preceding vowels instead have a greater effect on the tongue shape for English laterals. Despite smaller degrees of initial-final contrasts in the acoustic data when /i/ preceded /l/, JP03M's tongue shapes were different throughout the anterior-posterior dimensions of the tongue, which was also

somewhat evident from the visual inspection of Figure 2. This could suggest that it was not the TD gesture that JP03M uses to differentiate initial and final /l/. As well as in the case of English /r/ [10], the articulation of English /l/ in L2 speech may also be diverse across speakers, making it difficult to pin down a single articulatory property (e.g., TD gesture) that is crucial in the initial-final contrast in English /l/.

One possible explanation for this mismatch between acoustics and articulation may be the dynamic nature of acoustics and articulatory realisations of English liquids over time [28], [29]. While the mid-point analysis can adequately represent the broad phonetic quality of English /l/, dynamic analyses could provide a greater amount of information, particularly regarding co-articulatory influence of the flanking vowels [28]. Previous research finds that the articulatory gesture is often present before changes in acoustic signals were observed [29]. It could also be true that the participants might utilise several articulatory strategies to achieve a certain acoustic target (e.g., along the F2-F1 dimension), which was implicated in an earlier study of English /r/ [10]. In future research, auditory analysis could be incorporated to investigate what acoustic/articulatory properties are important for English laterals, which may not necessarily be manifested in the acoustics or articulatory data at the temporal mid-point.

The current focus on the midsagittal tongue shapes may have reduced the dimensionality of the complex articulatory properties involved in English /l/, particularly regarding tongue lateralisation that could only be captured on the tongue's coronal plane. Tongue lateralisation has been proposed to be an active articulatory gesture in English /l/ [30]. However, the transfer of the lateral gesture does not always happen from L1 to L2, such that a Japanese speaker who lateralises the tongue in L1 does not always succeed in producing English /l/ in a native-like manner [31]. In investigating tongue lateralization, obtaining coronal tongue data would be helpful.

Finally, an anonymous reviewer pointed that it would be helpful to include data collected from L1 English speakers. Whereas the current study makes reference to the English variety that exhibits the clear-dark allophony, such as Standard Southern British English (SSBE), the degree of the clear-dark contrast and the degree of 'darkness' are known to vary from one accent to another [7], [15]. The current results would be more enhanced if L1 English data were to be included.

5. Conclusions

This study investigated the effects of the syllabic position and vowel environment on the acoustics and articulation of English laterals produced by Japanese speakers. The results demonstrated that the picture may be more complicated than a mere syllable-initial and -final distinction, particularly in the articulatory dimension. The current study also adds novel evidence into the articulatory properties in L2 speech, in which L2 speakers' articulatory strategies may be diverse.

6. Acknowledgements

I thank all the participants who gave up their time for the data collection. I thank Dr Sam Kirkham and Dr Claire Nance for helpful comments and support throughout the research project. The work has improved thanks to comments from three anonymous reviewers. This research is financially supported by Graduate Scholarship for Degree Seeking Students, Japan Student Services Organization (JASSO) awarded to the author.

7. References


- [1] J. E. Flege, ‘Second Language Speech Learning Theory, Findings and Problems’, in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, W. Strange, Ed. York Press, 1995.
- [2] C. T. Best and M. D. Tyler, ‘Nonnative and second-language speech perception: Commonalities and complementarities’, in *Language experience in second language speech learning: In honor of James Emil Flege*, O.-S. Bohn and M. J. Munro, Eds. Amsterdam: John Benjamins Publishing Company, 2007, pp. 13–34. doi: 10.1075/llt.17.07bes.
- [3] J. E. Flege and O.-S. Bohn, ‘The Revised Speech Learning Model (SLM-r)’, in *Second Language Speech Learning: Theoretical and Empirical Progress*, 1st ed., R. Wayland, Ed. Cambridge University Press, 2021, pp. 3–83. doi: 10.1017/9781108886901.002.
- [4] P. Carter and J. Local, ‘F2 variation in Newcastle and Leeds English liquid systems’, *Journal of the International Phonetic Association*, vol. 37, no. 2, pp. 183–199, Aug. 2007, doi: 10.1017/S0025100307002939.
- [5] D. Recasens, ‘A cross-language acoustic study of initial and final allophones of /l/’, *Speech Communication*, vol. 54, no. 3, pp. 368–383, Mar. 2012, doi: 10.1016/j.specom.2011.10.001.
- [6] R. Sproat and O. Fujimura, ‘Allophonic variation in English /l/ and its implications for phonetic implementation’, *Journal of Phonetics*, vol. 21, no. 3, pp. 291–311, Jul. 1993, doi: 10.1016/S0095-4470(19)31340-3.
- [7] D. Turton, ‘Categorical or gradient? An ultrasound investigation of /l/-darkening and vocalization in varieties of English’, *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 8, no. 1, p. 13, May 2017, doi: 10.5334/labphon.35.
- [8] S. S. Narayanan, A. A. Alwan, and K. Haker, ‘Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals’, *The Journal of the Acoustical Society of America*, vol. 101, no. 2, pp. 1064–1077, Feb. 1997, doi: 10.1121/1.418030.
- [9] N. Yamane and P. Howson, ‘An ultrasound examination of taps in Japanese’, in *Proceedings of the 18th International Congress of Phonetic Sciences*, Aug. 2015, p. 5. [Online]. Available: <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0815.pdf>
- [10] J. Moore, J. Shaw, S. Kawahara, and T. Arai, ‘Articulation strategies for English liquids used by Japanese speakers’, *Acoust. Sci. & Tech.*, vol. 39, no. 2, pp. 75–83, Mar. 2018, doi: 10.1250/ast.39.75.
- [11] B. Gick, F. Campbell, S. Oh, and L. Tamburri-Watt, ‘Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids’, *Journal of Phonetics*, vol. 34, no. 1, pp. 49–72, Jan. 2006, doi: 10.1016/j.wocn.2005.03.005.
- [12] F. Campbell, B. Gick, I. Wilson, and E. Vatikiotis-Bateson, ‘Spatial and Temporal Properties of Gestures in North American English /r/’, *Lang Speech*, vol. 53, no. 1, pp. 49–69, Mar. 2010, doi: 10.1177/0023830909351209.
- [13] Articulate Instruments, *Articulate Assistant Advanced version 2.18*. Edinburgh: Articulate Instruments, 2019.
- [14] P. Boersma and D. Weenink, *Praat: doing Phonetics by Computer*. 2022. Accessed: Feb. 21, 2022. [Online]. Available: <https://www.fon.hum.uva.nl/praat/>
- [15] S. Kirkham, D. Turton, and A. Leemann, ‘A typology of laterals in twelve English dialects’, *The Journal of the Acoustical Society of America*, vol. 148, no. 1, pp. EL72–EL76, Jul. 2020, doi: 10.1121/10.0001587.
- [16] C. Nance, ‘Phonetic variation in Scottish Gaelic laterals’, *Journal of Phonetics*, vol. 47, pp. 1–17, Nov. 2014, doi: 10.1016/j.wocn.2014.07.005.
- [17] S. Barreda, ‘Fast Track: fast (nearly) automatic formant-tracking using Praat’, *Linguistics Vanguard*, vol. 7, no. 1, p. 20200051, Jan. 2021, doi: 10.1515/lingvan-2020-0051.
- [18] H. Wickham *et al.*, ‘Welcome to the Tidyverse’, *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, Nov. 2019, doi: 10.21105/joss.01686.
- [19] D. Bates, M. Mächler, B. Bolker, and S. Walker, ‘Fitting Linear Mixed-Effects Models Using lme4’, *Journal of Statistical Software*, vol. 67, pp. 1–48, Oct. 2015, doi: 10.18637/jss.v067.i01.
- [20] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2021. [Online]. Available: <https://www.R-project.org/>
- [21] B. Winter, ‘Statistics for Linguists: An Introduction Using R’, *Routledge & CRC Press*, 2020. <https://www.routledge.com/Statistics-for-Linguists-An-Introduction-Using-R/Winter/p/book/9781138056091> (accessed Mar. 19, 2022).
- [22] S. N. Wood, *Generalized Additive Models: An Introduction with R*, 2nd ed. New York: Chapman and Hall/CRC, 2017, doi: 10.1201/9781315370279.
- [23] J. van Rij, M. Wieling, R. H. Baayen, and H. van Rijn, *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*. 2020. Accessed: Mar. 20, 2022. [Online]. Available: <https://CRAN.R-project.org/package=itsadug>
- [24] S. Coretta, *rticulate: Ultrasound Tongue Imaging in R*. 2021. Accessed: Mar. 20, 2022. [Online]. Available: <https://github.com/stefanocoretta/rticulate>
- [25] S. Coretta, ‘Vowel duration and consonant voicing: A production study’, PhD Thesis, University of Manchester, Manchester, 2020. Accessed: Mar. 20, 2022. [Online]. Available: <https://stefanocoretta.github.io/phd-thesis/index.html>
- [26] M. Wieling, ‘Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English’, *Journal of Phonetics*, vol. 70, pp. 86–116, Sep. 2018, doi: 10.1016/j.wocn.2018.03.002.
- [27] D. Turton, ‘Variation in English /l/: Synchronic reflections of the life cycle of phonological processes’, PhD Thesis, University of Manchester, Manchester, 2014. Accessed: Mar. 20, 2022. [Online]. Available: [https://www.research.manchester.ac.uk/portal/en/theses/variation-in-english-l-synchronic-reflections-of-the-life-cycle-of-phonological-processes\(dfa11693-a112-45e2-99f6-2a08cf5f117b\).html](https://www.research.manchester.ac.uk/portal/en/theses/variation-in-english-l-synchronic-reflections-of-the-life-cycle-of-phonological-processes(dfa11693-a112-45e2-99f6-2a08cf5f117b).html)
- [28] S. Kirkham, C. Nance, B. Littlewood, K. Lightfoot, and E. Groarke, ‘Dialect variation in formant dynamics: The acoustics of lateral and vowel sequences in Manchester and Liverpool English’, *The Journal of the Acoustical Society of America*, vol. 145, no. 2, pp. 784–794, Feb. 2019, doi: 10.1121/1.5089886.
- [29] J. Ying, J. A. Shaw, C. Kroos, and C. T. Best, ‘Relations Between Acoustic and Articulatory Measurements of /l/’, in *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, Sydney, Dec. 2012, pp. 109–112.
- [30] J. Ying, J. A. Shaw, C. Carignan, M. Proctor, D. Derrick, and C. T. Best, ‘Evidence for active control of tongue lateralization in Australian English /l/’, *Journal of Phonetics*, vol. 86, p. 101039, May 2021, doi: 10.1016/j.wocn.2021.101039.
- [31] M. Morimoto, ‘Articulatory Preference in Japanese Liquids and F3 in English: A Preliminary Report’, *ICU Working Papers in Linguistics: Selected Papers from the 5th Asian Junior Linguists Conference (AJL5)*, vol. 15, pp. 1–6, Mar. 2021.

Chapter 7

Study 1: Formant dynamics in second language speech: Japanese speakers' production of English liquids

The pilot study presented in Chapter 5 demonstrates that L1 Japanese speakers may use tongue dorsum differently from L1 English speakers when producing L2 English liquids. Extending this, this chapter investigates differences in time-varying changes in spectral properties (formant dynamics) in production of word-initial liquid-vowel sequences between L1 Japanese and L1 English speakers. This study demonstrates that L1 Japanese speakers' production is more variable depending on the vowel contexts than that of L1 English speakers, providing further grounding to the studies presented in the subsequent chapter. This study has been published from the *Journal of Acoustical Society of America* as of the 22nd of January, 2024.

Formant dynamics in second language speech: Japanese speakers' production of English liquids

Takayuki Nagamine^{a)} 

Department of Linguistics and English Language, County South, Lancaster University, Lancaster, LA1 4YL, United Kingdom

ABSTRACT:

This article reports an acoustic study analysing the time-varying spectral properties of word-initial English liquids produced by 31 first-language (L1) Japanese and 14 L1 English speakers. While it is widely accepted that L1 Japanese speakers have difficulty in producing English /l/ and /ɹ/, the temporal characteristics of L2 English liquids are not well-understood, even in light of previous findings that English liquids show dynamic properties. In this study, the distance between the first and second formants (F_2-F_1) and the third formant (F_3) are analysed dynamically over liquid-vowel intervals in three vowel contexts using generalised additive mixed models (GAMMs). The results demonstrate that L1 Japanese speakers produce word-initial English liquids with stronger vocalic coarticulation than L1 English speakers. L1 Japanese speakers may have difficulty in dissociating F_2-F_1 between the liquid and the vowel to a varying degree, depending on the vowel context, which could be related to perceptual factors. This article shows that dynamic information uncovers specific challenges that L1 Japanese speakers have in producing L2 English liquids accurately. © 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.1121/10.0024351>

(Received 10 May 2023; revised 4 December 2023; accepted 5 December 2023; published online 22 January 2024)

[Editor: Susanne Fuchs]

Pages: 479–495

I. INTRODUCTION

A. Acquisition of English /l/ and /ɹ/ by L1 Japanese speakers

The current study investigates time-varying spectral properties of English liquids produced by first-language (L1) Japanese speakers. Numerous studies have shown that the acquisition of English liquids is particularly challenging for L1 Japanese speakers (e.g., Aoyama *et al.*, 2019; Best and Strange, 1992; Flege *et al.*, 1995; Saito and Munro, 2014; Sheldon and Strange, 1982). They typically perceive English /l/ and /ɹ/ as instances of a single L1 category of Japanese /r/ (e.g., Best and Strange, 1992; Guion *et al.*, 2000). This corresponds to the learning of “similar” phones between L1 and L2 in the Speech Learning model (SLM) (Flege, 1995; Flege and Bohn, 2021) and the single-category (SC) or the category-goodness (CG) assimilation scenarios in the Perceptual Assimilation model of Second Language (L2) Speech Learning (PAM-L2) (Best and Strange, 1992; Best and Tyler, 2007; Hattori and Iverson, 2009), predicting a moderate to substantial difficulty in acquisition of the L2 sounds. SLM posits that perceptual accuracy lays the foundation for accurate L2 speech production because L2 learners develop articulatory rules in the L2 phonetic categories that are established over the course of L2 speech learning (Flege and Bohn, 2021).

The difficulty L1 Japanese speakers face in acquiring English /l/ and /ɹ/ is associated with their sensitivity to the

phonetic cues used to distinguish the contrast. The key spectral dimension that contrasts English /l/ and /ɹ/ is the frequency of the third formant (F_3); American English /ɹ/ is associated with a notably low F_3 at 1300 Hz for male speakers and 1800 Hz for female speakers whereas laterals show a high F_3 at approximately 2500–2800 Hz (Espy-Wilson, 1992; Stevens, 2000). The F_2 frequency is associated with the resonance of the vocal tract cavity posterior to the primary constriction for both laterals and rhotics, which are commonly produced with a backed tongue body configuration (Stevens, 2000). Laterals are generally characterised by clear-dark allophony according to syllabic position; “clear” /l/s are often associated with laterals in pre-vocalic, syllable-initial position, and they typically have higher F_2 values and a greater separation between F_2 and F_1 (F_2-F_1) than the post-vocalic “dark” counterpart (Carter and Local, 2007; Recasens, 2012). American English exhibits relatively darker realisations of liquids than British English overall, but syllable-initial laterals in American English are still somewhat “clearer” than syllable-final counterparts (Recasens, 2012). This clear-dark allophony according to the syllable position results from different articulatory configurations, such that the degree of the tongue body retraction is greater for the final laterals than for the initial laterals (Recasens, 2012).

L1 Japanese speakers tend to rely on the less reliable cue of F_2 in their perception of English /l ɹ/ than a more robust cue of F_3 (Iverson *et al.*, 2003; Saito and Munro, 2014). As a result, they tend to produce the distinction along the F_2 dimension instead of learning to make a contrast

^{a)}Email: t.nagamine@lancaster.ac.uk

along F_3 (Aoyama *et al.*, 2019; Saito and van Poeteren, 2018). For instance, they produce word-initial English /l/ with a somewhat higher F_2 (approximately 1500–1800 Hz) than L1 English speakers (approximately 1200–1500 Hz), whereas F_2 frequencies for English /ɹ/ are similar between the two speaker populations (Aoyama *et al.*, 2019; Flege *et al.*, 1995). As for F_3 , they produce English /ɹ/ with a relatively high F_3 (2000–2600 Hz) but produce /l/ with F_3 values comparable to L1 English speakers (Aoyama *et al.*, 2019; Flege *et al.*, 1995; Saito and Munro, 2014). Nevertheless, previous research claims that L1 Japanese speakers could learn to use the acoustic cues as L1 English speakers would do, especially for F_1 and F_2 ; several studies reported similar F_1 values in production of English liquids between L1 Japanese and L1 English speakers (Aoyama *et al.*, 2019; Flege *et al.*, 1995; Saito and Munro, 2014). Saito and Munro (2014) also argue that the use of F_2 is easier for L1 Japanese speakers to acquire than that of F_3 for English /ɹ/ based on findings that L1 Japanese speakers who resided in Canada for longer than 2.5 months produced native-like F_2 values for English /ɹ/ compared to those who had less overseas experience.

The degree of difficulty in L1 Japanese speakers' acquisition of English liquids also varies depending on the vowel context, in which they are better at correctly identifying word-initial English liquids adjacent to front vowels compared to back vowels in perception (Shimizu and Dantsuji, 1983). This might be because L1 Japanese speakers may also perceive English /l/ and /ɹ/ as a sequence of a back vowel and a tap (i.e., [tʌr]), possibly due to the vocalic nature of English liquids (Guion *et al.*, 2000). L1 Japanese speakers are more likely to hear a /w/-like percept when perceiving English /l/ and /ɹ/ than L1 English speakers (Best and Strange, 1992; Mochizuki, 1981; Yamada and Tohkura, 1992). These results overall suggest that L1 Japanese speakers are sensitive not only to the phonemic status but also phonetic details of English /l/ and /ɹ/. In particular, Shimizu and Dantsuji (1983) speculate that coarticulatory properties may play a role in explaining the vocalic contextual effects in L1 Japanese speakers' correct identification of English /l/ and /ɹ/.

B. Dynamic analysis of English liquids

Although the errors in segmental realisation in L2 speech are claimed to be rooted in perception, accurate perception does not always entail accurate production (Flege and Bohn, 2021; Sheldon and Strange, 1982). While this does not mean that the role of perceptual accuracy should be discounted, it implies that L2 speech production may be shaped by a combination of factors in addition to perceptual accuracy.

One such possible factor includes the dynamic nature involved in the production of English liquids. Articulation of English liquids requires coordination of multiple articulatory gestures for accurate production (Campbell *et al.*, 2010; Sproat and Fujimura, 1993). English laterals, for instance,

involve coordination of tongue tip and dorsum gestures, and the timing and magnitude interact with the syllabic position; a tongue tip gesture precedes a tongue dorsum gesture with a greater magnitude for clear /l/ whereas the two gestures could be timed synchronously for the dark /l/ (Sproat and Fujimura, 1993). English rhotics show similar patterning of gestural timing and magnitude, where labial gestures precede the tongue tip and tongue body gestures (Campbell *et al.*, 2010; Proctor *et al.*, 2019). The dynamic nature of articulation in English liquids suggests that the acoustic characteristics of English liquids are inherently non-static, and it is, therefore, often challenging to select a single point in time that adequately represents liquid quality (Kirkham *et al.*, 2019; Ying *et al.*, 2012).

In addition, acoustic realisations of liquids interact with the neighbouring segments as a result of coarticulation. While coarticulation is often viewed as a consequence of the physiological mechanisms in the transition between segmental targets, some aspects of coarticulation may be language-specific and thus need to be learned (Beristain, 2022; Keating, 1985). Word-initial /ɹ/ in English, for instance, shows lower F_3 values when followed by back vowels compared to other vowel conditions (King and Ferragne, 2020). Similarly, vowel context influences realisations of American English /l/, particularly among word-initial /l/s, such that F_2 values are higher in the /i/ context than in the /a/ context (Recasens, 2012). Coarticulatory effects of liquids could span longer term than the domain of liquid segment itself and provide perceptual basis for listeners to distinguish English /l/ and /ɹ/ (West, 1999a,b).

The findings regarding the dynamic nature of liquid production and liquid-vowel coarticulation may account for the specific difficulties that L1 Japanese speakers have in producing English /l/ and /ɹ/. L1 Japanese speakers tend to substitute English /l/ and /ɹ/ with an alveolar tap or flap [ɾ], a canonical realisation of Japanese /ɾ/ (Riney *et al.*, 2000). Previous articulatory studies show that alveolar taps/flaps show stronger coarticulatory effects with the neighbouring vowels than English laterals and rhotics; while the tongue dorsum gesture is actively involved in the production of English /l/ and /ɹ/, taps and flaps [ɾ] show either less involvement of the tongue dorsum or a “stabilization” tongue dorsum gesture, resulting in stronger coarticulation with the vowel (Morimoto, 2020; Proctor, 2011; Recasens, 1991; Yamane *et al.*, 2015). Furthermore, an x-ray study suggests that L1 Japanese speakers' articulation of English liquids shows greater variability according to the vocalic environment (Zimmermann *et al.*, 1984). In sum, Japanese and English liquids differ in the way they are coarticulated with the vowels, and it can be predicted that L1 Japanese speakers exhibit different liquid-vowel coarticulatory patterns from that of L1 English speakers.

Despite the findings regarding the complexity involved in the production of English liquids, our understanding remains relatively limited regarding the specific mechanism whereby L1 Japanese speakers struggle to produce English /l/ and /ɹ/. This may be because previous research

commonly evaluates liquid quality based on a single-point measurement, in which formant frequencies are measured at one point in time, such as the F_3 minima, the spectral onset, or the spectral release (Aoyama *et al.*, 2019; Flege *et al.*, 1995; Saito and Munro, 2014). Analysis of liquids based on a single measurement, however, inevitably averages out temporal information that may be important for understanding the dynamic characteristics of English liquids.

In the current study, I show that dynamic formant measurement of English liquids allows us to better understand specific challenges that L1 Japanese speakers have in producing English /l/ and /ɹ/. Previous research suggests that (1) L1 Japanese speakers' acquisition of English liquids may be influenced by the phonetic details, such as vowel environments, and (2) English liquids show dynamic characteristics and interactions with the neighbouring vowels. Given these, I hypothesise that L1 Japanese speakers' production of English liquids will exhibit different dynamic acoustical properties compared to L1 English speakers. This study therefore asks what dynamic acoustic properties L1 Japanese speakers would show in their production of English /l ɹ/ compared to L1 English speakers.

I combine static and dynamic analyses of the acoustic properties of English liquids in this study. The static analysis investigates the distance between second and first formants (F_2-F_1) and the third formant (F_3) extracted at the liquid midpoint. The inclusion of this measure allows me to discuss the results in light of previous research in which the single-measurement analysis has been widely used (e.g., Aoyama *et al.*, 2019; Flege *et al.*, 1995; Saito and Munro, 2014; Saito and van Poeteren, 2018). In addition, the time-varying changes in the F_2-F_1 and F_3 values will capture the complex nature of liquid acoustics and the coarticulatory interactions between the liquid and the vowel (Howson and Redford, 2021; Kirkham *et al.*, 2019; Sproat and Fujimura, 1993).

II. METHODS

A. Participants

The data for the current study are obtained from 45 speakers: 31 L1 Japanese learners of English (17 female and 14 male) aged between 18 and 22 years [$M = 19.81$ years, standard deviation (SD) = 1.05] and 14 L1 North American L1 English speakers (11 female and three male) aged between 21 and 43 years ($M = 28.93$ years, $SD = 6.08$).

All of the L1 Japanese speakers were undergraduate university students recruited from two universities in Japan, located near the cities of Nagoya and Kobe, respectively. Their profile is considered to be typical for average Japanese university students who study English as a foreign language; all of them studied English primarily through the school curriculum in either or both primary and secondary schools, and continued it at the tertiary level, with a mean length of English study being 9.31 years ($SD = 2.42$). They did not have an extended stay in an English-speaking

country, with the length of overseas experience ranging from none to 4.25 months ($M = 0.77$ months, $SD = 1.35$).

In evaluating L1 Japanese speakers' L2 English proficiency, participants were asked to report their perception on their own oral fluency on a scale of seven, with 1 being "I do not speak English at all." to 7 being "No problems in using English in daily life." This is because there was no common measure available across participants to estimate their English proficiency due to the fact that students have taken different kinds of tests or that first-year students had not yet taken any English language test. Nevertheless, judging from the test scores that some of the participants were able to provide and observations by the researcher who has experience in English language teaching in Japan, their English proficiency is considered to be lower to upper intermediate, which largely agrees with their subjective evaluation of their fluency in English ($M = 3.84$, $SD = 1.10$) (see supplementary material for further details about the participants).¹

The 14 L1 English speakers identify themselves as fluent L1 speakers of North American English who grew up using English until 13 years of age. Five of them are from Canada and nine are from the United States. They resided in the United Kingdom (UK) at the time of recording; six of them were postgraduate students enrolled at a UK university and the rest worked in companies in the UK. Recruitment of L1 North American English speakers reflects the situation that American English tends to be chosen as a pedagogical model in English language teaching in Japan and therefore it is appropriate for L1 Japanese speakers' production to be compared to that of L1 North American English speakers (Setter and Jenkins, 2005).

B. Data collection

The audio recordings analysed in this study are a subset of data collection for a larger study, in which both articulatory and acoustic data were obtained in a simultaneous high-speed ultrasound-audio recording setting. For this reason, the participants wore an ultrasound headset while recording stimuli for the current study. The participants were recorded in a sound-attenuated booth at universities in the UK for L1 North American English speakers and in a quiet room at universities in Japan for L1 Japanese speakers. In recording some of the L1 Japanese speakers, however, there was minor background fan noise because of the Covid-19 restrictions mandating air ventilation at the time of recording. Acoustic signals were pre-amplified, digitized, and recorded onto a laptop computer via a Sound Devices (Reedsburg, WI) USB-Pre2 audio interface at 44.1 kHz with 16 bit quantisation.

The participants were asked to sit in front of the laptop screen and read the stimuli words in isolation that were displayed one by one orthographically using Articulate Assistant Advanced (AAA) (Edinburgh, UK) software version 220.4.1 (Articulate Instruments, 2022). No carrier phrases were used here because (1) the use of carrier phrases would impose additional difficulty on L1 Japanese speakers,

especially those who were less proficient in English, and (2) the experiment had to be as short as possible due to time constraints in the data collection sessions.

In light of the language mode hypothesis (Grosjean, 2008) that the language setting in an experiment can influence the participants' speech perception and possibly production, the recording sessions for the L1 Japanese speakers were structured as follows. The first half of the experiment, including briefing, equipment setup, and recording of the Japanese words (not presented in this paper), was conducted while I was giving instructions in Japanese. Then, I switched the language of instructions to English and the participant engaged in a short English conversation activity. This included a semi-structured dialogue in which I asked five simple questions to the participants (e.g., "What do you study?," "What do you like the best about the university?," etc.) Finally, the Japanese participants recorded the English words while I gave all the instructions in English. While it would have been theoretically desirable to have someone else who was an L1 English speaker lead the data collection session for English words, it was challenging for reasons of time and room availability given that each session for L1 Japanese speakers took up to 90 min.

The recording session with the L1 North American English speakers did not require such considerations because they recorded English words only. All the procedures were, therefore, conducted in English and each session took up to approximately 60 min. The participants were compensated for their time and participation with the amount of 2000 Japanese yen or 15 British pound stirlings in the form of cash or vouchers commensurate with the regulations at each of the recording venues. The research project has been reviewed and approved by the ethics committees at Lancaster University, Kobe Gakuin University, and Meijo University. Informed consent to take part in the study was obtained in written form from all participants.

C. Materials

Word-initial English /l/ and /ɹ/ were elicited from 16 monosyllabic words (eight minimal pairs), followed by a close front /i/, an open front /æ/, or a close back vowel /u/ (see Table I). The coda consonants were restricted to bilabials /p b m/ or labiodentals /f v/ to minimise the anticipatory coarticulatory effects on the word-initial liquids. All the target words were checked using the Longman Pronunciation Dictionary (Wells, 2008) to ensure that they have the intended vowel environment in American English.

TABLE I. Word list per vowel context.

Vowel context	Words		
/i/	leap / reap	leaf / reef	leave / reeve
/æ/	lap / rap	lamb / ram	lamp / ramp
/u/	lube / rube	loom / room	

D. Segmentation and data processing

Prior to segmentation, audio recordings were low-pass filtered at 11 000 Hz and downsampled to 22 050 Hz. Automatic segmentation was carried out at phoneme level with a Montreal Forced Aligner (MFA) version 2.0.6 (McAuliffe *et al.*, 2017). I then inspected the aligned data visually and manually corrected the segmentation using Praat where necessary (Boersma and Weenink, 2022).

I classified the liquid tokens into two broad categories: approximants and non-approximants, based on the spectrographic representations aided by auditory impressions. This decision reflects the consideration that the L1 Japanese speakers' production of liquids might show a wide range of variations due to the allophonic variation of Japanese /r/ and their articulatory strategies for English /l/ and /ɹ/. Realisations for Japanese /r/ include other types of approximants than English liquids, such as the canonical [r], retroflex flap [ɽ], retroflex lateral approximant [ɭ], and a lateral flap [ɺ] (Akamatsu, 1997; Arai, 2013). They may also use a single strategy or produce a reversed realisation for English /l ɹ/. It could be the case, for instance, that they produce a lateral liquid for both English /l ɹ/. It is also possible that they use [l] for English /ɹ/ and [ɺ] for English /l/. Classifications based on these two broad categories: approximants and non-approximants, therefore, guide me to choose an appropriate type of analysis while maximising the chance of capturing diverse acoustic properties in the L1 and L2 English liquids.

Based on these considerations, I first broadly labelled tokens as approximants if the liquid token in question shows a vowel-like formant structure (Ladefoged and Johnson, 2010). The spectral analysis focuses only on the tokens that are classified here as approximants; it thus excludes 281 non-approximants tokens (e.g., taps or flaps [ɾ]) out of a total of 2914 tokens, leaving 2633 tokens for further processing. The spectrographic examples of an approximant and a non-approximant token are shown in Figs. 1 and 2.

Following this, I segmented the liquid approximant tokens based on the primary cues of a steady state or an approximately steady state of the F_2 and an abrupt change in amplitude in the waveform (Lawson *et al.*, 2011). Laterals and rhotics in English involve various stages, including the transition into the liquid, the steady state, and the transition into the following vowel (Carter and Local, 2007). The current study uses the steady-state portion to define the liquid as in previous studies (Flege *et al.*, 1995; Kirkham, 2017). Although the liquid steady-state is an approximation given the various stages involved in the liquid acoustics mentioned above, this issue can be minimised in the dynamic analysis because it shows holistic time-varying trajectories across the liquid and vowel.

E. Acoustic analysis

This study analyses 2306 liquid tokens for mid-point analysis and 2515 liquid-vowel tokens for dynamic analysis. The detailed breakdown is shown in Table II. The current

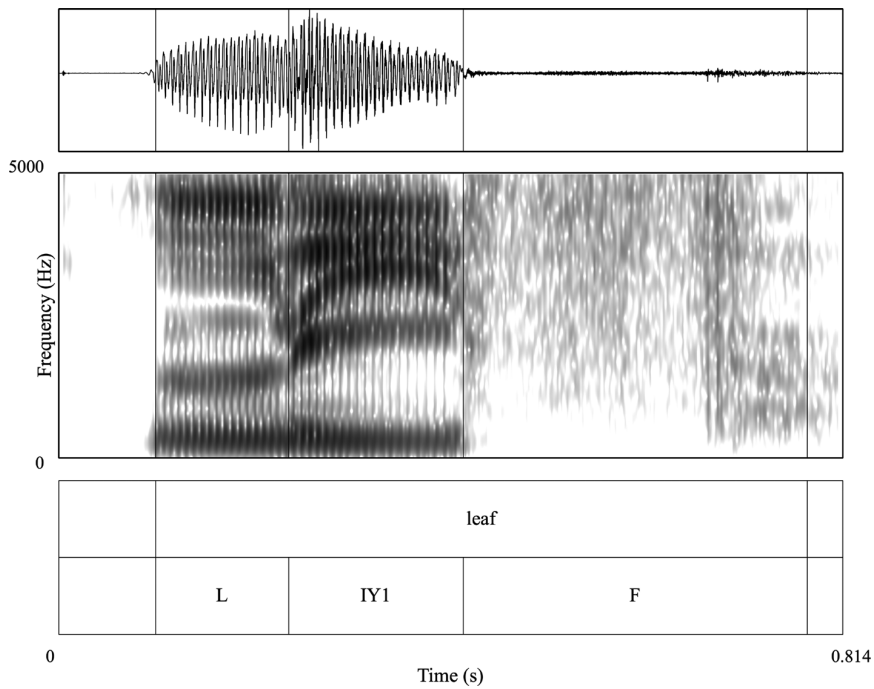


FIG. 1. Example spectrogram of an L1 North American L1 English speaker's production of *leaf*. Labels show phonetic segments in ARPABET, in which "IY1" indicates a stressed high front unrounded vowel /i/.

study compares two acoustic parameters between L1 Japanese and L1 English speakers' production of English liquids: (1) the distance between second (F_2) and first (F_1) formants (F_2-F_1) and (2) the third formant (F_3). F_2-F_1 is used as a measure to evaluate acoustic liquid quality; lower F_2-F_1 values can be related to darker realisations of liquids, resulting from a greater degree of tongue retraction (Howson and Redford, 2021; Sproat and Fujimura, 1993). F_3 is a primary acoustic dimension that distinguishes

English /l/ and /ɫ/, and previous research reports robust differences between L1 Japanese and L1 English speakers' production of English liquids.

F_1 , F_2 , and F_3 values were estimated and extracted with Fast Track, an automatic formant estimation Praat plug-in (Barreda, 2021). Fast Track samples formant frequencies every 2 ms throughout the interval, resulting in smooth trajectories between F_1 and F_3 . It then outputs the estimated formant frequencies while aggregating them in a specified

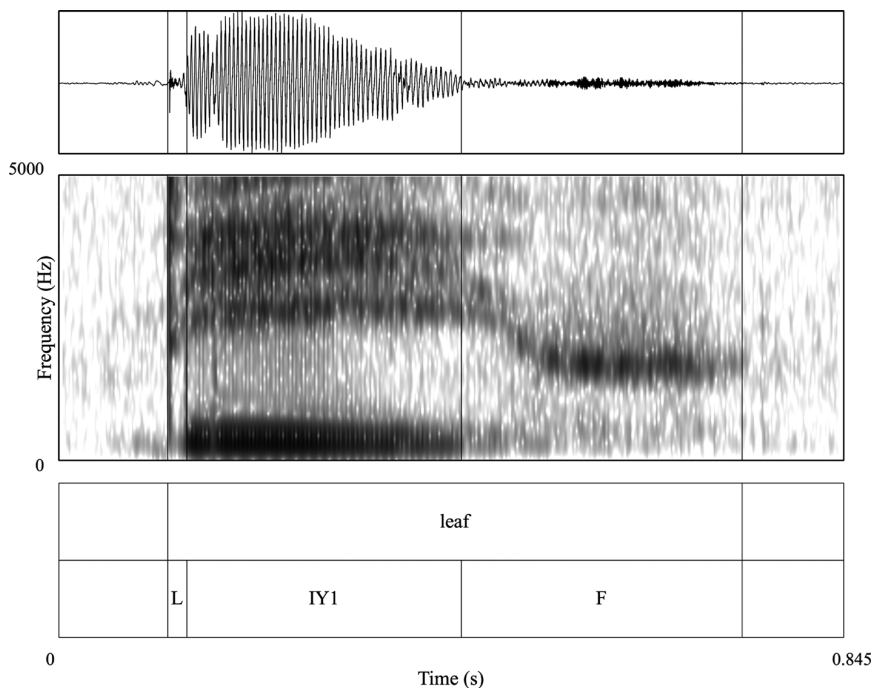


FIG. 2. Example spectrogram of a "definitely a tap" token of *leaf* produced by an L1 Japanese speaker. Labels show phonetic segments in ARPABET, in which "IY1" indicates a stressed high front unrounded vowel /i/.

TABLE II. The number of tokens per vowel context.

Vowel context	/i/	/æ/	/u/
L1 English			
Liquid ^a	155 / 187	188 / 173	119 / 130
Liquid-vowel ^b	177 / 197	199 / 192	130 / 134
L1 Japanese			
Liquid ^a	205 / 246	298 / 286	149 / 170
Liquid-vowel ^b	231 / 284	312 / 310	169 / 180

^a /l/ tokens on the left; /ɹ/ tokens on the right.

^b /l/+vowel tokens on the left; /ɹ/+vowel tokens on the right.

number of bins. The current analysis uses 11 data points throughout the liquid-vowel interval for each formant trajectory. The advantage of using Fast Track is that it performs multiple-step formant estimations by adjusting the maximum formant frequency and obtains the “best-winning” analysis based on regression analyses predicting the formant frequency as a function of time (Barreda, 2021). This achieves increased formant estimation accuracy by specifying different formant frequency ranges according to speakers’ age and gender.

In the current study, the female and male speakers were analysed separately with different ranges of the upper formant frequency ceiling: between 5000–7000 Hz for female speakers and between 4500–6500 Hz for male speakers. Fast Track then performs 24-step formant estimations with varying upper-frequency ceilings and estimates the formant frequencies at 11 equidistant points during (1) the liquid and (2) the liquid-vowel intervals with a 25 ms window padded before and after the segment. After formant tracking, formant estimation errors can be corrected based on visual inspection of the 24-step analyses. Using this, I visually inspected all the tokens one by one and either improved the formant measurement by nominating a different winning analysis or omitted the tokens when none of the analyses looked reasonable. At this visual inspection stage, 118 tokens out of the 2633 tokens (see Sec. IID) were excluded due to poor formant estimation accuracy.

Finally, Fast Track automatically omits tokens when they are shorter than 30 ms as formant estimation can be challenging for extremely short tokens. As a result, 209 tokens were excluded from the dataset for the static analysis, leaving 2306 tokens for static analysis and 2515 tokens for dynamic analysis. The difference in the number of tokens reflects the greater number of liquid-only tokens being omitted automatically by Fast Track as they were inevitably shorter than liquid-vowel intervals (see supplementary material for the data processing procedure described here).¹

F. Statistical analysis

All statistical analyses were performed using R version 4.2.2 (R Core Team, 2022) and data visualisation was performed using the *tidyverse* suite (Wickham et al., 2019). Prior to the statistical analysis, the formant values were transformed into Bark scale using the *bark* function in the

emuR package to allow for cross-speaker comparisons (Jochim et al., 2023).

For the static analysis, Bark-converted F_2-F_1 (Bark F_2-F_1) and F_3 (Bark F_3) at liquid midpoint were modelled using linear mixed-effect models (LME) using the *lme4::lmer* function (Bates et al., 2015). Separate models were constructed for /l/ and /ɹ/, respectively. The fixed effects included (1) the speaker’s first language (*L1*: i.e., English vs Japanese), (2) vowel context (*vowel*), and (3) the speaker’s gender (*gender*). No interactions were included because initial explorations suggested that the current dataset does not have the statistical power to detect interactions.

Furthermore, an anonymous reviewer suggested classifying the participants into groups according to their English proficiency and including this variable for analysis. Following this, I classified the participants into four groups based on the distribution of their subjective fluency rating scores. L1 Japanese speakers are classified into the *advanced* (rating 5–6, $n = 7$), *intermediate* (rating 4, $n = 14$), and *beginner* (rating 1–3, $n = 10$) groups. L1 English speakers constitute a group on their own (*L1 English*; rating 7, $n = 14$). The L1 English speaker group, however, confounds the *proficiency* variable with the *L1* variable, making the inclusion of the *proficiency* variable problematic. The issue is manifested in the rank-deficient warning for LMEs when both *L1* and *proficiency* are included in the same model, suggesting that two or more variables are not linearly independent from each other. A further analysis using the *caret::findLinearCombos* function shows co-linearity between *L1* and *proficiency* and suggests excluding the level of L1 English speakers from the *proficiency* variable.

For this reason, I perform a separate analysis focussing only on the L1 Japanese speakers’ data to investigate the effects of *proficiency* and summarise the results at the end of the static analysis. I have included L1 Japanese speakers only here because inclusion of L1 English speakers might reduce the magnitude of between-group differences among L1 Japanese speakers. The visualisation includes L1 English speakers’ data only for the purpose of comparison. I will not explore this extensively as this is not the main focus of the study (see supplementary material for further details of the analysis and results).¹

The random effect structure for the linear models included by-participant varying slopes and by-participant varying intercepts for vowel contexts and by-word varying intercepts. As a result, the following specification is used for four final models (i.e., models predicting Bark F_2-F_1 and Bark F_3 for /l/ and /ɹ/):

$$\text{lmer}(\text{Bark } F_2-F_1 \text{ or Bark } F_3 \sim L1 + \text{vowel} + \text{gender} + (1 | \text{word}) + (1 + \text{vowel} | \text{speaker})).$$

The significance of the fixed effects was tested via likelihood ratio testing by comparing the full model and the nested model excluding the fixed effect in question (Winter, 2020). If the full model significantly improved the model fit, I concluded that the main effect significantly influenced the outcome variable. The patterns associated with the vowel

contexts are interpreted via data visualisation for the sake of model simplicity (see supplementary material for additional statistical comparisons).¹

Second, the dynamic formant analysis used generalised additive mixed models (GAMMs) using the *mgcv::bam* function (Wood, 2017). The non-linear differences between contours can be evaluated in light of *height* and *shape* of the trajectories; the *height* dimension can be modelled via parametric terms, and the *shape* dimension via so-called *smooth terms* that specify the degree of wiggleness of contours (Sóskuthy *et al.*, 2018). Differences between a set of contours can also be directly modelled by incorporating a *reference smooth* (i.e., a contour at the reference level) and the *difference smooth* (i.e., a contour that models the degree of by-group difference of contours) (Sóskuthy, 2017). For more details about GAMMs, please be referred to the existing tutorial papers (e.g., Sóskuthy, 2017; Sóskuthy *et al.*, 2018; Wieling, 2018).

In the current study, I focus on differences in trajectory height and shape between the speaker groups (i.e., English vs Japanese). Separate models were constructed for each combination of the liquid-vowel pairings. Each model predicts the formant values, either Bark F_2-F_1 or Bark F_3 , by a parametric term of the speaker's first language and gender, as well as a time-varying reference smooth, a time-varying by-L1 difference smooth, and a time-varying by-gender smooth. It also includes time-by-speaker and time-by-word random smooths.

Note, again, that English proficiency was not included in the GAMMs models together with *L1* as this resulted in inaccurate predictions of the formant trajectories compared to the visualisations of the raw data. Instead, similarly to the linear mixed-effect model analysis, I conducted a separate analysis for the effects of *proficiency* using the L1 Japanese speakers' data only and summarise the relevant results at the end of the dynamic analysis. The choice of including L1 Japanese speakers only reflects the consideration that L1 English speakers' trajectories may be different in both shape and height, which would make it difficult for me to interpret whether statistically significant differences result from speakers' L1 or L1 Japanese speakers' proficiency. This is clear in the visualisations in Figs. 9 and 10, in which L1 English speakers' trajectories are distinct from the three groups of L1 Japanese speakers (see supplementary material for further details).¹

Residual autocorrelations in the trajectories were corrected using the autoregressive error model (AR model). The autoregressive parameter (ρ) was set as the amount of autocorrelation at lag 1 in the model, estimated using the *start_value_rho* function in the *itsadug* package (van Rij *et al.*, 2020). While this is usually an adequate estimate, the residual autocorrelations were negative in some cases, indicating that a lower value would be optimal (Sóskuthy *et al.*, 2018; Wieling, 2018). In such cases, the new ρ value was determined by exploring a range of values and visualising the autocorrelations at lag 1 for each ρ value. The final model specification across 12 models (two outcome

variables, i.e., Bark F_2-F_1 and Bark F_3) for two liquids (i.e., /l/ and /ɹ/) in three vowel contexts (i.e., /æ/, /i/ and /u/) is

$$\text{bam}(\text{Bark } F_2-F_1 \text{ or Bark } F_3 \sim LI + \text{gender} + s(\text{time}, \text{bs} = \text{"cr"}) + s(\text{time}, \text{by} = LI, \text{bs} = \text{"cr"}) + s(\text{time}, \text{by} = \text{gender}, \text{bs} = \text{"cr"}) + s(\text{time}, \text{speaker}, \text{bs} = \text{"fs,"} \text{xt} = \text{"cr,"} \text{m} = 1) + s(\text{time}, \text{word}, \text{bs} = \text{"fs,"} \text{xt} = \text{"cr,"} \text{m} = 1), \text{method} = \text{"ML"}).$$

Trajectory height and shape were compared through model comparisons using the *itsadug::compareML* function following the previous research (Kirkham *et al.*, 2019; Sóskuthy, 2017; Sóskuthy *et al.*, 2018) as follows:

- (1) I first compared (1) the full model and (2) the nested model excluding the parametric and the smooth terms associated with the speaker's *L1* or *gender*. This allows a comparison of the overall differences associated with these effects in both height and shape between the two contours.
- (2) If the above comparison showed a significantly improved model fit of the full model, I then compared (1) the full model and (2) the nested model including the parametric term of *L1* or *gender* but still excluding the by-L1 or by-gender smooth term. This tests whether the two contours differ significantly in shape.

If the full model was still better in the model fit after procedure 2 above, I concluded that both trajectory height and shape were different at a statistically significant level. If the full model improved the model fit for procedure 1 but not for procedure 2, then there was only a difference in trajectory height. Otherwise, I concluded that there was little evidence that the two trajectories are significantly different.

III. RESULTS

A. Liquid static analysis

In this section, I first present the liquid midpoint analysis of F_2-F_1 and F_3 using LMEs in order to investigate the overall trends in liquid quality. The static analysis tests the main effects of *L1*, *vowel*, and *gender* while the liquid-vowel interactions are interpreted via data visualisation. Note that the baseline participant population (i.e., intercept) is the female L1 English speakers in the /æ/ context but the gender is referred to only when the *gender* effect is discussed (see supplementary material for an additional analysis of vowel midpoints).¹

1. F_2-F_1 midpoint

The model summaries for the F_2-F_1 models are shown in Table III. The lateral F_2-F_1 model predicts that L1 Japanese speakers produce laterals higher at 8.83 Bark than L1 English speakers (6.74 Bark). F_2-F_1 for laterals slightly varies according to the vowel context; F_2-F_1 is the highest in the /i/ context with an averaged F_2-F_1 being at 8.02 Bark, followed by /u/ (7.54 Bark) and /æ/ (6.74 Bark). Male speakers produce laterals with lower F_2-F_1 values at 6.06 Bark.

TABLE III. LME summary: Liquid F_2-F_1 (Bark).

Variable	β	SE	t	$p(\chi^2)$
Lateral /l/				
Intercept	6.74	0.33	20.36	
L1				<0.001
Japanese	1.99	0.38	5.25	
Vowel				<0.001
/i/	1.28	0.16	8.23	
/u/	0.80	0.18	4.50	
Gender				0.072
Male	-0.68	0.36	-1.86	
Rhotic /ɹ/				
Intercept	6.38	0.34	18.53	
L1				<0.001
Japanese	1.86	0.40	4.68	
Vowel				<0.001
/i/	1.15	0.16	7.09	
/u/	0.60	0.13	4.58	
Gender				0.070
Male	-0.75	0.38	-1.97	

The rhotic F_2-F_1 model predicts that L1 English speakers produce rhotics in the /æ/ context at 6.38 Bark and L1 Japanese speakers overall produce 8.24 Bark. It also predicts higher F_2-F_1 overall in the /i/ context (7.53 Bark) and in the /u/ context (6.98 Bark) than in the /æ/ context. Similar to the laterals, male speakers produce rhotics with lower F_2-F_1 values at 5.63 Bark.

Overall, L1 Japanese speakers produce both English /l/ and /ɹ/ with consistently higher F_2-F_1 than L1 English speakers across vowel contexts (Fig. 3), and this is supported by the significant main effect of L1 for both /l/ [$\chi^2(1) = 17.58, p < 0.001$], and /ɹ/ [$\chi^2(1) = 15.68, p < 0.001$]. The main effect of vowel is also shown to be significant for both /l/ [$\chi^2(2) = 22.74, p < 0.001$] and /ɹ/ [$\chi^2(1) = 22.35, p < 0.001$]. While male speakers produce liquids with lower F_2-F_1 values

than female speakers, this difference was not shown to be statistically significant for either laterals [$\chi^2(1) = 3.23, p = 0.073$], or rhotics [$\chi^2(1) = 3.28, p = 0.070$].

2. F_3 midpoint

The model summaries for the F_3 models are shown in Table IV. The lateral F_3 model predicts that L1 English speakers produce F_3 at 15.83 Bark for /l/ while L1 Japanese speakers have a slightly lower F_3 at 15.54 Bark. Although model comparisons suggest significant effects of vowel for /l/ [$\chi^2(2) = 13.05, p = 0.001$], the difference seems to be quite minor; the model predicts 15.65 Bark for /l/ in the /i/ context and 15.44 Bark in the /u/ context. Finally, female speakers produce laterals with higher F_3 values by 1.12 Bark than male speakers overall.

The rhotic F_3 model predicts that L1 English speakers produce 12.17 Bark for /ɹ/ where L1 Japanese speakers produce higher F_3 at 14.05 Bark. Similar to the laterals, slight differences are found for /ɹ/ in the /i/ and /u/ contexts compared to /æ/; the model predicts 12.54 Bark in the /i/ context and 12.21 Bark in the /u/ context. The main effect of vowel is also significant here [$\chi^2(2) = 13.78, p = 0.001$].

While the main effect of vowel influences the F_3 values only slightly for both /l/ and /ɹ/, the effects of L1 are suggested to be significant for /ɹ/ [$\chi^2(1) = 30.62, p < 0.001$] but not for /l/ [$\chi^2(1) = 1.97, p = 0.161$]. Figure 4 seems to suggest a bimodal distribution in F_3 (Bark) for L1 English speakers, especially for /l/ in the /i/ and /u/ contexts. This seems to result from gender-related differences, in which male speakers produced liquids with lower F_3 values than female speakers. Indeed, the effects of gender are shown to be statistically significant for both laterals [$\chi^2(1) = 22.70, p < 0.001$] and rhotics [$\chi^2(1) = 15.87, p < 0.001$].

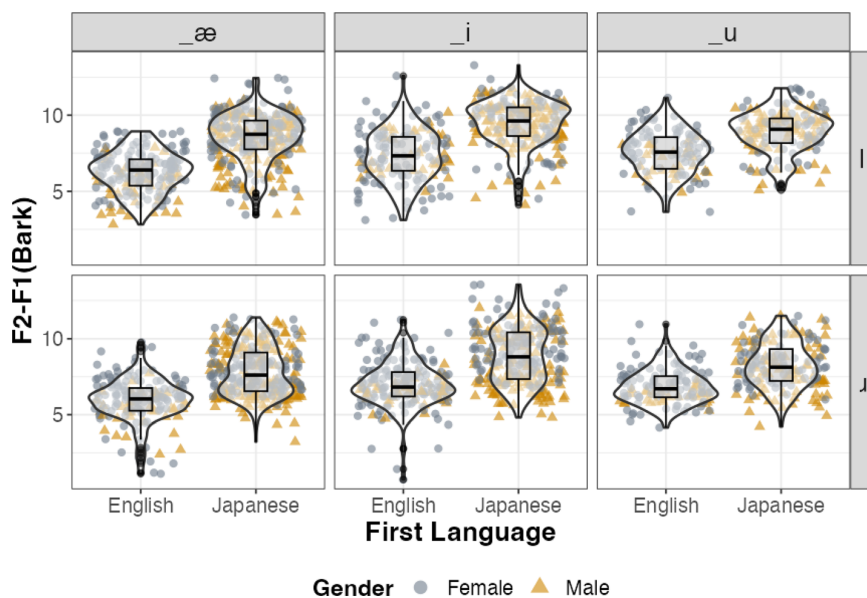


FIG. 3. (Color online) F_2-F_1 (Bark) at liquid midpoint. Each column shows vowel contexts for /l/ (top row) and /ɹ/ (bottom row). Each panel shows distributions of F_2-F_1 (Bark) for L1 English (left) and L1 Japanese (right) speakers. Overlaid is the scatterplot indicating speaker's gender: female (gray circles) and male (yellow triangles) speakers.

TABLE IV. LME summary: Liquid F_3 (Bark).

Variable	β	SE	t	$p(\chi^2)$
Lateral /l/				
Intercept	15.83	0.18	89.35	
L1				0.016
Japanese	-0.29	0.20	-1.44	
Vowel				0.001
/i/	-0.18	0.08	-2.12	
/u/	-0.39	0.08	-4.82	
Gender				<0.001
Male	-1.12	0.19	-5.79	
Rhotic /ɹ/				
Intercept	12.17	0.25	48.56	
L1				<0.001
Japanese	1.88	0.26	7.18	
Vowel				0.001
/i/	0.37	0.08	4.47	
/u/	0.04	0.10	0.41	
Gender				<0.001
Male	-1.15	0.25	-4.53	

3. Effects of L2 proficiency on the midpoint formant measurement

In addition to the main analysis, the effects of *proficiency* are tested for the three groups of L1 Japanese speakers. Grouping is based on their subjective fluency judgement scores: beginner ($n = 10$, rating 1–3), intermediate ($n = 14$, rating 4), and advanced ($n = 7$, rating 5–6). Similarly to the main analysis, separate LME were specified in which Bark F_2-F_1 or Bark F_3 are predicted by fixed effects of *proficiency*, *vowel*, and *gender* with by-item random intercepts and by-speaker random slopes and intercepts for vowels. The results are visualised in Figs. 5 and 6.

The F_2-F_1 models suggested statistically significant effects of *proficiency* on Bark F_2-F_1 for /ɹ/ [$\chi^2(2) = 7.52$,

$p = 0.002$], in which the advanced L1 Japanese learners of English produce rhotics with lower F_2-F_1 than those in the beginner and intermediate groups. No statistically significant *proficiency* effects are found for /l/ [$\chi^2(2) = 0.12$, $p = 0.94$]. For Bark F_3 , no statistically significant effects of *proficiency* are found for either /l/ [$\chi^2(2) = 0.81$, $p = 0.67$] or /ɹ/ [$\chi^2(2) = 0.057$, $p = 0.97$].

4. Summary: Static analysis

L1 Japanese speakers produce higher F_2-F_1 for both /l/ and /ɹ/ across vowel contexts. F_3 values for /l/ are only slightly lower for L1 Japanese speakers while they produce /ɹ/ with higher F_3 than L1 English speakers across vowel contexts. Male speakers produce liquids with lower F_2-F_1 and F_3 values, and this was particularly the case for F_3 . Finally, L1 Japanese speakers in the advanced group produced lower F_2-F_1 than the other groups for /ɹ/.

B. Dynamic analysis

Dynamic analysis in this section now focuses on variation in F_2-F_1 and F_3 trajectories across the liquid-vowel interval using GAMMs. In the visualisation of the liquid-vowel trajectories (Figs. 7 and 8), the liquid portion corresponds roughly to the first third of the interval whereas the vowel corresponds to the second two-thirds. Note the visualisation shows the predictions based on the full models.

1. F_2-F_1 liquid-vowel trajectory

The results of the model comparisons for the F_2-F_1 dynamic analysis are shown in Table V for laterals and Table VI for rhotics. The visualisations are shown in Fig. 7. The model comparisons show that the height and shape of the F_2-F_1 trajectories are significantly different between L1 English and L1 Japanese speakers for both liquids in all vowel contexts. The visualisations of the GAMMs show that



FIG. 4. (Color online) F_3 (Bark) at liquid midpoint. Each column shows vowel contexts for /l/ (top row) and /ɹ/ (bottom row). Each panel shows distributions of F_3 (Bark) for L1 English (left) and L1 Japanese (right) speakers. Overlaid is the scatterplot indicating speaker's gender: female (gray circles) and male (yellow triangles) speakers.

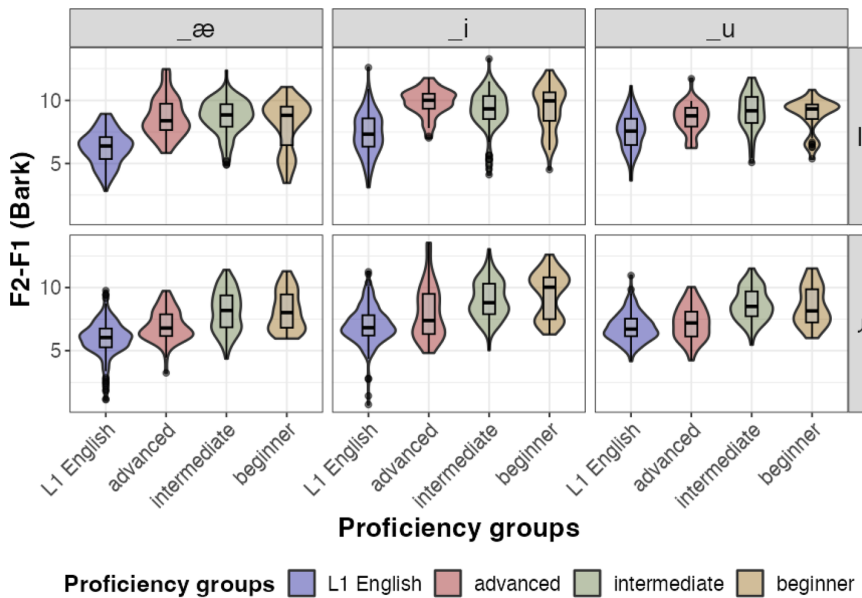


FIG. 5. (Color online) F_2-F_1 (Bark) at liquid midpoint by proficiency groups. Each column shows vowel contexts for /l/ (top row) and /l/ (bottom row). Each panel shows distributions of F_2-F_1 (Bark) for L1 English speakers and three groups of L1 Japanese speakers: advanced, intermediate, and beginner, from left to right.

the trajectories for L1 English and L1 Japanese speakers are similar in the /i/ context (the middle panels in Fig. 7) but look quite different in the /æ/ (left) and /u/ (right) contexts. L1 English speakers follow a similar tendency across the vowel contexts such that they start from lower F_2-F_1 values at the onset of the liquid, showing an increase towards the vowel target and a slight decrease towards the offset of the vowel.

L1 Japanese speakers, on the other hand, show distinct trajectory patterns depending on vowel context. In the /i/ context, their trajectories follow a similar tendency to that of L1 English speakers, but with an earlier rise from the liquid onset towards the vowel resulting in a consistently higher trajectory than L1 English speakers in the first half of the interval. In the /æ/ context, on the other hand, the L1

Japanese speakers show an opposite pattern to L1 English speakers, in which F_2-F_1 values are the highest earlier during the first third of the interval and decrease to the vowel with a small rise towards the end of the interval. Finally, the L1 Japanese speakers' trajectories in the /u/ context show smaller fluctuations than that of L1 English speakers; the trajectory shows almost a linear and monotonic decrease in this vowel context.

Differences associated with *gender* are statistically significant for trajectory height but not for shape for both laterals and rhotics across the vowel contexts. This suggests almost linear differences between female and male speakers' trajectories, in which female speakers show constantly higher trajectories than male speakers, and this is evident in Fig. 7.

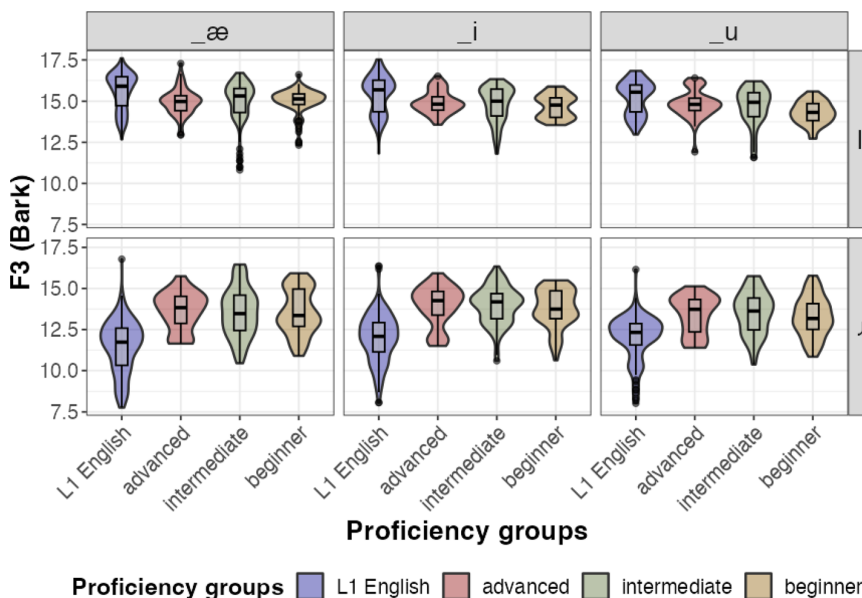


FIG. 6. (Color online) F_3 (Bark) at liquid midpoint by proficiency groups. Each column shows vowel contexts for /l/ (top row) and /l/ (bottom row). Each panel shows distributions of F_2-F_1 (Bark) for L1 English speakers and three groups of L1 Japanese speakers: advanced, intermediate, and beginner, from left to right.

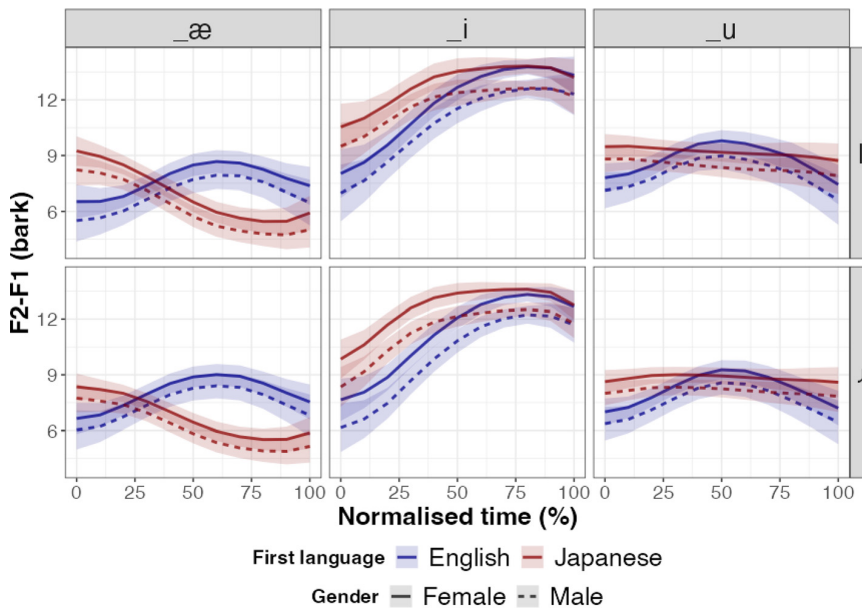


FIG. 7. (Color online) The F_2-F_1 (Bark) trajectories predicted by GAMMs over the liquid-vowel intervals for each liquid (rows) in each vowel context (columns). Each panel shows predictions based on the full model with a mean smooth and 95% confidence interval for L1 English (blue) and L1 Japanese (red) speakers and for female (solid) and male (dashed) speakers.

2. F_3 liquid-vowel trajectory

The model comparisons for F_3 are shown in Table VII for laterals and in Table VIII for rhotics. The visualisations are shown in Fig. 8. The lateral-vowel trajectories (the top half of Fig. 8) show similarities between L1 English and L1 Japanese speakers. The model comparisons suggest that, while the trajectory shape and height are different between L1 English and L1 Japanese speakers in the /i/ context, the trajectories in the /æ/ and /u/ contexts are not statistically significantly different, with the L1 Japanese speakers' trajectories being slightly lower, especially in the first half of the interval.

Even in the lateral-/i/ context where trajectory height and shape are statistically significant, however, a closer look

at the GAMMs model specifications and the model comparisons suggest that the difference between the two trajectories is marginal. Neither parametric or smooth terms associated with the $L1$ difference were statistically significant in the model summary [$\beta=0.18$, standard error (SE)=0.10, $t=1.83$, $p=0.07$ for the parametric term; $F(6.05)=1.79$, $p=0.09$ for the difference smooth]. The model comparison also suggests only a marginal improvement in the Akaike Information Criterion (AIC) values (1561.42 for the full model and 1565.84 for the nested model). Figure 8 also shows that the 95% confidence intervals of two trajectories overlap substantially throughout the liquid-vowel interval.

The /r/-vowel trajectories for F_3 , on the other hand, show statistically significant differences in both trajectory

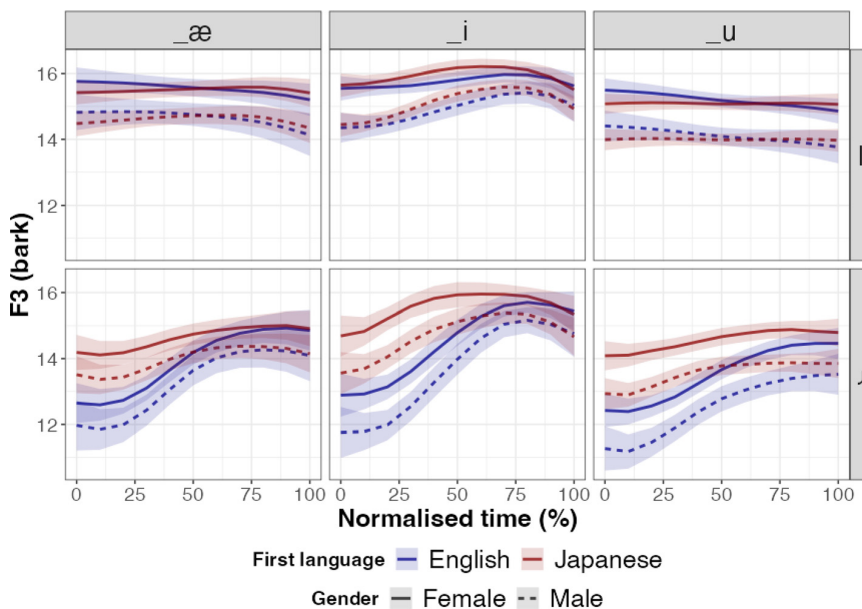


FIG. 8. (Color online) The F_3 (Bark) trajectories predicted by GAMMs over the liquid-vowel intervals for each liquid (rows) in each vowel context (columns). Each panel shows predictions based on the full model with a mean smooth and 95% confidence interval for L1 English (blue) and L1 Japanese (red) speakers and for female (solid) and male (dashed) speakers.

TABLE V. Model comparisons for F_2-F_1 GAMMs for laterals.

Comparison	χ^2	<i>df</i>	$p(\chi^2)$
/l/: /æ/ context			
Overall: L1	69.88	3	<0.001
Shape: L1	66.63	2	<0.001
Overall: gender	6.67	3	0.004
Shape: gender	1.10	2	0.333
/l/: /i/ context			
Overall: L1	16.54	3	<0.001
Shape: L1	9.68	2	<0.001
Overall: gender	18.91	3	<0.001
Shape: gender	0.34	2	0.712
/l/: /u/ context			
Overall: L1	25.41	3	<0.001
Shape: L1	23.67	2	<0.001
Overall: gender	4.02	3	0.045
Shape: gender	0.07	2	0.929

TABLE VI. Model comparisons for F_2-F_1 GAMMs for rhotics.

Comparison	χ^2	<i>df</i>	$p(\chi^2)$
/ɹ/: /æ/ context			
Overall: L1	53.57	3	<0.001
Shape: L1	45.94	2	<0.001
Overall: gender	4.10	3	0.042
Shape: gender	0.06	2	0.938
/ɹ/: /i/ context			
Overall: L1	39.40	3	<0.001
Shape: L1	24.09	2	<0.001
Overall: gender	21.90	3	<0.001
Shape: gender	0.33	2	0.723
/ɹ/: /u/ context			
Overall: L1	21.62	3	<0.001
Shape: L1	17.83	2	<0.001
Overall: gender	4.00	3	0.046
Shape: gender	0.02	2	0.985

TABLE VII. Model comparisons for F_3 GAMMs for laterals.

Comparison	χ^2	<i>df</i>	$p(\chi^2)$
/l/: /æ/ context			
Overall: L1	3.12	3	0.100
Shape: L1	—	—	—
Overall: gender	17.57	3	<0.001
Shape: gender	1.22	2	0.295
/l/: /i/ context			
Overall: L1	4.43	3	0.031
Shape: L1	2.53	2	0.080
Overall: gender	33.71	3	<0.001
Shape: gender	5.67	2	0.003
/l/: /u/ context			
Overall: L1	1.81	3	0.306
Shape: L1	—	—	—
Overall: gender	29.91	3	<0.001
Shape: gender	0.00	2	1.000

TABLE VIII. Model comparisons for F_3 GAMMs for rhotics.

Comparison	χ^2	<i>df</i>	$p(\chi^2)$
/ɹ/: /æ/ context			
Overall: L1	17.36	3	<0.001
Shape: L1	10.32	2	<0.001
Overall: gender	8.26	3	<0.001
Shape: gender	1.05	2	0.350
/ɹ/: /i/ context			
Overall: L1	43.55	3	<0.001
Shape: L1	26.89	2	<0.001
Overall: gender	22.40	3	<0.001
Shape: gender	3.21	2	0.041
/ɹ/: /u/ context			
Overall: L1	27.42	3	<0.001
Shape: L1	8.31	2	<0.001
Overall: gender	25.96	3	<0.001
Shape: gender	2.87	2	0.057

height and shape in all vowel contexts, although both L1 English and L1 Japanese speakers share a similar trend in the visualisation in Fig. 8. Both groups show lower F_3 values at the liquid onset, which then increase towards the vowel, where L1 English and L1 Japanese speakers' trajectories seem to converge. L1 Japanese speakers' trajectories are overall flatter and higher than that of L1 English speakers across all the vowel contexts.

Finally, similarly to the F_2-F_1 results, the *gender* effect seems to be statistically significant only for the trajectory height. This again suggests that the difference between trajectories for female and male speakers is close to linear (see Fig. 8).

3. Effects of L2 proficiency on formant trajectories

Similar to the static analysis, the effects of L1 Japanese speakers' proficiency have been tested separately from the main analysis. For each liquid-vowel pairing, the models predicts Bark F_2-F_1 or Bark F_3 with parametric terms of *proficiency* and *gender*, a time-varying reference smooth, a time-varying by-proficiency difference smooth, and a time-varying by-gender difference smooth. The random effect is accounted for by time-by-speaker and time-by-word random smooths. The visualisations are shown in Figs. 9 and 10; please note that the predictions shown in these figures are based on the models excluding parametric and smooth terms associated with *gender* because the plots would be too crowded to interpret otherwise.

The analyses for Bark F_2-F_1 suggest a statistically significant effect of *proficiency* on the trajectory height for /ɹ/ in the /u/ context [$\chi^2(6) = 10.24, p = 0.002$], in which the F_2-F_1 trajectory for the advanced group is lower than the beginner or intermediate groups. The visualisation in Fig. 9, however, shows that the trajectory shape is quite different between L1 English speakers and the advanced L1 Japanese speakers. For Bark F_3 , no statistically significant effects of *proficiency* are found for either /l/ or /ɹ/ for the L1 Japanese speakers.

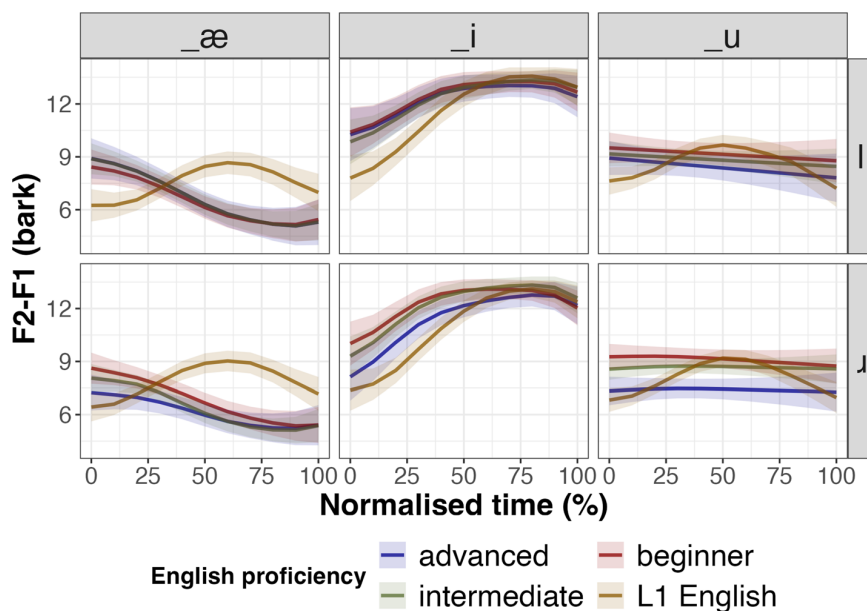


FIG. 9. (Color online) The F_2-F_1 (Bark) trajectories illustrating differences between the different proficiency groups among L1 Japanese speakers predicted by GAMMs over the liquid-vowel intervals for each liquid (rows) in each vowel context (columns). Each panel shows predictions based on the model excluding parametric and smooth terms associated with *gender* for simplicity, with a mean smooth and 95% confidence interval for advanced (blue), intermediate (red), beginner (green) L1 Japanese speakers and L1 English speakers (orange).

4. Summary: Dynamic analysis

The dynamic analysis shows substantial variability in the liquid-vowel realisations between L1 English and L1 Japanese speakers. Shape and height are significantly different for the F_2-F_1 trajectories for both /l/ and /ɹ/, with differences associated not only with the liquid portion corresponding to the first third of the interval but also with the transition patterns into the vowel. The F_3 trajectories for /l/ are largely comparable between L1 English and L1 Japanese speakers with little evidence of statistically significant differences. The F_3 trajectories for /ɹ/, on the other hand, differ substantially in the first half of the interval corresponding to the liquid portion. The effects of *gender* are manifested almost exclusively on the trajectory height,

meaning a linear difference between trajectories for female and male speakers. Although advanced L1 Japanese speakers produced the lower F_2-F_1 trajectories in the /ɹ/-/u/ context than the beginner and intermediate groups, the trend is quite different from that of L1 English speakers.

IV. DISCUSSION

A. Spectro-temporal variability in L2 English liquids

The current paper aims to capture time-varying acoustic properties of English liquids produced by L1 English and L1 Japanese speakers. It combines two analyses of F_2-F_1 and F_3 : the static analysis at the liquid midpoint and the dynamic analysis over the liquid-vowel interval. The liquid midpoint

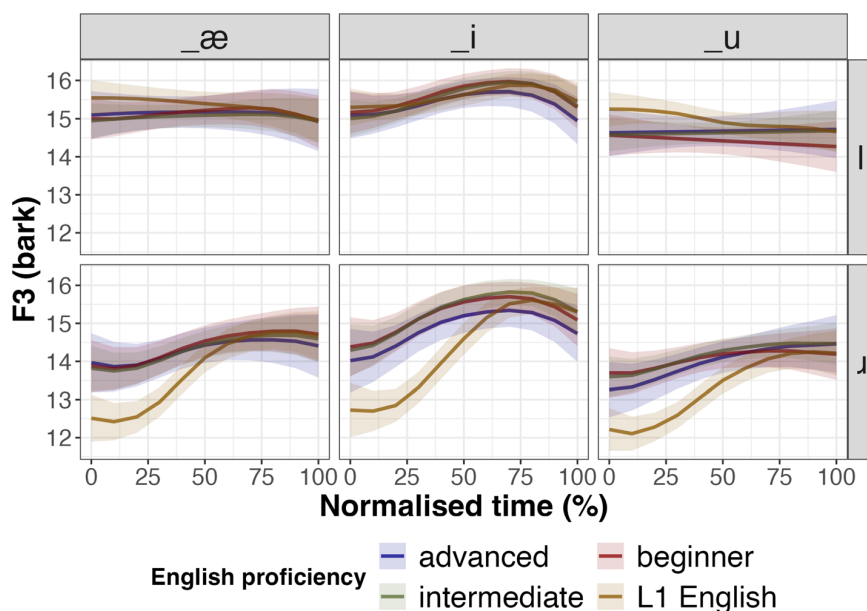


FIG. 10. (Color online) The F_3 (Bark) trajectories illustrating differences between the different proficiency groups among L1 Japanese speakers predicted by GAMMs over the liquid-vowel intervals for each liquid (rows) in each vowel context (columns). Each panel shows predictions based on the model excluding parametric and smooth terms associated with *gender* for simplicity, with a mean smooth and 95% confidence interval for advanced (blue), intermediate (red), beginner (green) L1 Japanese speakers and L1 English speakers (orange).

analysis suggests that L1 Japanese speakers constantly produce higher F_2-F_1 for both English /l/ and /ɹ/ and higher F_3 for /ɹ/ than L1 English speakers across vowel contexts. The dynamic analysis, on the other hand, shows that the between-L1 differences are non-linear, highlighting the complexity associated with the production of liquids and liquid-vowel coarticulation.

Comparing the effects of speaker gender and L1 demonstrate the importance of dynamic information in the liquid-vowel sequences. The static analysis shows that male speakers generally produce English liquids with lower F_2-F_1 and F_3 frequencies than female speakers, and the gender difference is statistically significant for F_3 . The dynamic analysis further shows clearly that the spectral difference between female and male speakers seems to be linear; GAMMs model comparisons suggest statistically significant differences in trajectory height but not in trajectory shape, and it is quite clear from the visualisations in Figs. 7 and 8 that the differences in trajectories between female and male speakers are (almost) linear.

The dynamic difference associated with speaker L1, on the other hand, draws a much more complicated picture. While the time-varying analysis of F_3 for /l/ indicates little difference between L1 English and L1 Japanese speakers, the F_3 values for /ɹ/ show a clear between-L1 difference in the first half of the interval, indicating differences in acoustic realisations of liquids and the transition into the vowel. Also, the trajectory shape associated with L1 Japanese speakers' /ɹ/ is flatter, resulting in a smaller distinction between /ɹ/ and the vowels. The two language groups slightly differ in the point in time at which F_3 achieves its maximum, such that L1 Japanese speakers seem to achieve the vowel target earlier than L1 English speakers do.

The F_2-F_1 trajectories further highlight the non-linear between-L1 differences in the trajectory (Fig. 7). In particular, L1 Japanese speakers show distinct trajectory patterns across vowel contexts, suggesting that their production of English liquids is subject to greater influence from the following vowels than that of L1 English speakers. The liquid-/i/ trajectories, for example, suggest that L1 Japanese speakers reach the vowel target earlier, given the early onset of the plateau, than L1 English speakers despite a similar trajectory pattern. The linear trend for the liquid-/u/ trajectories also indicates that L1 Japanese speakers do not clearly distinguish the liquid and the vowel on F_2-F_1 .

The separate static analyses on the effects of L1 Japanese speakers' English proficiency demonstrated that advanced L1 Japanese-speaking learners of English produced lower F_2-F_1 values for /ɹ/ than the other two groups. Given that L1 English speakers produced lower F_2-F_1 values for /ɹ/ at the liquid midpoint, the findings support the previous claims that English /ɹ/ is easier for L1 Japanese speakers to learn than English /l/ (Aoyama *et al.*, 2004), and that the use of F_2 and F_1 may be easier for them to acquire than that of F_3 (Saito and Munro, 2014). The dynamic analysis in the current study further demonstrates that advanced L1 Japanese speakers' F_2-F_1 trajectory is statistically

significantly lower in the /ɹ-/u/ context than the other two groups. While this could be taken as evidence of the *proficiency* effects, the linear trend of the trajectories across proficiency groups also suggests that even advanced L1 Japanese speakers do not seem to differentiate /ɹ/ and /u/. Fundamentally, this lack of liquid-vowel differentiation might demonstrate a general influence from their L1 (i.e., Japanese). Further research is clearly needed to investigate the effects of L2 proficiency on the formant dynamics by employing more rigorous measures of L2 proficiency, especially given that acoustic profiles of L2 English liquids can be complex (Aoyama *et al.*, 2019).

Overall, the dynamic analysis suggests that L1 Japanese speakers seem to differ not only in acoustic targets of English liquids, as captured in the static analysis, but also in the transition between the liquid and the vowel. The results are in line with the previous findings that the magnitude and timing of spectral changes differ in the production of English liquids by L1 English-speaking children (Howson and Redford, 2021) and by L2 learners of English (Espinal *et al.*, 2020) from that of adult L1 English speakers. These non-linear between-language differences could point to some possible mechanisms whereby L1 Japanese speakers struggle to produce English liquids accurately in light of L2 speech learning.

B. Acquisition of English /l/ and /ɹ/ by L1 Japanese speakers

The overarching question in this study concerns how L1 Japanese speakers differ from L1 English speakers in dynamic acoustic realisations of word-initial English liquids as a function of following vowels. The static analysis suggests that both speaker's L1 and vowel context influence the acoustic realisations of word-initial English /l/ and /ɹ/. The L1 effect is unsurprising, given that it largely agrees with previous findings that L1 Japanese speakers produce both English /l/ and /ɹ/ with higher F_2 and F_3 values than L1 English speakers (Aoyama *et al.*, 2019; Flege *et al.*, 1995; Saito and van Poeteren, 2018). Regarding the vowel effect, the static analysis suggests a general tendency that liquids in the /i/ context are produced with higher F_2-F_1 values than in the /u/ context, whereas the /æ/ context seems to facilitate the lowest F_2-F_1 values for liquids. This could be explained in light of previous findings that the F_2 values in English liquids tend to be higher when preceding a high vowel /i/ than a low vowel /a/ due to different articulatory demands on the tongue dorsum configurations (Recasens, 2012).

The dynamic results demonstrate that L1 Japanese speakers show different patterns of liquid-vowel coarticulatory patterns depending on the following vowel compared to L1 English speakers whose trajectory patterns are consistent across the vowel contexts. The liquid-/u/ trajectories, in particular, suggest that L1 Japanese speakers make a less clear distinction between the liquid and the vowel in the /u/ context. This could corroborate previous perceptual findings that L1 Japanese speakers are more likely to perceive a /w/-like percept when perceiving English /l/ and /ɹ/, resulting in a confusion between English /l ɹ/ and other categories

(e.g., /w/ or [wɹ]) and therefore in less success in identifying word-initial liquids in the back vowel context than in the front vowel context (Best and Strange, 1992; Guion *et al.*, 2000; Mochizuki, 1981; Shimizu and Dantsuji, 1983). The data in this study demonstrate that such confusion arising from the vocalic component of English liquids in perception could also be observed in L1 Japanese speakers' production.

Generally, L1 Japanese speakers produce higher F_3 for English /ɹ/ (Aoyama *et al.*, 2019; Flege, 1995; Saito and Munro, 2014). This is apparent in both static and dynamic analyses; in particular, the dynamic analysis for F_3 in Fig. 8 shows that by-group difference largely lies during the liquid portion, suggesting that the difference in F_3 would be attributed to the liquid realisations. Previous research claims that F_2 is an easier acoustic cue for L1 Japanese speakers to acquire (e.g., Saito and Munro, 2014; Saito and van Poeteren, 2018). While variations in F_1 could be negligible between the two speaker populations (e.g., Flege *et al.*, 1995; Saito and Munro, 2014), this claim does not explain well why the F_2 - F_1 trajectories, which could derive from variations of F_2 , are significantly different both in height and shape between L1 Japanese and L1 English speakers (see Fig. 7). It could therefore be argued that the static analysis only captures a snapshot of acoustical realisations of English liquids, when, in fact, L1 Japanese speakers differ from L1 English speakers in the dynamic spectral characteristics during the liquid-vowel interval.

In addition, an anonymous reviewer suggested a possibility that L1 Japanese speakers might use different dynamic strategies to make a contrast (e.g., through F_2) compared to L1 English speakers. It would, therefore, be worthwhile to investigate how L1 Japanese speakers use dynamic information to make such a phonological contrast, given especially that the Perceptual Assimilation model of L2 Speech Learning (PAM-L2) makes predictions about how L2 speakers assimilate L2 phonological contrasts into their L1 phonology (Best and Tyler, 2007).

Theoretically, the Speech Learning model (SLM) posits that L2 learners store representations of the L2 sounds at the level of the position-sensitive allophones (Flege, 1995; Flege and Bohn, 2021), and previous studies show that L1 Japanese speakers' perception of English /l/ and /ɹ/ is highly subject to the phonetic context and the coarticulatory effects with neighbouring segments (Mochizuki, 1981; Sheldon and Strange, 1982). Taken together, the current results demonstrate that L1 Japanese speakers are influenced by the phonetic details of L2 English liquids, not only in perception but also in production; L1 Japanese speakers show different patterns in the way they dissociate the liquid and vowel clearly, especially in the /u/ context, manifested in their production as different patterns of liquid-vowel coarticulation.

To summarise, the present study shows that the temporal spectral changes during the liquid-vowel intervals are significantly different between L1 English and L1 Japanese speakers along F_2 - F_1 for both liquids and F_3 for /ɹ/. The liquid-vowel trajectories of F_2 - F_1 in the /i/ and /u/ contexts highlight particularly notable temporal variability in the L1

Japanese speakers' data, suggesting that the liquid-vowel coarticulation could be considered as one of the production properties that L1 Japanese speakers need to acquire in production of English liquids.

V. CONCLUSION

The present study examines the acoustics of L1 Japanese and L1 English speakers' production of word-initial English liquids. The key findings include that L1 Japanese speakers differ in the coarticulatory pattern between the liquid and vowel from L1 English speakers. The dynamic analysis using GAMMs not only generally agrees with the findings from the static analysis but also highlights the robust yet complicated differences between L1 and L2 speech in the formant dynamics. Overall, this study illustrates that the dynamic characteristics are important aspects involved in production of English liquids in the context of L2 speech learning. Directly studying formant dynamics opens discussions around the specific underlying mechanism of L2 speech production under the influence of speakers' L1, and future research will complement the current results using articulatory methods for a better understanding of the factors that may underlie differences in acoustic dynamics shown in this study.

ACKNOWLEDGMENTS

I thank Professor Claire Nance and Dr. Sam Kirkham for their comments and support. Professor Noriko Nakanishi, Professor Yuri Nishio, and Dr. Bronwen Evans helped me with data collection. The research is financially supported by Graduate Scholarship for Degree-Seeking Students by Japan Student Services Organization (JASSO) and the 2022 Research Grant by the Murata Science Foundation. Data and codes that support the findings of this study are openly available on the Open Science Foundation (OSF) repository at <https://osf.io/2phx5/>. The author has no conflicts to disclose. This research is approved by ethics committees at Lancaster University, Kobe Gakuin University, and Meijo University. Informed consent was obtained from all participants.

¹See supplementary material at <https://osf.io/2phx5/> for further details about the participants; the data processing procedure; further details of the analysis and results; additional statistical comparisons; and an additional analysis of vowel midpoints.

Akamatsu, T. (1997). *Japanese Phonetics: Theory and Practice* (Lincom Europa, München, Newcastle).

Aoyama, K., Flege, J. E., Akahane-Yamada, R., and Yamada, T. (2019). "An acoustic analysis of American English liquids by adults and children: Native English speakers and native Japanese speakers of English," *J. Acoust. Soc. Am.* **146**(4), 2671–2681.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (2004). "Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /ɹ/," *J. Phon.* **32**(2), 233–250.

Arai, T. (2013). "On why Japanese /r/ sounds are difficult for children to acquire," in *Proceedings Interspeech 2013, ISCA*, Lyon, France, pp. 2445–2449.

- Articulate Instruments (2022). "Articulate Assistant Advanced" (version 220).
- Barreda, S. (2021). "Fast Track: Fast (nearly) automatic formant-tracking using Praat," *Linguistics Vanguard* 7(1), 20200051.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.* 67, 1–48.
- Beristain, A. M. (2022). "The acquisition of acoustic and aerodynamic patterns of coarticulation in second and heritage languages," Ph.D. thesis, University of Illinois Urbana-Champaign, Urbana, IL.
- Best, C. T., and Strange, W. (1992). "Effects of phonological and phonetic factors on cross-language perception of approximants," *J. Phon.* 20(3), 305–330.
- Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins Publishing Company, Amsterdam), pp. 13–34.
- Boersma, P., and Weenink, D. (2022). "Praat: Doing phonetics by computer."
- Campbell, F., Gick, B., Wilson, I., and Vatikiotis-Bateson, E. (2010). "Spatial and temporal properties of gestures in North American English /r/," *Lang. Speech* 53(1), 49–69.
- Carter, P., and Local, J. (2007). "F2 variation in Newcastle and Leeds English liquid systems," *J. Int. Phon. Assoc.* 37(2), 183–199.
- Espinal, A., Thompson, A., and Kim, Y. (2020). "Acoustic characteristics of American English liquids /l/, /l/, /l/ produced by Korean L2 adults," *J. Acoust. Soc. Am.* 148(2), EL179–EL184.
- Espy-Wilson, C. Y. (1992). "Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English," *J. Acoust. Soc. Am.* 92(2), 736–757.
- Flege, J. E. (1995). "Second language speech learning theory, findings and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Baltimore, MD), pp. 233–277.
- Flege, J. E., and Bohn, O.-S. (2021). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress*, 1st ed., edited by R. Wayland (Cambridge University Press, Cambridge, UK), pp. 3–83.
- Flege, J. E., Takagi, N., and Mann, V. (1995). "Japanese adults can learn to produce English /ɹ/ and /l/ accurately," *Lang. Speech* 38(1), 25–55.
- Grosjean, F. (2008). "The bilingual's language mode," in *Studying Bilinguals, Oxford Linguistics* (Oxford University Press, Oxford, New York), pp. 36–66.
- Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (2000). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," *J. Acoust. Soc. Am.* 107(5), 2711–2724.
- Hattori, K., and Iverson, P. (2009). "English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy," *J. Acoust. Soc. Am.* 125(1), 469–479.
- Howson, P. J., and Redford, M. A. (2021). "The acquisition of articulatory timing for liquids: Evidence from child and adult speech," *J. Speech. Lang. Hear. Res.* 64(3), 734–753.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). "A perceptual interference account of acquisition difficulties for non-native phonemes," *Cognition* 87(1), B47–B57.
- Jochim, M., Winkelmann, R., Jaensch, K., Cassidy, S., and Harrington, J. (2023). "emuR - Main package of the EMU Speech Database Management System," R package version 2.4.2, <https://CRAN.R-project.org/package=emuR>.
- Keating, P. A. (1985). "Universal phonetics and the organization of grammars," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. Fromkin (Academic Press, Orlando, FL), pp. 115–132.
- King, H., and Ferragne, E. (2020). "Loose lips and tongue tips: The central role of the /r/-typical labial gesture in Anglo-English," *J. Phon.* 80, 100978.
- Kirkham, S. (2017). "Ethnicity and phonetic variation in Sheffield English liquids," *J. Int. Phon. Assoc.* 47(1), 17–35.
- Kirkham, S., Nance, C., Littlewood, B., Lightfoot, K., and Groarke, E. (2019). "Dialect variation in formant dynamics: The acoustics of lateral and vowel sequences in Manchester and Liverpool English," *J. Acoust. Soc. Am.* 145(2), 784–794.
- Ladefoged, P., and Johnson, K. (2010). *A Course in Phonetics, International Edition*, 6th ed. (Wadsworth, Boston, MA).
- Lawson, E., Stuart-Smith, J., Scobbie, J. M., Yaeger-Dror, M., and Maclagan, M. (2011). "Liquids," in *Sociophonetics: A Student's Guide*, edited by M. Di Paolo and M. Yaeger-Dror (Routledge, Abingdon, VA), pp. 72–86.
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). "Montreal Forced Aligner: Trainable text-speech alignment using Kaldi," in *Proceedings Interspeech 2017*, pp. 498–502.
- Mochizuki, M. (1981). "The identification of /r/ and /l/ in natural and synthesized speech," *J. Phon.* 9(3), 283–303.
- Morimoto, M. (2020). "Geminated liquids in Japanese: A production study," Ph.D. thesis, University of California, Santa Cruz, CA.
- Proctor, M. (2011). "Towards a gestural characterization of liquids: Evidence from Spanish and Russian," *Lab. Phonol.* 2(2), 451–485.
- Proctor, M., Walker, R., Smith, C., Szalay, T., Goldstein, L., and Narayanan, S. (2019). "Articulatory characterization of English liquid-final rimes," *J. Phon.* 77, 100921.
- R Core Team (2022). "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing.
- Recasens, D. (1991). "On the production characteristics of apicoalveolar taps and trills," *J. Phon.* 19(3-4), 267–280.
- Recasens, D. (2012). "A cross-language acoustic study of initial and final allophones of /l/," *Speech Commun.* 54(3), 368–383.
- Riney, T. J., Takada, M., and Ota, M. (2000). "Segmentals and global foreign accent: The Japanese flap in EFL," *TESOL Quart.* 34(4), 711–737.
- Saito, K., and Munro, M. J. (2014). "The early phase of /l/ production development in adult Japanese learners of English," *Lang. Speech* 57(4), 451–469.
- Saito, K., and van Poeteren, K. (2018). "The perception–production link revisited: The case of Japanese learners' English /ɹ/ performance," *Int. J. Appl. Linguist.* 28(1), 3–17.
- Setter, J., and Jenkins, J. (2005). "Pronunciation," *Lang. Teach.* 38(1), 1–17.
- Sheldon, A., and Strange, W. (1982). "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception," *Appl. Psycholinguist.* 3(3), 243–261.
- Shimizu, K., and Dantsuji, M. (1983). "A study on the perception of /r/ and /l/ in natural and synthetic speech sounds," *Stud. Phonologica* 17, 1–14.
- Sóskuthy, M. (2017). "Generalised additive mixed models dynamic analysis linguistics: A practical introduction," [arXiv:1703.05339](https://arxiv.org/abs/1703.05339).
- Sóskuthy, M., Foulkes, P., Hughes, V., and Haddican, B. (2018). "Changing words and sounds: The roles of different cognitive units in sound change," *Top. Cogn. Sci.* 10(4), 787–802.
- Sproat, R., and Fujimura, O. (1993). "Allophonic variation in English /l/ and its implications for phonetic implementation," *J. Phon.* 21(3), 291–311.
- Stevens, K. N. (2000). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).
- van Rij, J., Wieling, M., Baayen, R. H., and van Rijn, H. (2020). "Itsadug: Interpreting time series and autocorrelated data using GAMMs."
- Wells, J. C. (2008). *Longman Pronunciation Dictionary*, 3rd ed. (Pearson Education Ltd., Harlow, UK).
- West, P. (1999a). "The extent of coarticulation of English liquids: An acoustic and articulatory study," in *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS-14)*, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey (San Francisco, CA), pp. 1901–1904.
- West, P. (1999b). "Perception of distributed coarticulatory properties of English /l/ and /r/," *J. Phon.* 27(4), 405–426.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemond, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). "Welcome to the Tidyverse," *J. Open Source Softw.* 4(43), 1686.
- Wieling, M. (2018). "Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English," *J. Phon.* 70, 86–116.
- Winter, B. (2020). *Statistics for Linguists: An Introduction Using R* (Routledge, London, UK).
- Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*, 2nd ed. (Chapman and Hall/CRC, New York).

Yamada, R. A., and Tohkura, Y. (1992). "The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners," *Percept. Psychophys.* **52**(4), 376–392.

Yamane, N., Howson, P., and Po-Chun, W. (2015). (Grace) "An ultrasound examination of taps in Japanese," in *Proceedings of the 18th International Congress of Phonetic Sciences* Glasgow, UK (August 10–14, 2015), pp. 1–5.

Ying, J., Shaw, J. A., Kroos, C., and Best, C. T. (2012). "Relations between acoustic and articulatory measurements of /l/," in *Proceedings of the 14th Australasian International Conference on Speech Science and Technology* (Sydney), pp. 109–112.

Zimmermann, G. N., Price, P., and Ayusawa, T. (1984). "The production of English /r/ and /l/ by two Japanese speakers differing in experience with English," *J. Phon.* **12**(3), 187–193.

Chapter 8

Study 2: Articulatory dynamics in second language speech

So far, it has been demonstrated that L1 Japanese speakers' production of L2 English liquids is more susceptible to the coarticulatory effects of the vowel contexts in the word-medial vowel-liquid-vowel sequence (Chapter 6) and in the word-initial liquid-vowel sequence (Chapter 7). While Chapter 7 clearly shows between-group coarticulatory differences, the findings in the articulatory study presented in Chapter 6 were based on qualitative observation of the PC trajectories, calling for a more formal comparison of tongue movement. This chapter therefore extends these two studies by presenting an articulatory study in which I use the Principal Component Analysis (PCA) to identify salient lingual dimensions involved in midsagittal tongue shape change during the word-initial liquid-vowel sequences and formally model between-group variability in tongue dorsum movement using Bayesian hierarchical regression modelling. The findings suggest that L1 Japanese speakers show greater variability in tongue dorsum movement than L1 English speakers, supporting the earlier observations that L1 Japanese speakers' production is susceptible to coarticulatory influences. This manuscript has been submitted to *Journal of Phonetics* and received a major revision decision.

In the first round of review, the editors and reviewers have suggested following

main changes. One major point is to clarify whether by-group differences shown in this study indeed result from coarticulatory patterns or they result simply from differences in articulatory targets for the liquids and vowels. Second, the link between perception and production presented in RQ1 seems to blur the focus of the paper; since it is not a central focus of the paper, it was suggested that the framing of the paper could be more centred around elaborating on the production patterns. Finally, further elaborations have been suggested regarding the quantitative analysis methods, particularly the combination of PCA and FPCA that projects the data into very abstract space.

I am planning to address the comments by:

- reducing the element concerning perceptual accuracy from the literature review and the methods to clarify the paper's focus on speech production,
- elaborating on the relationships between articulatory target and coarticulation in the literature review section,
- explaining the methods and data more clearly, including (1) why FPCA has been used instead of other methods such as GAMMs, (2) how the interpretations of each PC have emerged, and (3) how coarticulatory patterns shown in Figure 3 could be interpreted in light of potential between-group differences in liquid and vowel targets; and,
- linking the discussion section more closely to the results and the materials reviewed in the literature review section.

1 Learning to resist: Japanese speakers' production of
2 liquid-vowel coarticulation in L2 English

3 Takayuki Nagamine^a

4 ^a*Department of Linguistics and English Language, County South, Lancaster*
5 *University, Lancaster, LA1 4YL, United Kingdom*

6 **Abstract**

7 This article reports an articulatory study using ultrasound tongue imag-
8 ing investigating first-language (L1) Japanese and L1 English speakers' pro-
9 duction of word-initial English /l/ and /ɹ/. Articulatory properties of L1
10 Japanese speakers' production of English /l/ and /ɹ/ are not well-understood,
11 with previous research almost exclusively focussing on acoustics and/or static
12 measures of an inherently dynamic signal. The current study therefore com-
13 pares time-varying dynamics of midsagittal tongue shape in the liquid-vowel
14 sequence in English between 29 L1 Japanese and 14 L1 English speakers. The
15 results demonstrate clear by-group differences in the liquid-vowel coarticula-
16 tory patterns in tongue dorsum movement, whereby L1 Japanese speakers'
17 production is more variable depending on the vowel contexts than that of
18 L1 English speakers. This could be due to the transfer of the coarticulatory
19 patterns from Japanese /r/ and differences in the degree of coarticulatory
20 resistance. The findings overall demonstrate that time-varying properties fa-
21 cilitate a better understanding as to why certain sounds are difficult for L2
22 learners to produce accurately.

23 *Keywords:* second language speech production, English liquids,
24 coarticulation, ultrasound tongue imaging, dynamic analysis, Principal
25 Component Analysis

26 **1. Introduction**

27 This article reports an articulatory study investigating how first-language
28 (L1) Japanese speakers differ from L1 English speakers in producing English
29 /l/ and /ɹ/. It is well-known that L1 Japanese speakers have difficulty in

30 producing English /l/ and /ɭ/ in a similar manner as L1 English speakers.
31 Despite implicit postulations that L1 Japanese speakers’ difficulty in produc-
32 tion of English liquids might be derived from L1-L2 differences in articulation
33 or ‘articulatory routines’ in their L1 (Bradlow, 2008; Davidson, 2011; Olsen,
34 2012; Colantoni and Steele, 2008), it remains unclear how L1 Japanese speak-
35 ers articulate English /l/ and /ɭ/ because the majority of previous research
36 focusses on static acoustic properties (e.g., Flege et al., 1995; Aoyama et al.,
37 2019, 2023; Saito and Munro, 2014).

38 In this study, I show that investigating time-varying dynamics in articu-
39 lation offers a new account in explaining L1 Japanese speakers’ difficulty in
40 articulating English /l/ and /ɭ/. This echoes with an increasing amount of
41 evidence in L2 speech production research suggesting that L2 speech produc-
42 tion needs to be characterised with time-varying, dynamic properties (Espinal
43 et al., 2020; Schwartz and Kaźmierski, 2020; Beristain, 2022). English liquids
44 exhibit inherently dynamic acoustic and articulatory properties (Kirkham
45 et al., 2019; Campbell et al., 2010; Sproat and Fujimura, 1993). A previ-
46 ous study highlights a substantial variability in spectro-temporal properties
47 in production of word-initial English /l/ and /ɭ/ as a function of following
48 vowels between L1 English and L1 Japanese speakers (Nagamine, 2024). In
49 addition, a previous articulatory study demonstrates that L2 learners show
50 L1 transfer in articulation to produce L2 vowels that are similar to the equiv-
51 alent L1 vowels (Oakley, 2021). Extending this line of research, the current
52 study demonstrates that L1 Japanese speakers transfer articulatory strate-
53 gies from L1 Japanese /r/ to produce English /l/ and /ɭ/ and that such a
54 transfer may span across multiple segments, realised as coarticulation. This
55 articulatory transfer may account for possible specific challenges that L1
56 Japanese speakers face that hinder their “native-like” production of English
57 /l/ and /ɭ/.

58 The paper is structured as follows. The next section presents a brief
59 review of previous L2 speech production research, with a particular focus
60 on the liquid-vowel coarticulation in Section 2. It is followed by the de-
61 tails of the experiment in Section 3, which presents an articulatory study
62 using ultrasound tongue imaging. The ultrasound data are analysed quan-
63 titatively through a combination of statistical techniques as shown in Sec-
64 tion 3.5, including the Principal Component Analysis (PCA), the Functional
65 PCA, and the Bayesian hierarchical regression analysis. Finally, the results
66 are presented in Section 4, followed by a discussion of findings around the
67 coarticulatory properties involved in the production of English liquids in

68 Section 5. This paper is accompanied by an online repository that deposit
69 supplementary materials, including the code and data sets used in the study,
70 supplementary data visualisation and additional data analysis, available at
71 <https://osf.io/h3zf4/>.

72 **2. Previous research**

73 *2.1. Theoretical models in L2 speech learning*

74 The broader context of the current research lies in the acquisition of L2
75 segments by adult L2 learners (Flege, 1995; Flege and Bohn, 2021; Best and
76 Tyler, 2007). L2 learners must learn to perceive and produce L2 sounds ac-
77 curately to fully acquire L2 speech (Archibald, 2021). Models of L2 speech
78 learning commonly posit that the L1 phonological/phonetic structure influ-
79 ences L2 speech perception and production. It is also hypothesised that
80 L2 perceptual accuracy influences the degree of production accuracy in L2
81 speech (Chang, 2019). In the context of L1 Japanese speakers’ production of
82 English liquids, it could, therefore, be argued that the primary source of dif-
83 ficulty is the underlying interaction between the liquid categories in English
84 and Japanese (Archibald, 2021; Bradlow et al., 1999; Chang, 2019).

85 The Speech Learning Model (SLM; Flege, 1995; Flege and Bohn, 2021)
86 posits that L1 and L2 categories co-exist in a common phonetic space in
87 long-term memory. The core mechanism in L2 speech learning is ‘equiv-
88 alence classification’ between the L2 sounds and the closet L1 categories,
89 meaning that L2 speech learning takes place by L2 learners classifying L2
90 sounds as instances of the equivalent L1 categories based on ‘perceived sim-
91 ilarity’ of phonetic detail at the level of position-sensitive allophones (Flege,
92 1995; Flege and Bohn, 2021). Given that many production errors in L2
93 speech have a perceptual basis, SLM assumes that perception and produc-
94 tion are linked (although without a strict precedence), arguing that phonetic
95 categories store language-specific realisation rules that specify articulatory
96 commands to produce L2 sounds (Flege, 1995). The realisation rules dic-
97 tate “the amplitude and duration of muscular contractions that position the
98 speech articulators in space and time” (Flege, 1992, p. 165). This suggests
99 that L2 learners who have a better perceptual accuracy might be able to ar-
100 ticulate L2 sounds accurately than those who have a less accurate perceptual
101 realisation, resulting in a greater accuracy in their L2 speech production.

102 The Perceptual Assimilation Model for Second Language Learning (PAM-
103 L2; Best, 1995; Best and Tyler, 2007) is a theory of speech perception, which

104 posits that L2 learners directly perceive articulatory gestures of the speaker
105 for a given L2 sound and use the gestural information to phonologically
106 assimilate L2 phonemic contrasts into L1 categories. According to PAM-L2,
107 L1 and L2 sounds are identified as functionally equivalent to each other when
108 they are recognised as involving the same set of articulatory gestures (Best
109 and Tyler, 2007). PAM-L2 shares a similar view with SLM that L2 sounds are
110 perceived in relation to L1 categories, but differs in the levels of assimilation
111 (i.e., phonetic in SLM vs both phonetic and phonological in PAM-L2) and
112 the smallest unit postulated for L2 speech learning (i.e., positional allophones
113 for SLM and articulatory gestures for PAM-L2). Despite being a theory of
114 speech perception, PAM-L2 might be potentially useful in conceptualising
115 interactions between L1 and L2 articulation because of its direct reference to
116 articulatory gestures (Nagle and Baese-Berk, 2022; Bradlow et al., 1999).

117 2.2. Japanese speakers' perception and production of English /l/ and /ɹ/

118 L1 Japanese speakers perceive English /l/ and /ɹ/ in relation to the
119 Japanese /r/ category. This corresponds to the scenario of learning similar
120 phones in SLM and a Single-Category scenario in PAM-L2 (Flege and Bohn,
121 2021; Aoyama et al., 2004; Guion et al., 2000; Best and Strange, 1992), in
122 which both models predict a substantial difficulty in perception (but see Hat-
123 tori and Iverson (2009) for a different prediction under the PAM-L2 frame-
124 work). More specifically, L1 Japanese speakers perceive English /l/ and /ɹ/
125 with a different perceptual similarity, such that English /ɹ/ is perceived to
126 be more dissimilar to Japanese /r/ than is English /l/ (Aoyama et al., 2004).
127 As a result, it is easier for them to learn to perceive English /ɹ/ than En-
128 glish /l/ (Aoyama et al., 2004; Hattori and Iverson, 2009). Furthermore,
129 previous perception training studies demonstrate that such asymmetry in L1
130 Japanese speakers' perception of English /l/ and /ɹ/ can also be observed in
131 production, in which the magnitude of improvement in production is larger
132 for English /ɹ/ than for English /l/ (Bradlow et al., 1997; Shinohara and
133 Iverson, 2018).

134 L1 Japanese speakers' production of English /l/ and /ɹ/ show different
135 characteristics from that of L1 English speakers. Acoustically, they have
136 difficulty in lowering the F3 frequencies for English /ɹ/ as low as that of
137 L1 English speakers. F3 is suggested to be a difficult acoustic cue for L1
138 Japanese speakers to learn, which may result from different cue weighting
139 patterns in perception (Aoyama et al., 2019; Saito and Munro, 2014; Saito
140 and van Poeteren, 2018; Iverson et al., 2003). As a consequence, L1 Japanese

141 speakers seem to distinguish English liquids along the F2 dimension instead
142 of the more difficult F3, e.g., lower F2 for English /ɹ/ than for English /l/
143 (Aoyama et al., 2019; Nagamine, 2024).

144 The acoustic findings that L1 Japanese speakers rely on F2 more than
145 F3 are often used to infer articulatory configurations in their production of
146 English liquids. Theoretically, it is hypothesised that L2 speakers produce
147 L2 segments using the articulatory configurations for the closest L1 category
148 (Flege, 1987). Given that F2 seems to be an easier cue, it is argued that
149 L1 Japanese speakers rely more on the front-back tongue movement in dis-
150 tinguishing English /l/ and /ɹ/ than articulatory characteristics associated
151 with F3 (Aoyama et al., 2023; Saito and van Poeteren, 2018). Given this, it
152 can be argued that L1 Japanese speakers redeploy articulatory parameters
153 that already exist in L1 Japanese (i.e., degree of tongue retraction) instead
154 of acquiring new articulatory parameters, such as simultaneous constrictions
155 in the labial, palatal and pharyngeal areas (Saito and van Poeteren, 2018).

156 While the claim that L1 Japanese speakers redeploy articulatory proper-
157 ties for Japanese /r/ to produce English /l/ and /ɹ/ seems reasonable, one
158 issue here is that it does not receive any empirical support. Rather, a handful
159 of articulatory studies on the topic seems to suggest that such a view may
160 be oversimplified (Moore et al., 2018; Masaki et al., 1996). A previous study,
161 for example, shows that, while L1 American English speakers generally use
162 L1 tongue positions to articulate L2 French vowels /i, u, e, o/ that are sim-
163 ilar between the two languages, the extent of articulatory similarity seems
164 to depend on the vowel identity (Oakley, 2021). This overall suggests that
165 a closer look at articulatory data would uncover finer-grained articulatory
166 properties that may not always be transparent from the acoustic signals.

167 The complex acoustic-articulatory relationships are well-known for En-
168 glish liquids. Articulation of English /ɹ/ can, for instance, be classified into
169 two types according to the tongue shape: the tongue-tip-up ‘retroflex’ and
170 tongue-tip-down ‘bunched’ variants, that constitute the two ends of a contin-
171 uum consisting of various intermediate realisations (Delattre and Freeman,
172 1968; Zhou et al., 2008; King and Ferragne, 2020). Variability in tongue
173 shape, however, is usually not readily perceivable auditorily as far as the
174 lower three formants are concerned, as they all achieve lower F3 (Delattre
175 and Freeman, 1968). Acoustic differences indicating tongue shape variability
176 are suggested to lie in higher formants such as F4 and F5 (Zhou et al., 2008).
177 Furthermore, the degree to which palatal and pharyngeal constrictions con-
178 tributes to the F3 lowering seems to be speaker specific (Harper et al., 2020).

179 This illustrates a complex acoustic-articulatory relationship in English /ɹ/,
180 and it is indeed quite challenging to identify articulatory properties that
181 directly influence the acoustic output of low F3 (Hashi et al., 2003).

182 Such a substantial degree of articulatory variability is also found for L1
183 Japanese speakers’ production of English liquids. Masaki et al. (1996), for in-
184 stance, compared vocal tract configurations using magnetic resonance imag-
185 ing (MRI) and acoustic parameters in an intervocalic English /ɹ/ and a
186 syllable-final English /l/ produced by five L1 American English speakers and
187 nine L1 Japanese speakers with a varying degree of L2 English proficiency.
188 They found that the presence of sublingual cavity for English /ɹ/ corre-
189 lates with lowering of F3 and subsequently a more accurate identification
190 of English liquids by L1 English-speaking listeners (cf. Stevens, 2000; Alwan
191 et al., 1997). There is, however, a substantial degree of inconsistency between
192 acoustics and articulation in their study, demonstrating a substantial vari-
193 ability that exists in L1 Japanese speakers’ articulatory strategies for English
194 /l/ and /ɹ/. Similarly, using electromagnetic articulography (EMA), Moore
195 et al. (2018) found that L1 Japanese speakers use at least seven patterns of
196 midsagittal tongue shape when articulating English /l/ and /ɹ/, in which
197 participants achieved some degrees of F3 lowering for English /ɹ/ but none
198 of them showed native-like tongue shape. Overall, these findings highlight a
199 considerable variability in L1 Japanese speakers’ articulation of English /l/
200 and /ɹ/, suggesting a challenging nature of inferring articulatory properties
201 based on acoustic signals.

202 2.3. Coarticulation in L2 speech production

203 Recent L2 speech production studies increasingly suggest the importance
204 of time-varying properties for a better understanding of L2 segmental pro-
205 duction (e.g., Schwartz and Kaźmierski, 2020; Espinal et al., 2020; Beris-
206 tain, 2022). Dynamic properties in L2 speech production could span not
207 only within individual segments but also over multiple segments, manifested
208 through the process of coarticulation. Coarticulation can be defined broadly
209 as “patterns of coordination, between the articulatory gestures of neighbour-
210 ing segments, which result in the vocal tract responding at any one time to
211 commands for more than one segment” (Manuel, 1999, p. 179). While speech
212 is often seen primarily as a sequence of segmental targets, in which coarticula-
213 tory patterns may be determined by language-universal rules (Chomsky and
214 Halle, 1968), empirical evidence suggests that coarticulation is a language-
215 specific process which needs to be learnt during second language acquisition

216 (Keating, 1985; Beristain, 2022; Öhman, 1966).

217 English /l/ and /ɹ/ show different coarticulatory influences from the
218 neighbouring vowels. The degree of coarticulation is inversely correlated
219 with the degree of constraints imposed on tongue body, such that segments
220 whose tongue body is highly constrained should show a lesser degree of vo-
221 calic coarticulation, resulting in a greater degree of stability (Recasens, 2012).
222 For instance, English /ɹ/ shows a greater resistance to coarticulation than
223 English /l/ does, meaning that the tongue shape for English /ɹ/ is more sta-
224 ble regardless of the neighbouring vowels than English /l/ is (Proctor et al.,
225 2019). Also, ‘clear’ and ‘dark’ /l/s differ in the degree of coarticulatory resis-
226 tance, such that the latter shows a greater degree of coarticulatory resistance
227 than the former because of a greater degree of dorsal activity involved in ar-
228 ticulation (Recasens, 2012). Even though the degrees of ‘darkness’ differ from
229 one language/dialect to another; e.g., laterals in American English are over-
230 all ‘darker’ than that of British English, syllable-initial laterals still exhibit
231 relatively clearer realisations than the syllable-final counterparts (Recasens,
232 2012).

233 The coarticulatory patterns for English /l/ and /ɹ/ contrast with that for
234 Japanese /r/, canonically realised as an alveolar tap or flap [ɾ], whose articu-
235 lation shows a greater sensitivity to vocal context and thus a greater degree
236 of vocalic coarticulation. A greater susceptibility to phonetic contexts for the
237 articulation of alveolar taps is reported across languages including American
238 English (Derrick and Gick, 2011), Catalan (Recasens, 1991; Recasens and
239 Rodríguez, 2016, 2017) and Japanese (Sudo et al., 1982; Yamane et al., 2015;
240 Maekawa, 2023). In Catalan, the degree of lingual contact to the palate
241 varies according to the tongue height and backness of the neighbouring vowe-
242 ls, such that it is largest when flanked by high vowels [i] compared to [a]
243 or [u] (Recasens, 1991). Similarly, variability in the midsagittal tongue body
244 movement is greater for alveolar taps [ɾ] than for alveolar trills [r] in Catalan
245 (Recasens and Rodríguez, 2016). In Japanese, Maekawa (2023) demonstrates
246 using real-time magnetic resonance imaging that precise alveolar place of ar-
247 ticulation for Japanese [ɾ] correlates with adjacent vowels, arguing that the
248 global tongue body movement for [ɾ] is determined largely by neighbouring
249 vowels. Finally, comparing nonpalatalised and palatalised taps in Japanese,
250 Yamane et al. (2015) shows that nonpalatalised taps are subject to a greater
251 degree of vocalic coarticulation in tongue dorsum, arguing for lack of gestural
252 specifications for tongue dorsum. Taken together, these studies support the
253 view of Recasens (1991, p. 279) that “the positioning of the tongue body

254 does not involve much articulatory control” for alveolar taps/flaps [ɾ].

255 Although coarticulation is not fully addressed in the current theoretical
256 frameworks in L2 speech learning, their implicit theoretical assumptions
257 could possibly capture the role of time-varying properties in L2 speech learn-
258 ing. SLM, for instance, assumes that L2 speech learning takes place at po-
259 sitional allophonic level, which could potentially account for different pho-
260 netic realisations of segments due to context-specific coarticulatory influence
261 (Bradlow et al., 1999; Colantoni et al., 2015). PAM-L2 states that coartic-
262 ulation could be one of the factors that influence L2 learners’ accurate per-
263 ception of L2 sounds (Best and Tyler, 2007). In addition, previous research
264 suggests a possibility that L2 learners of English may differ in how they
265 distinguish English liquids from neighbouring vowels in perception (Wang
266 et al., 2023) and in production (Nagamine, 2024). This background overall
267 points to the importance of looking beyond the scope of individual segments
268 in understanding the L2 speech learning (cf. Beristain, 2022).

269 *2.4. Summary and research questions*

270 To summarise, the previous articulatory descriptions point to a clear dif-
271 ference between Japanese and English liquids especially in terms of coarticu-
272 lation, which could account for specific difficulties that L1 Japanese speakers
273 encounter in producing English /l/ and /ɾ/. This is a promising argument
274 that may advance our understanding of the nature of L2 speech production
275 mechanism, corroborating the recent findings in L2 speech learning research
276 arguing for the importance of dynamic properties in L2 speech learning. A
277 lack of articulatory data in this research context, however, hinders us from
278 validating this argument, as the majority of previous research has focussed
279 on acoustics. Given that the theoretical frameworks posit that the learner’s
280 L1 and L2 categories interact, it is possible that L1 Japanese speakers are
281 influenced by the coarticulatory patterns for L1 Japanese /ɾ/ when produc-
282 ing L2 English /l/ and /ɾ/. In addition, L1 Japanese speakers who have
283 better perceptual accuracy might be more likely to be able to overcome such
284 difference in coarticulation than those who have less accurate perception.

285 The objective of the current study is to better understand the articulatory
286 mechanisms involved in L1 Japanese speakers’ production of English liquids
287 by focussing on the liquid-vowel coarticulation. It considers the effects of L1
288 Japanese speakers’ perceptual accuracy as it has been shown that production
289 accuracy depends on perceptual accuracy in the previous research. In ad-
290 dition, it investigates time-varying properties in articulation spanning over

291 a word-initial liquid-vowel sequence in monosyllabic English words, taking
292 into account possible difference in liquid-vowel coarticulation by including
293 multiple vowel contexts. Specifically, the current study sets out to answer
294 the following research questions:

- 295 1. How does L1 Japanese speakers' perceptual identification accuracy in-
296 fluence their articulation of English /l/ and /ɹ/?
- 297 2. What difference can be observed in liquid-vowel coarticulatory patterns
298 in English /l/ and /ɹ/ produced by L1 Japanese speakers and L1 En-
299 glish speakers?

300 **3. Methods**

301 *3.1. Participants*

302 Data for this study are obtained from of 43 participants, including 14 L1
303 North American English speakers (11 female, 3 male) with a mean age of
304 29.05 years ($SD = 6.25$) and 29 L1 Japanese speakers (17 female, 12 male)
305 with a mean age of 19.60 years ($SD = 0.94$).

306 L1 Japanese speakers were all undergraduate students enrolled at univer-
307 sities in Japan at the time of recording. All of them identified themselves
308 as L1 speakers of Japanese growing up using only Japanese. They studied
309 English mainly through school curriculum from primary or secondary schools
310 and continued at university in Japan, with a mean length of their English
311 study being 8.82 years ($SD = 2.06$). They did not have experience of an
312 extended stay in English-speaking countries, with a mean length of overseas
313 stay being 0.65 months ($SD = 1.23$). Their mean self-perceived fluency is
314 3.71 ($SD = 1.04$) on a scale of one ("I do not speak English at all.") to
315 seven ("No problems in using English in daily life."). They rated the degree
316 of English use being 3.66 ($SD = 1.08$) on a scale between one ("I do not
317 use English at all.") to seven ("I use English every day."). Given all this
318 background information, I consider that they represent a typical population
319 of English-as-a-foreign-language (EFL) learners in Japan.

320 L1 North American English speakers were from the US ($n = 9$) and
321 Canada ($n = 5$). All of them grew up using English as a primary language
322 until at least 13 years of age and identified themselves as fluent L1 speakers
323 of North American English. They all rated their self-perceived proficiency
324 at seven whereas the degree of English use being 6.58 ($SD = 0.88$). They
325 resided in the UK at the time of recording for work or a postgraduate study.

326 Note that I consider this L1 English speaker population because it is a North
 327 American variety of English that is used as a model in English language
 328 teaching in Japan (Setter and Jenkins, 2005, p. 2).

329 3.2. Materials

330 3.2.1. Production

331 Target words for the production task consist of 16 English monosyllabic
 332 words (eight minimal pairs) that contain word-initial /l ɹ/ in three vowel
 333 environments /i/, /æ/ and /u/. The coda consonants are always bilabial
 334 or labiodental to minimise lingual anticipatory coarticulatory effects. All
 335 the English target words have been checked using Longman Pronunciation
 336 Dictionary (Wells, 2008) and ensured that they have the intended vowel en-
 337 vironment in American English. The three vowel environments /i/, /æ/ and
 338 /u/ were chosen while taking into account the maximal similarity between
 339 American English and Japanese vowels even in the case of L1 substitution
 340 that is likely to occur in L1 Japanese speakers’ production (Vance, 2008;
 341 Makino, 2009).

Table 1: Word list per vowel context

Vowel context	Words		
/i/	leap / reap	leaf / reef	leave / reeve
/æ/	lap / rap	lamb / ram	lamp / ramp
/u/	lube / rube	loom / room	

342 3.2.2. Perception

343 In order to investigate the effects of L2 perceptual accuracy on L2 speech
 344 production, I collected participants’ perceptual accuracy of English /l/ and
 345 /ɹ/. The perception task was conducted in a two-alternative forced choice
 346 (2AFC) design, in which participants heard a monosyllabic word containing
 347 either English /l/ or /ɹ/ word initially and chose what they heard by clicking
 348 one of the two buttons displayed on the computer screen. Ninety-six words
 349 (48 minimal pairs) served as experimental stimuli in this task as shown in
 350 Table 2.

351 The word list for the perception experiment was developed based on the
 352 audio recordings publicly available in Brekelmans et al. (2022). In the current
 353 study, only the word-initial liquids were used in light of the previous findings

354 that word-initial singleton liquids are difficult for L1 Japanese speakers to
355 accurately identify compared to the word-final singleton liquids (e.g., Brad-
356 low et al., 1997). The target words were chosen so that word-initial liquids
357 appeared in a range of vowel environment: front vowels /i:/, /ɪ/, /ε/, /eɪ/,
358 or /æ/ and back vowels /u:/, /ʊ/, /ɑ/, or /oʊ/.

359 In addition, the effects of lexical frequency were controlled as L1 Japanese
360 speakers are likely to have a smaller vocabulary than L1 English speakers
361 and this may influence their perception pattern (Flege et al., 1996). The
362 lexical frequency was checked using two corpora: the New JACET List of
363 8000 Basic Words (JACET list: Daigaku Eigo Kyoiku Gakkai Kihongo Kaitei
364 Tokubetsu Inkaï, 2016) and AmeE06 (Potts and Baker, 2012). The former
365 lists the most frequent 8 000 words from British National Corpus and Corpus
366 of Contemporary American English while also taking into account the word
367 occurrence in the school and university entrance exams in Japan. AmeE06 is
368 a relatively up-to-date corpus, containing 1 017 879 tokens from 500 sources
369 from texts published between 2004 and 2007, and the coverage has been
370 limited to contemporary American English, produced by the authors who
371 had been born or continuously lived for the majority of their lives in the
372 United States (Potts and Baker, 2012). In Table 2, the label ‘familiar’ means
373 that both words in a given minimal pair are included in the JACET list
374 whereas ‘unfamiliar’ refers to a pair in which one of the members is outside
375 the JACET list. Although two words, *loot* and *loom*, were not included in the
376 JACET list, they are classified as familiar words given a rather transparent
377 grapheme-phoneme relationship that may not impose substantial difficulty
378 for L1 Japanese speakers (Nogita, 2016).

379 3.3. Data collection procedure

380 Both production and perception tasks were conducted in a single experi-
381 mental session. Written consent was obtained from all participants and par-
382 ticipants were compensated with either ¥2 000 or £15. Ethics approval was
383 obtained from Lancaster University (UK), Kobe Gakuin University (Japan)
384 and Meijo University (Japan).

385 3.3.1. Production

386 Simultaneous high-speed midsagittal ultrasound tongue images and audio
387 recordings were collected from the participants. Recording was carried out in
388 a quiet room at universities in Japan for L1 Japanese speakers and in a sound-
389 proof recording booth at universities in the UK for L1 English speakers. At

Table 2: Stimuli for perception task. Familiar = both minimal pairs should be familiar with the L1 Japanese participants. Unfamiliar = one of the minimal pair words may be unfamiliar to the L1 Japanese-speaking participants. Italicised words fall outside the JACET8000 list, deemed to be unfamiliar.

Vowel	Front		Back	
	Familiar	Unfamiliar	Familiar	Unfamiliar
High	rim/limb	<i>reek/leak</i>	room/loom	<i>rude/lewd</i>
	read/lead	<i>reach/leech</i>	root/loot	<i>ruse/lose</i>
Mid	raid/laid	<i>rake/lake</i>	road/load	<i>robe/lobe</i>
	red/led	<i>raise/laze</i>		
Low	right/light	<i>rife/life</i>		
	rack/lack	<i>rice/lice</i>	rock/lock	<i>rot/lot</i>
	rag/lag	<i>rash/lash</i>	wrong/long	<i>rob/lob</i>

390 the time of recording L1 Japanese speakers, the Covid-19 measures were still
391 in place mandating air ventilation at all times, so there was minor fan noise
392 in the audio recording for some of the L1 Japanese speakers' data.

393 During the recording, participants sat in front of a laptop computer and
394 read aloud the target words displayed on the laptop screen in isolation. They
395 wore an UltraFit, a plastic headset to stabilise the ultrasound probe under-
396 neath participants' lower jaw relative to head movement (Spreafico et al.,
397 2018). Prior to the recording, participants bit a thin plastic plate to obtain
398 their bite plane information (Scobbie et al., 2011). Recording and stimuli
399 presentation was made on Articulate Assistant Advanced software version
400 220.4.1 (Articulate Instruments, 2022). Audio signals were recorded with an
401 Opus 55 MK ii condenser microphone attached to the UltraFit headset, pre-
402 amplified and digitised at 44.1 kHz with 16-bit quantisation using a Sound
403 Device USB Pre-2 audio interface. The parameters in ultrasound recording
404 setting were fixed for each speaker throughout the recording session but opti-
405 mised for each speaker due to differences in vocal tract size and image clarity,
406 within a range of probe frequency between 2-4 MHz, depth between 80-90
407 mm and field of view between 80-100%, achieving a frame rate of approxi-
408 mately 80 frames per second.

409 During the experiment, I controlled L1 Japanese speakers' language mode
410 by differentiating the language of instructions (Grosjean, 2008). I gave in-
411 structions in Japanese in the first half of the experiment, including briefing,

Table 3: The number of tokens produced by participants.

Vowel context	English /l/		English /ɹ/	
	L1 Japanese	L1 English	L1 Japanese	L1 English
/i/	301	193	301	194
/æ/	295	193	305	192
/u/	197	130	199	130

412 headset fitting and recording of Japanese words (not presented in this study).
413 Then, I switched the language of instruction into English and did a small con-
414 versation activity with the L1 Japanese participant, in which I asked them
415 simple questions such as “Where are you from?”, “What do you study at
416 university?”, “What do you like about the university?” etc. I then started
417 recording the English words while still providing instructions in English and
418 continued to do so in the following perception task. All instructions for L1
419 North American English speakers were given entirely in English.

420 Participants produced each word three to five times, resulting in a total
421 of 2 630 tokens of word-initial English liquids analysed in this study. This
422 includes 1 309 tokens of English /l/ and 1 321 tokens of English /ɹ/. The
423 detailed breakdown of the number of tokens is shown in Table 3. The pro-
424 duction task took up to 60 minutes for L1 Japanese speakers and 30 minutes
425 for L1 English speakers.

426 3.3.2. Perception

427 For the perception task, the participants heard a stimulus through head-
428 phones and chose the word they heard by clicking one of the buttons displayed
429 on the screen showing minimal pair words orthographically. The experiment
430 was set up and conducted using Gorilla (Anwyl-Irvine et al., 2018). The
431 perception task started with a trial session, in which they heard nine words
432 that did not contain word-initial English liquids to familiarise themselves
433 with the procedure. Feedback was provided in this practice session, in which
434 a smile face was displayed if their response was correct and a frown face if
435 incorrect. The main task phase contained four blocks, each of which con-
436 sisted of 24 tasks, covering all 96 tokens, with a short break between each
437 block. No feedback was provided in the main test phase. The order of stimuli
438 presentation was randomised between speakers within each block.

439 *3.4. Production data processing*

440 Tongue spline was estimated from ultrasound tongue images using the
441 DeepLabCut (DLC) plug-in via the AAA software version 220.5.1 (Articu-
442 late Instruments, 2022; Wrench and Balch-Tomes, 2022). The liquid-vowel
443 intervals were delimited based on forced alignment using Montreal Forced
444 Aligner version 2.0.6 (McAuliffe et al., 2017), which were later corrected
445 based on visual inspections wherever necessary using Praat (Boersma and
446 Weenink, 2022). Segmentation procedure for English liquid tokens is the
447 same as in a previous study (see Nagamine (2024) for detailed explanations).

448 Before tongue spline estimations, six participants out of 56 were excluded
449 from the analysis due to poor imaging quality, for whom no tongue spline
450 fitting was carried out. After the tongue spline estimation for the remain-
451 ing 50 participants, I further identified inaccurate tongue spline estimation
452 by plotting tongue splines using R version 4.3.2 (R Core Team, 2023). At
453 this stage, tongue spline estimation for additional six speakers appeared to
454 be problematic who were thus removed from further analysis. In addition,
455 one speaker was further excluded due to overall poor audio recording qual-
456 ity where no reliable acoustic segmentation was possible. As a result, the
457 data obtained from 43 participants were analysed, including 14 L1 North
458 American English speakers and 29 L1 Japanese speakers.

459 In the current analysis, I tracked the tongue movement in the interval
460 from the 350 ms prior to the acoustic onset of the word-initial liquid and
461 to acoustic vowel offset, which was deemed to be the most optimal interval
462 length after explorations in order to account for a mismatch between acoustic
463 and articulatory onsets (Palo et al., 2014; Ying et al., 2012). The interval
464 sufficiently captures a divergence of tongue shape for word-initial liquids from
465 the ‘mean rest posture’ as the participants read the stimuli words in isolation
466 and no preceding articulatory movement was recorded before the utterance
467 (Palo et al., 2014; Wilson and Kanada, 2014).

468 Finally, L1 Japanese speakers were grouped according to their perceptual
469 identification accuracy using the Gaussian mixture models (GMMs) with the
470 *Mclust* package on R (Fraley et al., 2023). GMMs is a clustering technique
471 that identifies the optimal number of independent Gaussian probability den-
472 sities underlying the data using the Bayesian Information Criterion (Winter,
473 2019). GMMs suggest two clusters for L1 Japanese speakers, in which the
474 first cluster consists of nine speakers and the second contains the remaining
475 20 speakers. The details of the participants’ performance are summarised in
476 Table 4. Overall, while both groups perform better for identifying English

Table 4: Mean proportion correct for perceptual identification task for each of the lexical and segmental categories (in %). Values in parenthesis indicate standard deviation.

	Familiar /l/	Unfamiliar /l/	Familiar /ɹ/	Unfamiliar /ɹ/
Cluster 1	52.94	35.71	52.47	44.44
($n = 9$)	(7.64)	(5.05)	(6.28)	(8.33)
Cluster 2	65.86	52.74	79.95	80.17
($n = 20$)	(20.06)	(23.43)	(13.79)	(14.52)

477 /ɹ/ than English /l/, cluster 2 outperforms cluster 1 across all four condi-
478 tions. Cluster 1 is also influenced by the lexical status to a greater extent
479 than cluster 1. These observations generally agree with the findings in the
480 previous study (Flege et al., 1996). In the remaining part of the current
481 study, I consider cluster 1 consisting of the intermediate learners of English
482 (“Intermediate”) and cluster 2 the advanced learners (“Advanced”).

483 3.5. Dynamic ultrasound analysis

484 Articulatory analysis in this study investigates time-varying changes in
485 midsagittal tongue shape. This means that it is necessary to identify (1)
486 spatial and (2) temporal properties that characterise liquid-vowel coartic-
487 ulation. First, the Principal Component Analysis (PCA) was run on the
488 midsagittal tongue splines to identify key spatial dimensions along the mid-
489 sagittal tongue shape. PCA can be considered to be a systematic approach
490 to identify major variation in articulation used in previous research using
491 ultrasound (e.g., Turton, 2017; Nance and Kirkham, 2022; Slud et al., 2002).
492 The 11 sets of x/y coordinates on the tongue splines were exported from
493 ultrasound images using DLC in the AAA software throughout the analysis
494 window (i.e., from the 350ms prior to the acoustic liquid onset to the vowel
495 offset). They were then within-speaker normalised into z -scores to facilitate
496 cross-speaker comparison. All the z -normalised x/y coordinate values in the
497 analysis window for all tokens for all speakers were, then, submitted to PCA
498 using the `stats::princomp` function in R.

499 The PCA suggests five dimensions that account for greater than 5% of the
500 variance observed in the data, a threshold recommended in Baayen (2008):
501 40.15% by PC1, 29.85% by PC2, 9.76% by PC3, 7.12% by PC4, and 5.16%
502 by PC5. Figure 1 suggests that PC1 captures tongue dorsum raising and its
503 associated variation along tongue root, whereas PC2 indicates a joint tongue

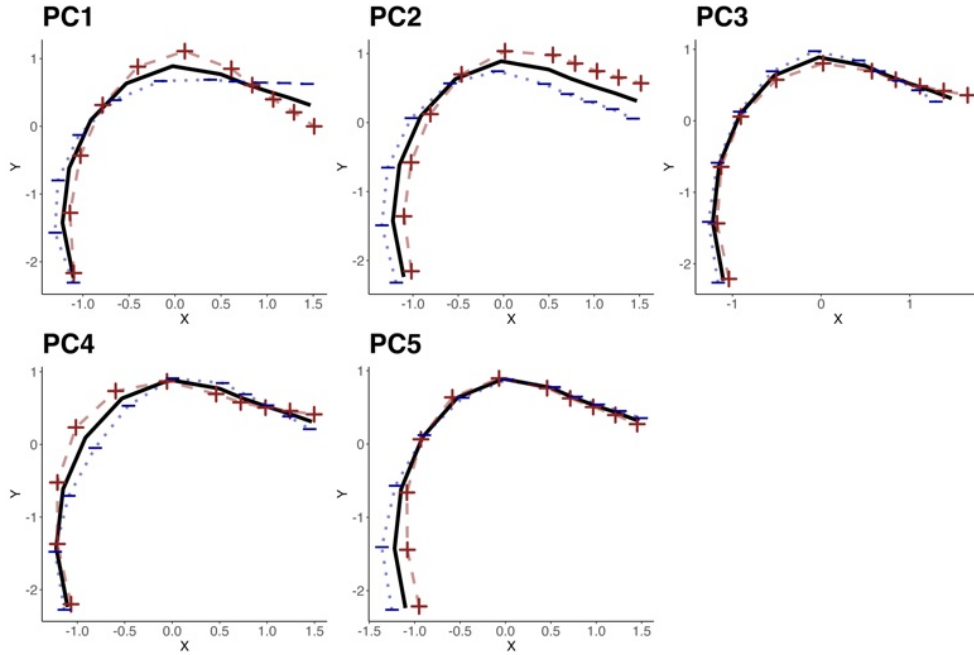


Figure 1: Variation captured by PCs 1 to 5. PCs 1 and 2 jointly account for approximately 70% of variance in the data. Variation is expressed by adding (red dashed line with the ‘+’ symbol) and subtracting (blue dotted line with the ‘-’ symbol) each PC loading using its associated standard deviation from the mean tongue curve (black solid line).

504 movement along the height and retraction dimension. Subsequent PCs ac-
 505 count for a minor variation around the tongue tip (PC3), tongue body (PC4)
 506 and tongue root (PC5). Since the first two PCs jointly capture nearly 70%
 507 of the variance in the data, I retain PCs 1 and 2 only in the subsequent
 508 analysis. Of the two PCs in particular, this article presents the findings
 509 along the PC1 dimension that captures the tongue dorsum movement, as it
 510 accounts for the largest variation of the data and is the most relevant articula-
 511 tory dimension determining the degree of vocalic coarticulation in Japanese
 512 and English liquids (e.g., Recasens, 1991; Proctor et al., 2019; Maekawa,
 513 2023). The PC2 analysis is included in the online supplementary materials
 514 (<https://osf.io/h3z4/>) as I do not have theoretical predictions regarding
 515 the PC2 dimension.

516 The PCA above associates a unique number (i.e., a PC score) along the
 517 PC1 dimension with tongue splines from each ultrasound frame. Since an

518 ultrasound video is a sequence of static ultrasound images, temporal changes
519 in the spatial articulatory properties can be inferred by tracking changes in
520 PC scores over time. Such dynamic changes in PC scores are expressed as
521 time-varying trajectories for each token as shown in the top panel in Figure 3,
522 with x -axis indicating proportional time from the onset to the offset of the
523 analysis window and y -axis the PC scores.

524 Main trends in the time-varying trajectories were identified using the
525 Functional Principal Component Analysis (FPCA) via the `fdapace::FPCA`
526 function (Zhou et al., 2022). FPCA takes functional data (e.g., time-varying
527 trajectories) as input and summarises key variability observed in the input
528 contours into numeric values (FPC scores), allowing for a further statistical
529 analysis (e.g., Asano and Gubian, 2018; Cronenberg et al., 2020). This is
530 conceptually similar to an approach using Discrete Cosine Transform (DCT)
531 in that it achieves a dimensionality reduction of the input data by summaris-
532 ing it into a number of coefficients (Watson and Harrington, 1999). DCT has
533 been used in previous ultrasound research in which the second coefficient ex-
534 plaining dynamic changes in midsagittal tongue shape (Kirkham and Nance,
535 2022).

536 The FPCA analysis here identifies two FPCs jointly accounting for 81.70%
537 of the data: FPC1 (57.00%) and FPC2 (24.70%). The analysis here focusses
538 on FPC1, as it explains the largest proportion of variance in the data and
539 it captures between-group variability in FPC scores across vowel contexts,
540 which is the most relevant dimension in this study, as shown in Figure 2 (cf.
541 Cronenberg et al., 2020).

542 3.6. Statistical analysis

543 The dynamic ultrasound analysis so far summarises the liquid-vowel coar-
544 ticulation along the tongue dorsum dimension in the FPC1 scores, allowing
545 for a quantitative analysis of the coarticulatory patterns. For this, I per-
546 formed the Bayesian hierarchical modelling using the `brms::brm` function
547 (Bürkner, 2017). I fitted separate models for each segment (i.e., English
548 /l/ and English /ɹ/) for a greater ease in model interpretation. The model
549 specifications include fixed effects of (1) vowel context (three levels: /æ/, /i/
550 and /u/), (2) speaker group (three levels: L1 English, Intermediate and Ad-
551 vanced) and an interaction between them in order to predict the FPC1 scores.
552 The factor variables are treatment coded, with the baseline level being /i/
553 for the vowel context and L1 English speakers for the speaker group.

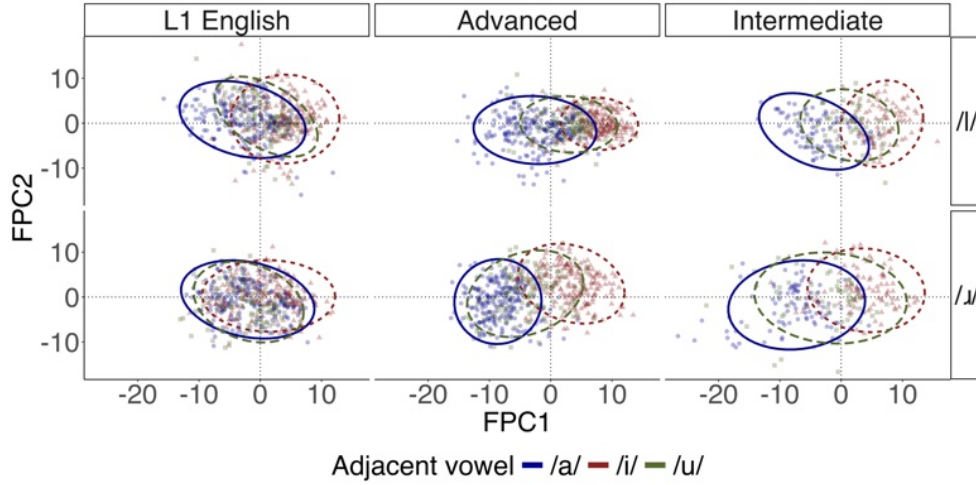


Figure 2: Variation explained by FPC1 (on x axis) and FPC2 (on y axis) for PC1. Speaker groups are indicated in columns and liquid segments in rows. Ellipsis shows multivariate t -distribution based on the 95% confidence interval. Dotted lines represent zero on each axis.

554 The random effect structure includes by-speaker random slopes and in-
 555 tercepts for each vowel condition and by-item random slopes and intercepts
 556 for each speaker group. Since I did not have much *a priori* information about
 557 possible FPC values, I used weakly Gaussian informed priors to allow for a
 558 wide range of possible parameter values. Priors are specified using normal
 559 distribution for the intercept ($\mu = 0, \sigma = 20$), regression coefficients ($\mu =$
 560 $0, \sigma = 100$), standard deviations associated with random effects ($\mu = 0, \sigma$
 561 $= 10$) and standard deviations for the likelihood function ($\mu = 0, \sigma = 10$).
 562 I used the LKJ-correlation prior (i.e., `lkj(2)`) for correlation coefficients for
 563 the interaction term (Vasishth et al., 2018; Franke and Roettger, 2019). The
 564 model specification of the full model is:

```
565 brm(FPC1 ~ vowel + group + vowel:group + (1 + vowel | speaker) +
566 (1 + group | word))
```

567 The interaction between the vowel contexts and the speaker groups is
 568 of particular interest for the current study, as it is expected that the FPC1
 569 scores would vary depending on the speaker groups. If the perceptual accu-
 570 racy influences L1 Japanese speakers' liquid-vowel coarticulation, then the

571 Intermediate and Advanced L1 Japanese speakers would exhibit different pat-
572 terns of FPC1 values. Similarly, L1 Japanese speakers would exhibit different
573 FPC1 values across vowel contexts from L1 English speakers if they employ
574 the coarticulatory strategy for L1 Japanese /r/ to produce L2 English /l/
575 and /ɹ/. The magnitude of the interaction effect is evident in the subsequent
576 analysis (see Sections 4.2.2 and 4.2.3), but additional model comparisons are
577 included in the online supplementary materials.

578 The models were run using four sampling chains for 12 000 iterations
579 with a warm-up period of 2 000 iterations for each model, resulting in ap-
580 proximately 40 000 samples for each parameter model. These specifications
581 were decided in order to obtain sufficient sample size for a reliable estima-
582 tion of posterior distribution (Vasishth et al., 2018; Kruschke, 2015). The
583 index for model convergence, *Rhat*, was 1.00 for all parameters across mod-
584 els, suggesting that all models converged successfully. The assessment of
585 each model was further carried out through posterior predictive check using
586 *brms::pp_check*. Posterior predictive check shows how closely simulated pos-
587 terior distribution aligns with actual observed data (Vasishth et al., 2018).
588 Based on visual inspections of the generated plot, I conclude that there is
589 no substantial divergence of the simulated posterior distributions from the
590 distributions of the actual data. Finally, I conducted prior sensitivity anal-
591 ysis to check the stability of the likelihood function of the Bayesian model
592 against perturbations in prior specifications using the *priorsense* package, in
593 which no prior-likelihood conflicts were diagnosed (Kallioinen et al., 2023).

594 The Bayesian posterior distribution is assessed based on the 95% high-
595 est density interval (HDI), the narrowest interval containing the probability
596 mass in the posterior distribution indicating the degree of uncertainty for
597 parameter estimations (McElreath, 2016). The HDI informs us of the degree
598 of uncertainty for a given contrast; values in the HDI are more probable given
599 the data than those outside the interval. In the statistical analysis section,
600 I first show the median and HDIs of the FPC1 values for each level instead
601 of directly reporting regression coefficients. This allows for a greater ease in
602 interpreting the results and a holistic evaluation of the overall tendency.

603 More direct evidence to the research questions is then provided by mod-
604 elling the posterior distributions encoding the difference in FPC1 values be-
605 tween conditions using the *emmeans* package (Lenth et al., 2022). This
606 includes (1) differences in FPC1 values between groups within each vowel
607 context (RQ1) and (2) differences in FPC1 values between vowel contexts
608 within each speaker group (RQ2). In this comparison, the smaller the dif-

609 ference in FPC values is, the closer the difference would be to zero and thus
610 the 95% HDI more likely contains zero in its distribution. In addition, the
611 probability of direction (PD) is calculated for each credible interval. The PD
612 is an index of effect existence, ranging from 0.50 to 1.00, in which the smaller
613 the PD is (and thus the closer it is to 0.50), the greater uncertainty the effect
614 is in a given direction and as a consequence the greater uncertainty whether
615 such an effect exists (Makowski et al., 2019).

616 4. Results

617 4.1. *Dynamic changes of midsagittal tongue shape*

618 Recall that the articulatory dimension that is being investigated in this
619 study is tongue dorsum raising. The tongue dorsum movement is inferred
620 by tracking time-varying changes in PC1 values during the analysis window.
621 As the top panel in Figure 3 shows, L1 English speakers show a relatively
622 converged cluster of trajectories during the liquid interval, whereas the two
623 groups of L1 Japanese speakers (i.e., Intermediate and Advanced) exhibit
624 distinct trajectory patterns that diverge from each other near the liquid on-
625 set. This suggests a greater degree of vocalic coarticulation for L1 Japanese
626 speakers than for L1 English speakers. The two groups of L1 Japanese speak-
627 ers seem to exhibit similar liquid-vowel coarticulatory patterns for English
628 /l/ and /ɾ/ judging from similar trajectory patterns. L1 English speakers,
629 on the other hand, show slightly different degrees of trajectory convergence
630 such that they are more converged for English /ɾ/ than for English /l/.

631 The bottom panel in Figure 3 shows reconstructed time-varying PC1
632 trajectories based on the FPC1 scores. Each thin line corresponds to an
633 individual token associated with a unique FPC1 score. Overall, the recon-
634 structed trajectories suggest that FPC1 mainly captures the vowel contextual
635 effects in the degree of tongue dorsum raising, in which higher FPC1 values
636 correspond to higher PC1 values and thus a greater degree of tongue dor-
637 sum raising. The height of trajectories for L1 Japanese speakers are clearly
638 separated according to the vowel contexts, in which higher FPC1 values are
639 encoded as trajectories associated with the /i/ context (shown in red in Fig-
640 ure 3), constantly showing higher PC1 values compared to trajectories for
641 the /u/ (green) and /æ/ (blue) contexts. Trajectories for L1 English speak-
642 ers, on the other hand, are separated less clearly across the vowel contexts,
643 although English /l/ seems to exhibit a clearer separation of the trajectories
644 than English /ɾ/.

645 Overall, the dynamic analysis shown here provides qualitative evidence re-
646 garding by-group differences in the liquid-vowel coarticulation patterns across
647 vowel contexts in tongue dorsum height. L1 Japanese-speaking groups ex-
648 hibit a greater variation across vowel contexts than L1 English speakers. L1
649 English speakers also seem to show a greater stability for English /ɹ/ than
650 for English /l/.

651 *4.2. Quantifying contrasts in trajectory shapes: Bayesian hierarchical regres-* 652 *sion analysis*

653 The section above illuminates some qualitative differences in the liquid-
654 vowel coarticulation patterns between the three speaker groups. In order
655 to provide direct quantitative evidence to the research questions, I analyse
656 the trajectory patterns encoded in FPC1 scores using Bayesian hierarchical
657 regression modelling. I first present the median values and posterior dis-
658 tributions for FPC values for each speaker group across vowel contexts to
659 understand the overall trends. I then examine the effects of (1) speaker
660 groups (i.e., reflecting differences in perceptual accuracy) and (2) vowel con-
661 texts by modelling the difference in FPC1 values and compare them across
662 conditions.

663 *4.2.1. Overall trend in estimated FPC1 values*

664 I first provide an overview of the statistical analysis. Figure 4 visualises
665 posterior distributions for estimated FPC1 values and Table 5 summarises the
666 median, the lower and upper limit of each distribution and the probability of
667 direction. Figure 4 shows that the FPC1 values are higher in the /i/ context,
668 followed by the /u/ and /æ/ contexts. Estimated median FPC1 values for
669 L1 English speakers are closer to zero compared to that of L1 Japanese
670 speakers. The HDIs for L1 English speakers contain zero for both /l/ and
671 /ɹ/ across all vowel contexts except for /l/ in the /i/ context, which suggests
672 that their tongue dorsum movement is largely similar across vowel contexts.
673 The case of /l/ in the /i/ context could reflect differences in coarticulatory
674 effects between English /l/ and /ɹ/, although the probability of direction
675 is still smaller than that of L1 Japanese speakers and the distribution is
676 close to zero in Figure 4. This overall suggests that L1 English speakers
677 show relatively stable coarticulatory patterns regardless of vowel contexts,
678 with English /l/ showing a slightly greater degree of susceptibility to vowel
679 contexts than English /ɹ/

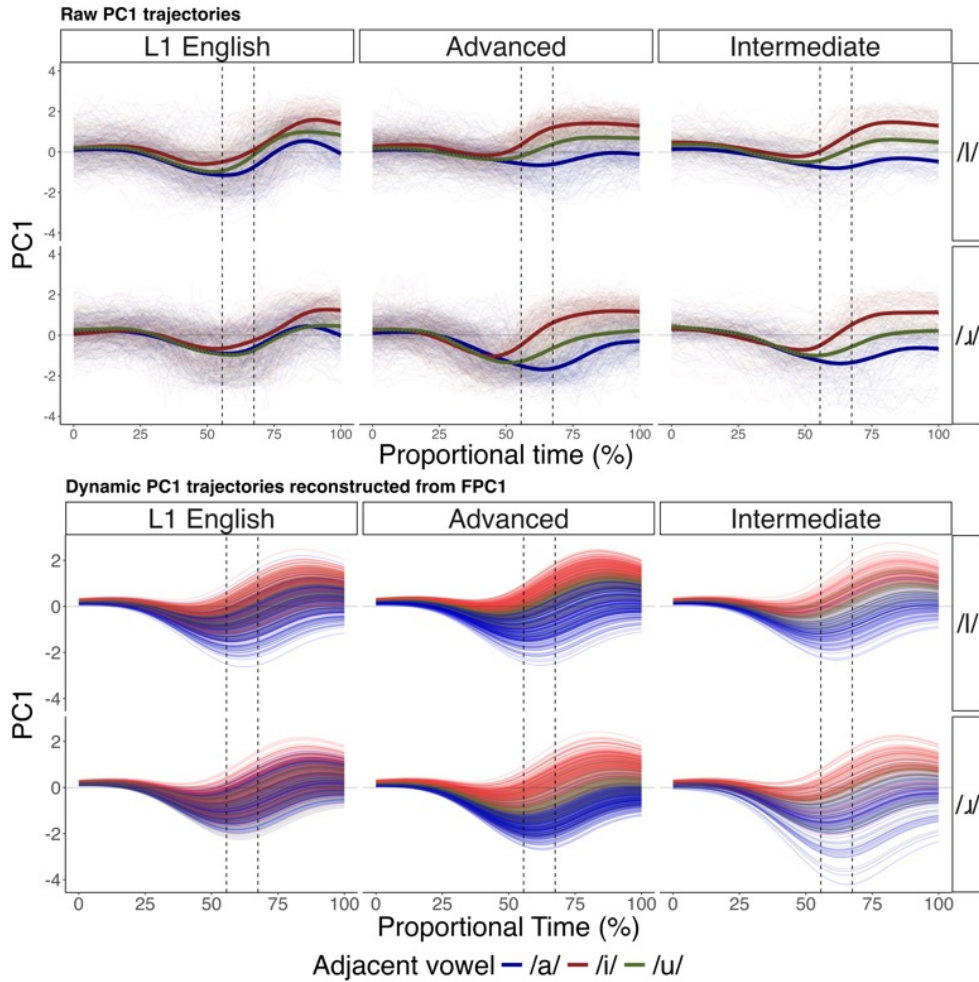


Figure 3: Time-varying changes of PC1 during the analysis window based on the raw data (top) and reconstructed trajectories from the FPC1 information (bottom). For the raw data, by-token trajectories are shown in faint thin lines on which aggregated smoothed trajectories are superimposed. Time is expressed proportionally from 0% (350 ms prior to acoustic liquid onset) to 100% (acoustic vowel offset). In each panel, columns indicate participant groups (left: L1 English speaker, middle: L1 Japanese-speaking intermediate learners of English, right: L1 Japanese-speaking advanced learners of English) and rows indicate segments (upper row: English /l/, lower row: English /ɹ/). Each trajectory represents changes of PC scores in different vowel contexts: /æ/ (blue), /i/ (red) and /u/ (olive green). The two vertical dashed lines represent mean acoustic interval for the liquid segment.

680 In contrast, the two groups of L1 Japanese speakers show the FPC1 scores
681 that are distinctively diverged from zero. This is particularly notable for /ɹ/
682 in the /æ/ context, in which the median values are substantially lower than
683 that of L1 English speakers. A similar tendency can be observed for both /l/
684 and /ɹ/ in the /i/ context, in which the median values and the overall distri-
685 butions are higher for the Intermediate and the Advanced speakers than for
686 L1 English speakers. The FPC1 values in the /u/ context are generally close
687 to zero and similar to that of L1 English speakers. Finally, the two groups
688 of L1 Japanese speakers behave similarly with little notable differences.

689 Overall, these results suggest that L1 English speakers are less susceptible
690 to vowel contexts in producing English liquids than L1 Japanese speakers,
691 with a greater stability in tongue dorsum movement observed for English /ɹ/
692 than for English /l/. L1 Japanese speakers, on the other hand, are influenced
693 by the vowel contexts when producing English liquids, especially in the /i/
694 and /æ/ contexts given that their FPC1 distributions distinctively diverge
695 from zero.

696 4.2.2. *Between-group differences*

697 As set out in research question 1, if the perceptual accuracy influences
698 tongue movement, it can be expected that the Intermediate and Advanced
699 group speakers would exhibit differences in dynamic PC1 trajectory patterns,
700 expressed as FPC1 scores. This is examined by quantifying the difference in
701 the FPC1 values between speaker groups, which is visualised in Figure 5 and
702 summarised in Table 6.

703 Overall, the two groups of L1 Japanese speakers show little difference in
704 FPC1 values across vowel contexts for both English /l/ and /ɹ/, suggesting
705 that they do not differ from each other in terms of the tongue dorsum move-
706 ment. The 95% HDI includes zero in all vowel contexts as shown in Figure
707 5. The probability of direction values in Table 6 are constantly small and
708 closer to 0.50, representing a great level of uncertainty in the presence of the
709 between-group difference in FPC1 values.

710 L1 English speakers differ from the two groups of L1 Japanese speakers,
711 but the magnitude varies depending on the liquid segment and the vowel con-
712 text. For English /l/, L1 English speakers differ from L1 Japanese speakers
713 in the /i/ context, with the probability of direction being close to 1.00. In
714 the other vowel contexts, however, the magnitude of difference is relatively
715 small, in which the 95% HDI includes zero and the probability of direction
716 being low, indicating a greater uncertainty in the presence of the speaker

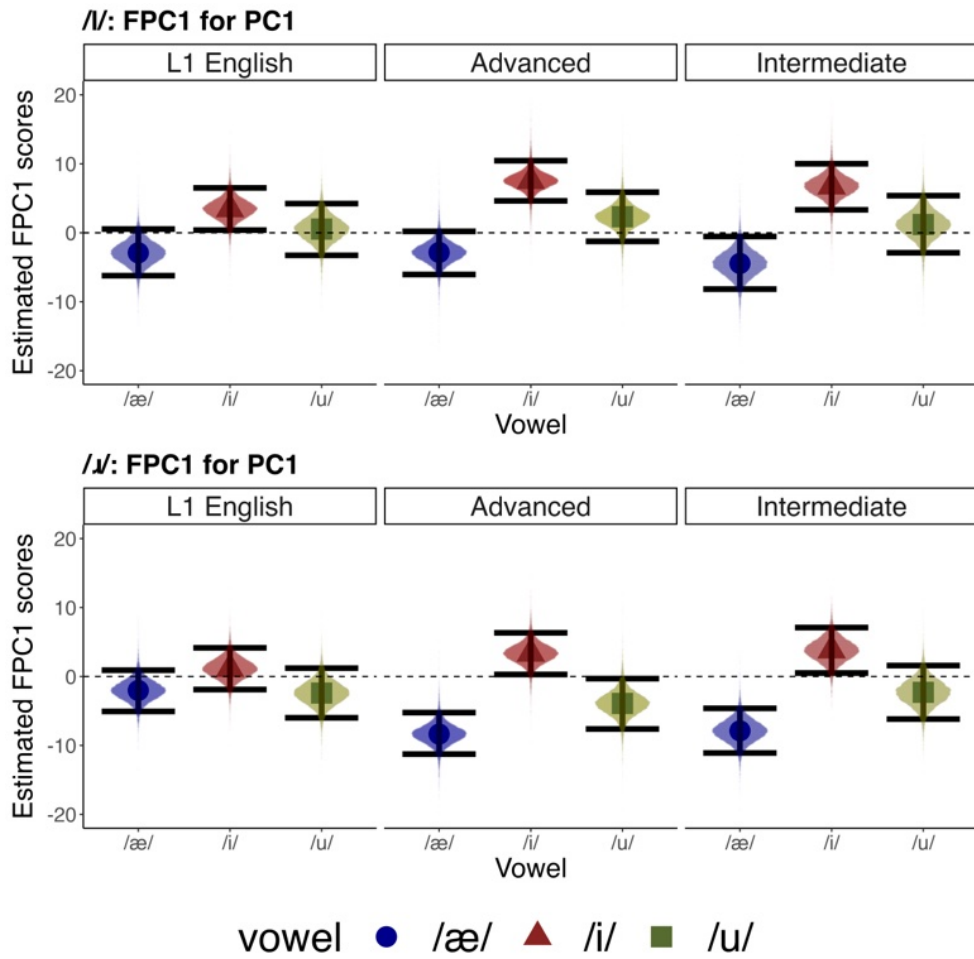


Figure 4: Posterior distributions of estimated FPC1 scores for PC1 for English /l/ (top) and /ɹ/ (bottom), showing distributions of posterior draws based on simulated values. Symbols indicate mean posterior estimates and the error bars represent 95% highest density intervals (HDIs).

Table 5: Bayesian posterior distributions of estimated FPC1 values for PC1.
HDI = highest density interval, PD = probability of direction.

Segment	Group	Vowel	Median	Lower HDI	Upper HDI	PD	
/l/	L1 English	/i/	3.45	0.39	6.50	0.98	
		/æ/	-2.91	-6.22	0.54	0.96	
		/u/	0.54	-3.27	4.24	0.62	
	Advanced	/i/	7.53	4.63	10.45	1.00	
		/æ/	-2.85	-6.06	0.21	0.97	
		/u/	2.33	-1.25	5.88	0.93	
	Intermediate	/i/	6.75	3.33	10.01	1.00	
		/æ/	-4.45	-8.17	-0.53	0.99	
		/u/	1.20	-2.90	5.38	0.75	
	/ɹ/	L1 English	/i/	1.13	-1.89	4.15	0.79
			/æ/	-2.02	-5.06	0.92	0.92
			/u/	-2.40	-5.98	1.22	0.92
Advanced		/i/	3.34	0.29	6.34	0.98	
		/æ/	-8.35	-11.25	-5.23	1.00	
		/u/	-3.91	-7.61	-0.30	0.98	
Intermediate		/i/	3.81	0.50	7.10	0.99	
		/æ/	-7.88	-11.09	-4.62	1.00	
		/u/	-2.29	-6.17	1.59	0.89	

717 group effect. For English /ɪ/, on the other hand, L1 English speakers differ
718 from L1 Japanese speakers not only in the /i/ context but also in the /æ/
719 context, with the probability of direction being 1.00 constituting strong evi-
720 dence for the presence of the difference. The /u/ context, however, does not
721 exhibit evidence for such between-group difference. Overall, these confirm
722 the qualitative observation in Figure 4 that the difference between L1 English
723 and L1 Japanese speakers is most notable in the /i/ context for both English
724 /l/ and /ɪ/, as well as in the /æ/ context for English /ɪ/ only.

725 4.2.3. Variability according to vowel context

726 The FPCA analysis visualised in Figure 3 suggests that the three speaker
727 groups differ in the extent to which the tongue dorsum movement is influ-
728 enced by the vowel context. In order to examine this vowel context effect, I
729 quantify the differences in FPC1 values between vowel contexts and compare
730 them across groups. It can be predicted that the greater the vowel context
731 effect is, the greater the difference in FPC1 values is between vowel contexts.
732 The FPC1 difference between vowel contexts are visualised in Figure 6 and
733 summarised in Table 7.

734 For English /l/, all speaker groups follow a similar pattern in the posterior
735 distributions of the FPC1 difference between vowel contexts. The probability
736 of direction is consistently high across all conditions, although L1 Japanese
737 speakers show higher values of the probability of direction, suggesting that
738 they are slightly more susceptible to the vowel context effects than L1 English
739 speakers. This overall suggests that both L1 English and L1 Japanese speak-
740 ers are influenced by the vowel context to a similar extent for articulation of
741 English /l/, with a slight indication that L1 English speakers' production is
742 less variable than that of L1 Japanese speakers.

743 For English /ɪ/, L1 English and L1 Japanese speakers show different ten-
744 dency in the way their FPC1 values differ between vowel contexts. L1 English
745 speakers generally exhibit weak evidence of such vowel context effects, with
746 the 95% HDI consistently overlapping zero across all vowel pairings (see Fig-
747 ure 6). The two groups of L1 Japanese speakers, on the other hand, show a
748 greater variability in FPC1 values across vowel contexts, with the 95% HDI
749 not including zero for all conditions (but in the /æ/-/u/ condition for the
750 Advanced speakers), and the probability of direction values span between
751 0.97 and 1.00 indicating small levels of uncertainty in the presence of the
752 vowel effects.

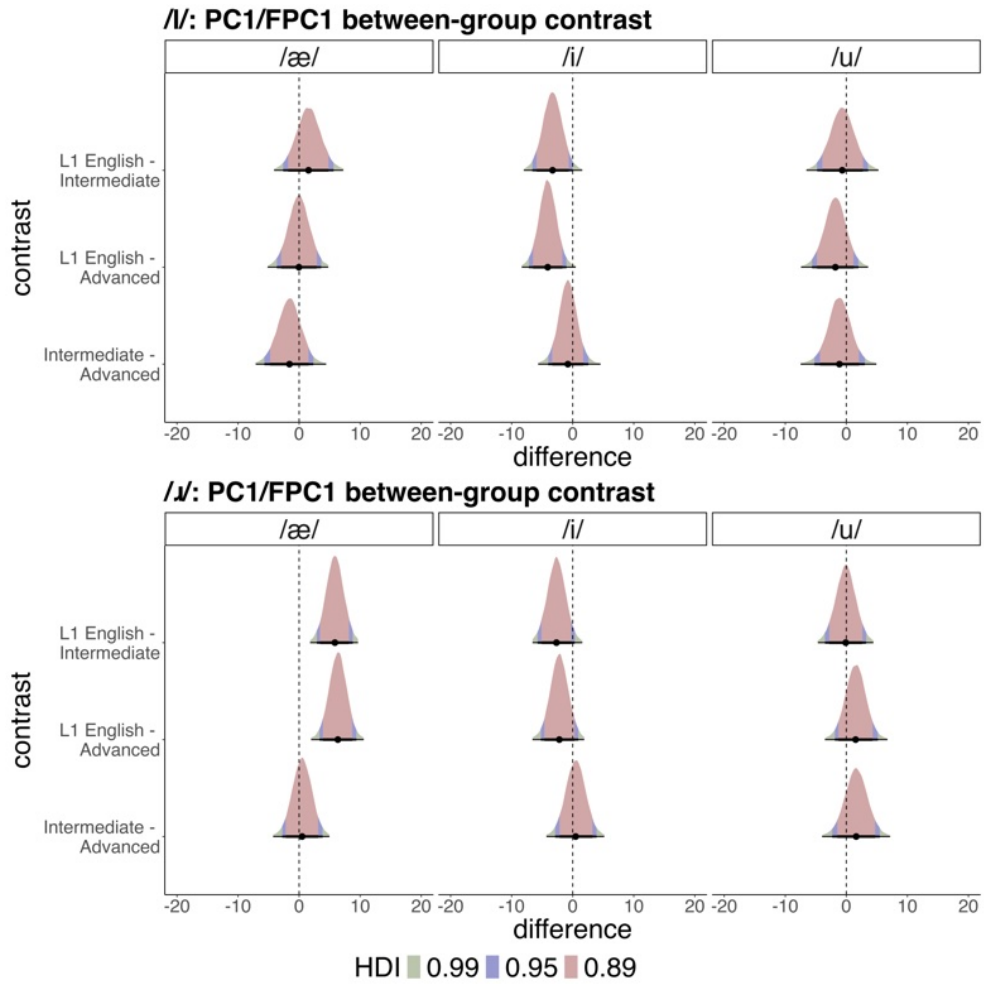


Figure 5: Posterior distributions for the degree of between-group difference in FPC1 values for the PC1 dimension for /l/ (top) and /ɹ/ (bottom). Colours show the highest density intervals (HDI) at the levels of 99% (green), 95% (blue) and 89% (pink). The black point indicates the median values for each posterior distribution with a thin and thick horizontal line corresponding to the 95% and 89% HDIs respectively.

Table 6: Bayesian posterior distributions for contrasts in FPC1 values between vowel contexts in the PC1 dimension.

Segment	Vowel	Group contrast	Estimate	Lower HDI	Upper HDI	PD
/l/	/i/	L1 English - Intermediate	-3.30	-6.60	0.02	0.97
		L1 English - Advanced	-4.10	-7.12	-0.99	0.99
		Intermediate - Advanced	-0.79	-4.06	2.53	0.70
	/æ/	L1 English - Intermediate	1.54	-2.56	5.69	0.78
		L1 English - Advanced	-0.04	-3.66	3.58	0.51
		Intermediate - Advanced	-1.58	-5.64	2.38	0.80
	/u/	L1 English - Intermediate	-0.67	-4.78	3.56	0.63
		L1 English - Advanced	-1.79	-5.58	2.00	0.84
		Intermediate - Advanced	-1.13	-5.16	3.09	0.73
/r/	/i/	L1 English - Intermediate	-2.67	-5.73	0.34	0.96
		L1 English - Advanced	-2.19	-5.28	0.86	0.92
		Intermediate - Advanced	0.48	-2.82	3.96	0.61
	/æ/	L1 English - Intermediate	5.86	2.91	8.79	1.00
		L1 English - Advanced	6.35	3.32	9.36	1.00
		Intermediate - Advanced	0.49	-2.77	3.85	0.62
	/u/	L1 English - Intermediate	-0.10	-3.49	3.24	0.52
		L1 English - Advanced	1.52	-1.89	5.16	0.81
		Intermediate - Advanced	1.62	-2.31	5.48	0.81

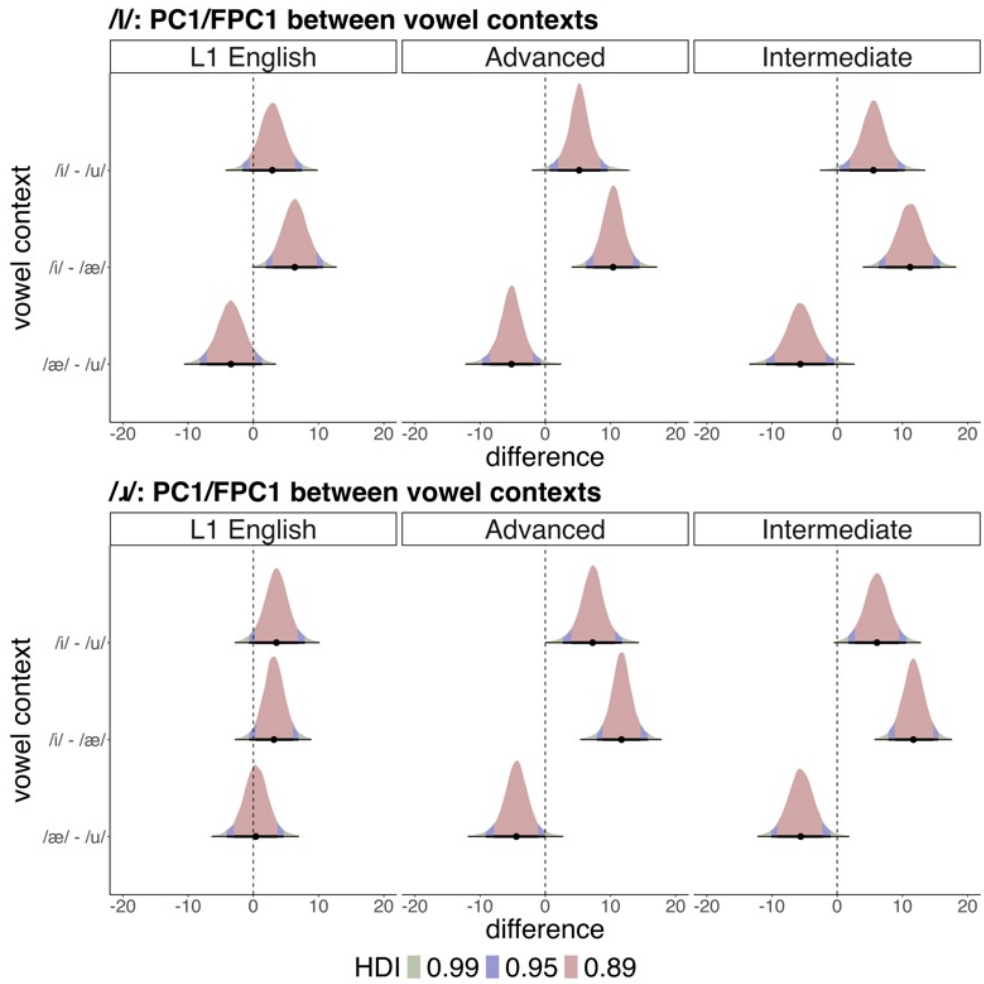


Figure 6: Posterior distributions for the degree of contrast in FPC1 values between vowel contexts for the PC1 dimension for /l/ (top) and /ɹ/ (bottom). Colours show the highest density intervals (HDI) at the levels of 99% (green), 95% (blue) and 89% (pink). The black point indicates the median values for each posterior distribution with a thin and thick horizontal line corresponding to the 95% and 89% HDIs respectively.

Table 7: Bayesian posterior distributions for contrasts in FPC1 values between vowel contexts in the PC1 dimension.

Segment	Group	Vowel contrast	Estimate	Lower HDI	Upper HDI	PD
/l/	L1 English	/i/ - /u/	2.90	-1.68	7.50	0.91
		/i/ - /æ/	6.35	1.92	10.69	0.99
		/æ/ - /u/	-3.44	-8.15	1.38	0.93
	Advanced	/i/ - /u/	5.21	0.63	9.59	0.98
		/i/ - /æ/	10.40	6.23	14.52	1.00
		/æ/ - /u/	-5.18	-9.69	-0.70	0.98
	Intermediate	/i/ - /u/	5.55	0.41	10.41	0.98
		/i/ - /æ/	11.19	6.26	15.88	1.00
		/æ/ - /u/	-5.65	-10.84	-0.37	0.98
/ɹ/	L1 English	/i/ - /u/	3.54	-0.68	7.91	0.95
		/i/ - /æ/	3.16	-0.62	6.93	0.96
		/æ/ - /u/	0.38	-4.07	4.62	0.58
	Advanced	/i/ - /u/	7.27	2.65	11.75	0.99
		/i/ - /æ/	11.70	7.82	15.75	1.00
		/æ/ - /u/	-4.43	-9.02	0.11	0.97
	Intermediate	/i/ - /u/	6.11	1.59	10.48	0.99
		/i/ - /æ/	11.69	7.78	15.48	1.00
		/æ/ - /u/	-5.59	-10.13	-1.03	0.99

753 *4.2.4. Summary*

754 The FPC1 scores on the PC1 dimension are generally higher in the /i/
755 context reflecting a greater degree of tongue dorsum raising than in the /u/
756 and /æ/ contexts. There is little evidence that L1 Japanese speakers' per-
757 ceptual accuracy influences their tongue movement (RQ1). Both L1 English
758 and L1 Japanese speakers are influenced by the vowel context to a similar
759 degree for English /l/, whereas L1 English speakers' production of English
760 /ɹ/ is less variable across vowel contexts than that of L1 Japanese speakers
761 (RQ2).

762 **5. Discussion**

763 The current study investigates word-initial liquid-vowel coarticulation
764 patterns in English produced by L1 English and L1 Japanese speakers, with
765 a particular focus on (1) the effects of L1 Japanese speakers' perceptual accu-
766 racy and (2) differences in coarticulatory patterns between L1 and L2 speech
767 production. The main articulatory dimension investigated here is tongue
768 dorsum movement, captured by PC1. Overall, while there is little evidence
769 of the effects of L1 Japanese speakers' perceptual accuracy, the analysis il-
770 luminates some key coarticulatory differences between L1 English and L1
771 Japanese speakers for both English /l/ and /ɹ/.

772 *5.1. Perceptual accuracy and liquid-vowel coarticulation*

773 Research Question 1 asked how L1 Japanese speakers' perceptual ac-
774 curacy influences the tongue movement patterns in producing word-initial
775 liquid-vowel sequences in English. This was investigated by classifying L1
776 Japanese speakers into two groups according to their performance in the per-
777 ceptual identification task of word-initial English liquids. Gaussian Mixture
778 Models suggests that they can be grouped into two groups, which I called
779 'Intermediate' and 'Advanced' learners based on the proportion of correct
780 responses in the perception task.

781 Overall, the current study provides little evidence that L1 Japanese per-
782 ceptual accuracy influences articulation of English liquids, as demonstrated
783 by the substantial degree of uncertainty in predicting the difference in FPC1
784 values between the Intermediate and Advanced groups across all vowel con-
785 texts. This provides little support for a broader claim that L1 Japanese
786 speakers' perceptual accuracy correlates with their production accuracy (Brad-
787 low et al., 1997; Saito and van Poeteren, 2018; Shinohara and Iverson, 2018;

788 Flege, 1995). Previous work predicts correlations between perception and
789 production based on an assumption that perceptual accuracy shapes accurate
790 mental representations for L2 sounds where detailed articulatory commands
791 are specified (Flege et al., 2002). It was also shown that the area of the brain
792 concerning acoustic-articulatory mapping was activated when L1 Japanese
793 speakers improved perceptual identification accuracy of English /l/ and /ɹ/
794 (Callan et al., 2003). Empirical work indeed demonstrates that L1 Japanese
795 speakers’ production of English /l/ and /ɹ/ was evaluated to be more accu-
796 rate after intensive perceptual training (Bradlow et al., 1997; Shinohara and
797 Iverson, 2018), suggesting “a unified, common mental representation that
798 underlies both speech perception and speech production” (Bradlow et al.,
799 1997, p. 2308).

800 A lack of a clear relationship between perception and production in this
801 study could be due to the complex perception-production relationship sug-
802 gested in recent studies (Hattori and Iverson, 2011; Flege et al., 2021). Al-
803 though accurate perception is understood as a necessary condition for L2
804 speech production development, it is not a sufficient factor that enables L2
805 learners to improve their production (Flege and Bohn, 2021; Sheldon and
806 Strange, 1982). In order to account for the complex perception-production re-
807 lationships, the Speech Learning Model (SLM) relaxes its strong perception-
808 precedence assumption, positing that perception and production coevolve
809 (Flege et al., 2021). Empirically, previous research also shows inconsistency
810 between perception and production, demonstrating that L1 Japanese speak-
811 ers’ perceptual accuracy did not show statistically significant correlations
812 with the use of F2 and F3 in production (Hattori and Iverson, 2011). Over-
813 all, the current findings add evidence to the complexity between perception
814 and production, demonstrating that L1 Japanese speakers’ perception skills
815 may not always be manifested as L1-like articulation of English /l/ and /ɹ/.

816 A possible factor that complicates the link between perception and pro-
817 duction could be the fact that English liquids are inherently dynamic, com-
818 plex segments (Sproat and Fujimura, 1993; Campbell et al., 2010). This
819 suggests that L2 learners may need not only to establish the perception-
820 production link but also to acquire appropriate temporal coordination pat-
821 terns both in acoustics and articulation. Previous research argues that L1
822 Japanese speakers can learn to produce English /l/ and /ɹ/ in a similar
823 manner as L1 English speakers, especially along the F2 dimension (Saito
824 and Munro, 2014; Flege et al., 1995). L1 Japanese speakers, however, realise
825 both F2 and F3 differently from L1 English speakers when time-varying for-

826 mant trajectories are taken into account for word-initial English /l/ and /ɭ/
827 (Nagamine, 2024). Crucially, L1 Japanese speakers exhibit a greater vari-
828 ability in the shape and height of the F2–F1 and F3 trajectories for word-
829 initial English /l/ and /ɭ/ across vowel contexts than L1 English speakers
830 (Nagamine, 2024). A lack of clear perceptual accuracy effect in this study
831 corroborates the argument that perceptual accuracy alone may not explain
832 the production accuracy in L2 speech production and that coarticulation is
833 one of the important aspects for L1 Japanese speakers to acquire when pro-
834 ducing English /l/ and /ɭ/ (Flege and Bohn, 2021; Sheldon and Strange,
835 1982).

836 5.2. *Liquid-vowel coarticulation between L1 English and L1 Japanese speak-* 837 *ers*

838 Research Question 2 seeks differences between L1 English and L1 Japanese
839 speakers in the coarticulatory patterns in the word-initial liquid-vowel se-
840 quence in English. The current study provides compelling evidence that L1
841 Japanese speakers differ from L1 English speakers in the liquid-vowel coar-
842 ticulation given a greater variability in FPC1 scores in their production. In
843 addition, such coarticulatory difference between L1 English and L1 Japanese
844 speakers is manifested differently for English /l/ and /ɭ/, which could be due
845 to the differences in the degree of inherent coarticulatory resistance between
846 these liquid segments.

847 Given the articulatory dimension identified from PCA, the results here
848 suggest that L1 Japanese speakers show different coarticulatory patterns from
849 L1 English speakers in the word-initial liquid-vowel sequence along the tongue
850 dorsum dimension. It is commonly understood that L2 learners acquire L2
851 segments in relation to the closest L1 counterpart (Flege, 1995; Best and
852 Tyler, 2007). L1 Japanese speakers acquire English /l/ and /ɭ/ as poor
853 instances of Japanese liquid category /r/, which is canonically realised as an
854 alveolar tap or flap [ɾ]. Given these, it could be argued that the L1 Japanese
855 speakers’ dorsal liquid-vowel coarticulation in their L2 English /l/ and /ɭ/
856 may be influenced by (co)articulation of Japanese /r/.

857 The current results demonstrate that L1 Japanese speakers exhibit a
858 greater variability in tongue dorsum movement across vowel contexts than
859 L1 English speakers, especially for English /ɭ/. This is shown clearly in Fig-
860 ure 6 and Table 7. For English /ɭ/, the 95% HDIs for L1 English speakers
861 encompass zero, whereas the 95% HDIs for L1 Japanese speakers are distinc-
862 tively away from zero with the probability of direction values close to 1.00 in

863 all possible combinations of vowel contexts. This overall suggests a credible
864 evidence for the difference in FPC1 values between vowel contexts for L1
865 Japanese speakers, whereas it is generally unclear whether such difference
866 exists for L1 English speakers.

867 In contrast to English /ɹ/, both L1 English and L1 Japanese speakers ex-
868 hibit similar patterns of FPC1 values for English /l/ across vowel contexts.
869 This could be explained in terms of coarticulatory properties of English /l/
870 and Japanese /r/. Previous research shows that laterals exhibit a lesser de-
871 gree of coarticulatory resistance than English /ɹ/, suggesting that production
872 of English /l/ is more influenced by the vowel contexts than that of English
873 /ɹ/ (Proctor et al., 2019). Coarticulatory resistance can be predicted by the
874 degree of involvement of the dorsal gesture in production, in which clearer
875 laterals usually exhibit a lesser degree of coarticulatory resistance than darker
876 laterals (Recasens, 2012). These previous descriptions overall agree with the
877 findings in the current study, as it includes clearer, syllable-initial /l/s and
878 demonstrates a greater variability can be predicted across vowel contexts in
879 L1 English speakers' production.

880 Without a direct comparison between English /l/ and Japanese /r/, how-
881 ever, it remains unclear how L1 Japanese speakers' production of English /l/
882 could be fully explained. On the one hand, it is possible that L1 Japanese
883 speakers acquire L1-like coarticulatory properties for English /l/, resulting
884 in a similar tendency as that of L1 English speakers. On the other hand,
885 L1 Japanese speakers' production of English /l/ may be influenced by the
886 articulatory strategies for Japanese /r/. Tentatively, however, I argue that
887 the latter account is more plausible than the former in the light of the current
888 findings on the articulation for English /ɹ/ and the previous findings related
889 to the relative difficulty of acquisition between English /l/ and /ɹ/.

890 The current study demonstrates clear difference in liquid-vowel coarticu-
891 lation for English /ɹ/. L1 English speakers, on the one hand, show relatively
892 stable FPC1 values across vowel contexts, agreeing with the previous find-
893 ings that English /ɹ/ exhibit a greater degree of coarticulatory resistance
894 resulting in stable tongue shape across vowel contexts (Proctor et al., 2019).
895 L1 Japanese speakers, on the other hand, showed a greater magnitude of
896 variability in FPC1 scores across vowel contexts, indicated in Figure 4 and
897 shown more clearly in terms of the posterior distributions and the probabilit-
898 y of direction in Figure 6 and Table 7. This could result from differences
899 in the degree of tongue dorsum activity between English /ɹ/ and Japanese
900 /r/; for the latter, previous research reports a smaller degree of dorsal ac-

901 tivity and therefore a greater susceptibility to the vowel contexts (Recasens
902 and Rodríguez, 2016; Maekawa, 2023; Yamane et al., 2015). Given this, L1
903 Japanese speakers’ greater variability in FPC1 values could be accounted
904 for such that they transfer L1 articulatory strategy to produce English /ɹ/,
905 which is manifested as stronger dorsal vocalic coarticulation than L1 English
906 speakers.

907 The results for English /ɹ/ provide further accounts for the English /l/
908 results. Previous research argues that L1 Japanese speakers learn to produce
909 English /ɹ/ more easily than English /l/ (Aoyama et al., 2004, 2019). They
910 perceive English /ɹ/ as more dissimilar to Japanese /r/ than English /l/
911 is, which facilitates a formation of new phonetic category for English /ɹ/
912 (Aoyama et al., 2004; Flege et al., 2021). Given these, it can be expected
913 that L1 Japanese speakers could learn the coarticulatory patterns for English
914 /ɹ/ before English /l/ if coarticulatory properties can be acquired in a similar
915 manner as articulation of individual segments (Beristain, 2022). The current
916 findings, however, do not provide much evidence that L1 Japanese speakers
917 acquire a similar coarticulatory pattern for English /ɹ/ as that of L1 English
918 speakers.

919 Furthermore, an ultrasound study reports a similar degree of coartic-
920 ulatory resistance between clear [l]s and alveolar taps [ɾ] in terms of the
921 tongue dorsum activity involved in the consonantal articulation (Recasens
922 and Rodríguez, 2016). Provided that L1 Japanese speakers find English /l/
923 more similar to Japanese /r/ than English /ɹ/ is, it is likely that they rede-
924 ploy the existing articulatory strategy for English /l/ than for English /ɹ/
925 (Saito and van Poeteren, 2018). These considerations overall suggest that
926 it is unlikely that L1 Japanese speakers in the current study produce En-
927 glish /l/ in a similar manner as L1 English speakers; instead, it points to a
928 possibility that L1 Japanese speakers produce English liquids under the ar-
929 ticulatory influence of the corresponding L1 category of Japanese /r/, which
930 is manifested more for English /ɹ/ than for English /l/ due to differences in
931 tongue dorsum movement and thus in coarticulatory resistance.

932 Note, however, that the above account is based on the assumption that
933 L1 Japanese speakers’ perceptual accuracy does influence their articulation
934 of English liquids, which seem to contradict to the findings discussed earlier.
935 A lack of perceptual accuracy effects in the current study could, however, be
936 due to the research design, in which the L1 Japanese-speaking participants
937 were recruited from a relatively homogeneous pool of English learners in
938 Japan. It is generally difficult to make theoretical predictions regarding the

939 success of L2 speech learning in the English as a foreign language (EFL)
940 context, as in the current study, compared to the English as a second language
941 (ESL) context found in the majority of previous research where a long-term
942 residence in an English-speaking environment is assumed (Tyler, 2019; Saito
943 and Munro, 2014; Flege, 1995; Aoyama et al., 2004). It is possible that
944 the participants in the current study can be considered under a common
945 label of ‘inexperienced Japanese speakers of English’ in the previous research
946 given their limited experience in overseas study experience (e.g., Flege et al.,
947 1995, 1996). The lack of perceptual accuracy effect in this study, therefore,
948 corroborates the view that L1 Japanese speakers in the current study are
949 under the influence of Japanese /r/ when producing English /l/ and /ɹ/.

950 Finally, the current study provides supporting evidence to the Bilingual
951 Coarticulatory Model (BCM; Beristain, 2022) in highlighting the role of coar-
952 ticulatory properties in L2 speech production. The difference between En-
953 glish /l/ and /ɹ/ also supports the model’s postulate that coarticulatory prop-
954 erties can be acquired in a similar manner as segmental properties (Beristain,
955 2022). On the other hand, the current study does not offer supporting evi-
956 dence to the BCM’s argument that proficient L2 learners tend to exhibit
957 more “native-like” coarticulatory patterns than those who are less proficient
958 (Beristain, 2022). This is likely to have resulted from differences in research
959 design in that the current study does not address participants’ linguistic pro-
960 ficiency directly. Nevertheless, it could be argued that perceptual accuracy
961 has a greater theoretical relevance to the other models of L2 speech learning,
962 and future research should seek a link between participants’ perceptual accu-
963 racy and linguistic proficiency, especially in the EFL context in which it is still
964 challenging to generalise the findings regarding these variables (Tyler, 2019).
965 Overall, the current study supports the previous findings that L2 learners use
966 L1 articulatory strategies to produce ‘similar’ sounds in L2 (Oakley, 2021)
967 and demonstrates further that such articulatory transfer could also be ob-
968 served in coarticulation, which could be a possible articulatory challenge for
969 L1 Japanese speakers in producing English liquids.

970 6. Conclusion

971 This study compared liquid-vowel coarticulation in English produced by
972 L1 English and two groups of L1 Japanese speakers, classified based on their
973 perceptual identification accuracy of English liquids. The study examined
974 dynamic changes in midsagittal tongue shapes recorded using ultrasound

975 and then analysed using the Principal Component Analysis and the Func-
976 tional Principal Component Analysis. Statistical analysis showed clear dif-
977 ferences in the liquid-vowel coarticulatory patterns between L1 English and
978 L1 Japanese speakers in terms of tongue dorsum movement. L1 Japanese
979 speakers' articulation of /ɹ/ show a greater degree of variability across vowel
980 contexts than that of L1 English speakers, which could result from the L1
981 transfer of articulatory mechanism from Japanese /r/. This suggests that
982 coarticulatory resistance is an important aspect in for L1 Japanese speakers'
983 production of English liquids.

984 The current study offers implications for future research mainly in the
985 following three aspects. First, L1 Japanese speakers' production of Japanese
986 /r/ could be directly compared to their production of L2 English liquids
987 to empirically confirm the L1 articulatory transfer. This would require a
988 separate statistical analysis design focussing solely on L1 Japanese speakers'
989 data given the difference in the number of variable levels (i.e., whereas both
990 L1 English and L1 Japanese speakers produced English /l/ and /ɹ/, only
991 L1 Japanese speakers produced Japanese /r/). Second, it will be interesting
992 to investigate the gestural coordination patterns focussing on the timing
993 and magnitude of coronal and dorsal gestures (Sproat and Fujimura, 1993;
994 Campbell et al., 2010; Proctor et al., 2019). Comparing relative gestural
995 timing between tongue tip and tongue dorsum between syllable-initial and
996 syllable-final English liquids will further highlight cross-linguistic influence
997 in gestural affinity in L1 and L2 speech production. Finally, articulation of
998 English /l/ and /ɹ/ needs to be investigated in a three-dimensional space. It
999 is possible, for example, that L1 Japanese speakers exhibit different behaviour
1000 of tongue lateral edges, a major articulatory property characterising English
1001 /l/ of which speakers may have an active control (Ying et al., 2021). While
1002 the current findings suggest similar tongue dorsum behaviour for English
1003 /l/ between L1 English and L1 Japanese speakers, the articulatory data on
1004 the coronal plane could help disentangle the acquisition of coarticulatory
1005 resistance from the L1 articulatory transfer for English /l/.

1006 Nevertheless, these limitations clearly lay out new avenues for the ar-
1007 ticulatory approach to L2 speech production research that may be made
1008 possible by the quantitative analysis as shown in the current findings. The
1009 study provides a clear evidence of L1 articulatory routine, attributed here to
1010 the liquid-vowel coarticulation patterns, while highlighting complexity in the
1011 perception-production link, articulatory variability and the L2 proficiency
1012 measurement in L2 speech learning research. Overcoming these challenges,

1013 the articulatory approach to the L2 speech production research would ulti-
1014 mately be able to offer accounts as to why foreign accents are the persistent
1015 nature of adult L2 learners' production of L2 speech.

1016 7. Acknowledgements

1017 This article is part of the author's PhD research, supervised by Professor
1018 Claire Nance and Dr Sam Kirkham at Lancaster University, UK. An earlier
1019 version of the study was presented at the Colloquium of the British Associ-
1020 ation of Academic Phoneticians (BAAP) 2024 on the 25th of March, 2024,
1021 in Cardiff, UK. I thank the participants for their time and willing participa-
1022 tion in the experiments. I thank Professor Noriko Nakanishi, Professor Yuri
1023 Nishio and Dr Bronwen Evans, for facilitating the data collection. Thank
1024 you also to Dr Alan Wrench for technical support regarding the ultrasound
1025 data analysis.

1026 This work is financially supported by the Graduate Scholarship for Degree-
1027 Seeking Students by the Japan Student Services Organization (JASSO) [ID:
1028 20SD10500601] and the 2022 Research Grant by the Murata Science Foun-
1029 dation [grant number: M22助入027] awarded to the author. Codes and
1030 data supporting the findings in this article are publicly available at <https://osf.io/h3zfq/>. The author has no conflicts to declare. This research is
1031 approved by ethics committees at Lancaster University, Kobe Gakuin Uni-
1032 versity, and Meijo University. Informed consent was obtained from all par-
1033 ticipants.
1034

1035 References

- 1036 Alwan, A., Narayanan, S., Haker, K., 1997. Toward articulatory-acoustic
1037 models for liquid approximants based on MRI and EPG data. Part II. The
1038 rhotics. *The Journal of the Acoustical Society of America* 101, 1078–1089.
1039 doi:10.1121/1.417972.
- 1040 Anwyl-Irvine, A.L., Massoné, J., Flitton, A., Kirkham, N.Z., Evershed, J.K.,
1041 2018. Gorilla Experiment Builder.
- 1042 Aoyama, K., Flege, J.E., Akahane-Yamada, R., Yamada, T., 2019. An acous-
1043 tic analysis of American English liquids by adults and children: Native En-
1044 glish speakers and native Japanese speakers of English. *The Journal of the*
1045 *Acoustical Society of America* 146, 2671–2681. doi:10.1121/1.5130574.

- 1046 Aoyama, K., Flege, J.E., Guion, S.G., Akahane-Yamada, R., Yamada, T.,
1047 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of
1048 Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32, 233–250.
1049 doi:10.1016/S0095-4470(03)00036-6.
- 1050 Aoyama, K., Hong, L., Flege, J.E., Akahane-Yamada, R., Yamada, T.,
1051 2023. Relationships Between Acoustic Characteristics and Intelligibil-
1052 ity Scores: A Reanalysis of Japanese Speakers’ Productions of American
1053 English Liquids. *Language and Speech* , 002383092211409doi:10.1177/
1054 00238309221140910.
- 1055 Archibald, J., 2021. Ease and Difficulty in L2 Phonology: A Mini-Review.
1056 *Frontiers in Communication* 6.
- 1057 Articulate Instruments, 2022. Articulate Assistant Advanced version 220.
1058 Articulate Instruments.
- 1059 Asano, Y., Gubian, M., 2018. “Excuse meeee!!”: (Mis)coordination of lexical
1060 and paralinguistic prosody in L2 hyperarticulation. *Speech Communica-*
1061 *tion* 99, 183–200. doi:10.1016/j.specom.2017.12.011.
- 1062 Baayen, R.H., 2008. *Analyzing Linguistic Data: A Practical Introduc-*
1063 *tion to Statistics Using R*. Cambridge University Press. doi:10.1017/
1064 CB09780511801686.
- 1065 Beristain, A.M., 2022. *The Acquisition of Acoustic and Aerodynamic Pat-*
1066 *terns of Coarticulation in Second and Heritage Languages*. Ph.D. thesis.
1067 University of Illinois Urbana-Champaign.
- 1068 Best, C.T., 1995. A Direct Realist View of Cross-Lanugage Speech Percep-
1069 tion., in: Strange, W. (Ed.), *Speech Perception and Linguistic Experience:*
1070 *Theoretical and Methodological Issues*. York Press, pp. 171–204.
- 1071 Best, C.T., Strange, W., 1992. Effects of phonological and phonetic factors
1072 on cross-language perception of approximants. *Journal of Phonetics* 20,
1073 305–330. doi:10.1016/S0095-4470(19)30637-0.
- 1074 Best, C.T., Tyler, M.D., 2007. Nonnative and second-language speech per-
1075 ception: Commonalities and complementarities, in: Bohn, O.S., Munro,
1076 M.J. (Eds.), *Language Experience in Second Language Speech Learning:*

- 1077 In Honor of James Emil Flege. John Benjamins Publishing Company, Am-
1078 sterdam, pp. 13–34. doi:10.1075/111t.17.07bes.
- 1079 Boersma, P., Weenink, D., 2022. Praat: Doing Phonetics by Computer.
- 1080 Bradlow, A.R., 2008. Training non-native language sound patterns, in: Ed-
1081 wards, J.G.H., Zampini, L, M. (Eds.), Phonology and Second Language
1082 Acquisition. John Benjamins Publishing Company, pp. 287–308.
- 1083 Bradlow, A.R., Akahane-Yamada, R., Pisoni, D.B., Tohkura, Y., 1999. Train-
1084 ing Japanese listeners to identify English /r/ and /l/: Long-term retention
1085 of learning in perception and production. *Perception & Psychophysics* 61,
1086 977–985. doi:10.3758/BF03206911.
- 1087 Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R., Tohkura, Y., 1997. Train-
1088 ing Japanese listeners to identify English /r/ and /l/: IV. Some effects of
1089 perceptual learning on speech production. *Journal of Acoustical Society*
1090 *of America* 101, 2299–2310. doi:10.1121/1.418276.
- 1091 Brekelmans, G., Lavan, N., Saito, H., Clayards, M., Wonnacott, E., 2022.
1092 Does high variability training improve the learning of non-native phoneme
1093 contrasts over low variability training? A replication. *Journal of Memory*
1094 *and Language* 126, 104352. doi:10.1016/j.jml.2022.104352.
- 1095 Bürkner, P.C., 2017. **Brms** : An *R* Package for Bayesian Multilevel Models
1096 Using *Stan*. *Journal of Statistical Software* 80. doi:10.18637/jss.v080.
1097 i01.
- 1098 Callan, D.E., Tajima, K., Callan, A.M., Kubo, R., Masaki, S., Akahane-
1099 Yamada, R., 2003. Learning-induced neural plasticity associated with
1100 improved identification performance after training of a difficult second-
1101 language phonetic contrast. *NeuroImage* 19, 113–124. doi:10.1016/
1102 S1053-8119(03)00020-X.
- 1103 Campbell, F., Gick, B., Wilson, I., Vatikiotis-Bateson, E., 2010. Spatial and
1104 Temporal Properties of Gestures in North American English /r/. *Language*
1105 *and Speech* 53, 49–69. doi:10.1177/0023830909351209.
- 1106 Chang, C.B., 2019. The phonetics of second language learning and bilin-
1107 gualism, in: Katz, W.F., Assmann, P.F. (Eds.), *The Routledge Handbook*

- 1108 of Phonetics. 1 ed.. Routledge, Abingdon, Oxon ; New York, NY : Rout-
 1109 ledge, 2019. | Series: Routledge handbooks in linguistics, pp. 427–447.
 1110 doi:10.4324/9780429056253-16.
- 1111 Chomsky, N., Halle, M., 1968. The Sound Pattern of English. Harper & Row
 1112 Publishers, New York, Evanston, and London.
- 1113 Colantoni, L., Steele, J., 2008. Integrating articulatory constraints into mod-
 1114 els of second language phonological acquisition. *Applied Psycholinguistics*
 1115 29, 489–534. doi:10.1017/S0142716408080223.
- 1116 Colantoni, L., Steele, J., Escudero, P., 2015. *Sec-
 1117 ond Language Speech: Theory and Practice*.
 1118 [https://www.cambridge.org/highereducation/books/second-language-
 1119 speech/CD3AF18D3D80D60C82CD68AEE1F8B87D](https://www.cambridge.org/highereducation/books/second-language-speech/CD3AF18D3D80D60C82CD68AEE1F8B87D). doi:10.1017/
 1120 CB09781139087636.
- 1121 Cronenberg, J., Gubian, M., Harrington, J., Ruch, H., 2020. A dynamic
 1122 model of the change from pre- to post-aspiration in Andalusian Spanish.
 1123 *Journal of Phonetics* 83, 101016. doi:10.1016/j.wocn.2020.101016.
- 1124 Daigaku Eigo Kyoiku Gakkai Kihongo Kaitei Tokubetsu Iinkai, 2016.
 1125 Daigaku Eigo Kyouiku Gakkai Kihongo Risuto: Shin JACET 8000 [The
 1126 Japan Association for College English Teachers Basic Word List: New
 1127 JACET 8000]. Kirihara Shoten.
- 1128 Davidson, L., 2011. Phonetic and Phonological Factors in the Second Lan-
 1129 guage Production of Phonemes and Phonotactics. *Language and Linguis-
 1130 tics Compass* 5, 126–139. doi:10.1111/j.1749-818X.2010.00266.x.
- 1131 Delattre, P., Freeman, D.C., 1968. A Dialect Study of American R's by X-ray
 1132 motion picture. *Linguistics* 6, 29–68. doi:10.1515/ling.1968.6.44.29.
- 1133 Derrick, D., Gick, B., 2011. Individual variation in English flaps and taps:
 1134 A case of categorical phonetics. *The Canadian Journal of Linguistics /
 1135 La revue canadienne de linguistique* 56, 307–319. doi:10.1353/cjl.2011.
 1136 0024.
- 1137 Espinal, A., Thompson, A., Kim, Y., 2020. Acoustic characteristics of
 1138 American English liquids /ɹ/, /l/, /ɹl/ produced by Korean L2 adults.

- 1139 The Journal of the Acoustical Society of America 148, EL179–EL184.
1140 doi:10.1121/10.0001758.
- 1141 Flege, J.E., 1987. The production of “new” and “similar” phones in a foreign
1142 language: Evidence for the effect of equivalence classification. *Journal of*
1143 *Phonetics* 15, 47–65. doi:10.1016/S0095-4470(19)30537-6.
- 1144 Flege, J.E., 1992. The intelligibility of English vowels spoken by British
1145 and Dutch talkers, in: Kent, R.D. (Ed.), *Studies in Speech Pathology and*
1146 *Clinical Linguistics*. John Benjamins Publishing Company, Amsterdam.
1147 volume 1, pp. 157–232. doi:10.1075/sspc1.1.06fle.
- 1148 Flege, J.E., 1995. Second Language Speech Learning Theory, Findings and
1149 Problems, in: Strange, W. (Ed.), *Speech Perception and Linguistic Expe-*
1150 *rience: Issues in Cross-Language Research*. York Press, pp. 233–277.
- 1151 Flege, J.E., Aoyama, K., Bohn, O.S., 2021. The Revised Speech Learning
1152 Model (SLM-r) Applied, in: Wayland, R. (Ed.), *Second Language Speech*
1153 *Learning*. 1 ed.. Cambridge University Press, pp. 84–118. doi:10.1017/
1154 9781108886901.003.
- 1155 Flege, J.E., Bohn, O.S., 2021. The Revised Speech Learning Model (SLM-r),
1156 in: Wayland, R. (Ed.), *Second Language Speech Learning: Theoretical and*
1157 *Empirical Progress*. 1 ed.. Cambridge University Press, pp. 3–83. doi:10.
1158 1017/9781108886901.002.
- 1159 Flege, J.E., Mackay, I.R.A., Piske, T., 2002. Assessing bilingual dominance.
1160 *Applied Psycholinguistics* 23, 567–598. doi:10.1017/S0142716402004046.
- 1161 Flege, J.E., Takagi, N., Mann, V., 1995. Japanese Adults can Learn to
1162 Produce English /I/ and /l/ Accurately. *Language and Speech* 38, 25–55.
1163 doi:10.1177/002383099503800102.
- 1164 Flege, J.E., Takagi, N., Mann, V., 1996. Lexical familiarity and English-
1165 language experience affect Japanese adults’ perception of /ɪ/ and /l/. *The*
1166 *Journal of the Acoustical Society of America* 99, 1161–1173. doi:10.1121/
1167 1.414884.
- 1168 Fraley, C., Raftery, A.E., Scrucca, L., Murphy, T.B., Fop, M., 2023. *Mclust:*
1169 *Gaussian Mixture Modelling for Model-Based Clustering, Classification,*
1170 *and Density Estimation*.

- 1171 Franke, M., Roettger, T.B., 2019. Bayesian regression modeling (for factorial
1172 designs): A tutorial. doi:10.31234/osf.io/cdxv3.
- 1173 Grosjean, F., 2008. The bilingual's language mode, in: *Studying Bilinguals*.
1174 Oxford University Press, Oxford ; New York. Oxford Linguistics, pp. 36–
1175 66.
- 1176 Guion, S.G., Flege, J.E., Liu, S.H., Yeni-Komshian, G.H., 2000. Age of
1177 learning effects on the duration of sentences produced in a second language.
1178 *Applied Psycholinguistics* 21, 205–228. doi:10.1017/S0142716400002034.
- 1179 Harper, S., Goldstein, L., Narayanan, S., 2020. Variability in individual
1180 constriction contributions to third formant values in American English
1181 /ɹ/. *The Journal of the Acoustical Society of America* 147, 3905–3916.
1182 doi:10.1121/10.0001413.
- 1183 Hashi, M., Honda, K., Westbury, J.R., 2003. Time-varying acoustic and
1184 articulatory characteristics of American English [ɹ]: A cross-speaker study.
1185 *Journal of Phonetics* 31, 3–22. doi:10.1016/S0095-4470(02)00062-1.
- 1186 Hattori, K., Iverson, P., 2009. English /r-/l/ category assimilation by
1187 Japanese adults: Individual differences and the link to identification ac-
1188 curacy. *The Journal of the Acoustical Society of America* 125, 469–479.
1189 doi:10.1121/1.3021295.
- 1190 Hattori, K., Iverson, P., 2011. Examination of the Relationship between L2
1191 Perception and Production: An Investigation of English /r-/l/ Perception
1192 and Production by Adult Japanese Speakers, in: Nakano, M. (Ed.),
1193 *Interspeech Workshop on Second Language Studies: Acquisition, Learn-
1194 ing, Education and Technology*, pp. 2–4.
- 1195 Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Ket-
1196 termann, A., Siebert, C., 2003. A perceptual interference account of ac-
1197 quisition difficulties for non-native phonemes. *Cognition* 87, B47–B57.
1198 doi:10.1016/S0010-0277(02)00198-1.
- 1199 Kallioinen, N., Paananen, T., Bürkner, P.C., Vehtari, A., 2023. Detecting
1200 and diagnosing prior and likelihood sensitivity with power-scaling. doi:10.
1201 48550/arXiv.2107.14054, arXiv:2107.14054.

- 1202 Keating, P.A., 1985. Universal Phonetics and the Organization of Grammars,
1203 in: *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Academic
1204 Press, pp. 115–132.
- 1205 King, H., Ferragne, E., 2020. Loose lips and tongue tips: The central role of
1206 the /r/-typical labial gesture in Anglo-English. *Journal of Phonetics* 80,
1207 100978. doi:10.1016/j.wocn.2020.100978.
- 1208 Kirkham, S., Nance, C., 2022. Diachronic phonological asymmetries and the
1209 variable stability of synchronic contrast. *Journal of Phonetics* 94, 101176.
1210 doi:10.1016/j.wocn.2022.101176.
- 1211 Kirkham, S., Nance, C., Littlewood, B., Lightfoot, K., Groarke, E., 2019.
1212 Dialect variation in formant dynamics: The acoustics of lateral and vowel
1213 sequences in Manchester and Liverpool English. *The Journal of the Acous-
1214 tical Society of America* 145, 784–794. doi:10.1121/1.5089886.
- 1215 Kruschke, J.K., 2015. *Doing Bayesian Data Analysis: A Tutorial with R,
1216 JAGS, and Stan*. 2 ed., Elsevier. doi:10.1016/B978-0-12-405888-0.
1217 09999-2.
- 1218 Lenth, R.V., Buerkner, P., Herve, M., Love, J., Miguez, F., Riebl, H.,
1219 Singmann, H., 2022. *Emmeans: Estimated Marginal Means, aka Least-
1220 Squares Means*.
- 1221 Maekawa, K., 2023. Articulatory characteristics of the Japanese /r/: A real-
1222 time MRI study., in: Radek Skarnitzl, Jan Volín (Eds.), *Proceedings of the
1223 20th International Congress of Phonetic Sciences*, Guarant International,
1224 Prague. pp. 992–996.
- 1225 Makino, T., 2009. Vowel substitution patterns in Japanese speakers' English,
1226 in: Čubrović, B., Paunovic, T. (Eds.), *Ta(l)King English Phonetics across
1227 Frontiers*. Cambridge Scholars, Newcastle, pp. 19–32.
- 1228 Makowski, D., Ben-Shachar, M.S., Chen, S.H.A., Lüdecke, D., 2019. Indices
1229 of Effect Existence and Significance in the Bayesian Framework. *Frontiers
1230 in Psychology* 10.
- 1231 Manuel, S., 1999. Cross-language studies: Relating language-particular coar-
1232 ticulation patterns to other language-particular facts, in: Hardcastle, W.J.,

- 1233 Hewlett, N. (Eds.), *Coarticulation*. 1 ed.. Cambridge University Press, pp.
1234 179–198. doi:10.1017/CB09780511486395.009.
- 1235 Masaki, S., Akahane-Yamada, R., Tiede, M.K., Shimada, Y., Fujimoto, I.,
1236 1996. An MRI-based analysis of the English /r/ and /l/ articulations,
1237 in: *Proceeding of Fourth International Conference on Spoken Language*
1238 *Processing.*, pp. 1581–1584. doi:10.1109/ICSLP.1996.607922.
- 1239 McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., Sonderegger, M., 2017.
1240 *Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi*,
1241 in: *Interspeech 2017, ISCA*. pp. 498–502. doi:10.21437/Interspeech.
1242 2017-1386.
- 1243 McElreath, R., 2016. *Statistical Rethinking: A Bayesian Course with Ex-*
1244 *amples in R and STAN*. 1 ed., Chapman and Hall/CRC, New York.
1245 doi:10.1201/9780429029608.
- 1246 Moore, J., Shaw, J., Kawahara, S., Arai, T., 2018. Articulation strategies
1247 for English liquids used by Japanese speakers. *Acoustical Science and*
1248 *Technology* 39, 75–83. doi:10.1250/ast.39.75.
- 1249 Nagamine, T., 2024. Formant dynamics in second language speech: Japanese
1250 speakers’ production of English liquids. *The Journal of the Acoustical*
1251 *Society of America* 155, 479–495. doi:10.1121/10.0024351.
- 1252 Nagle, C.L., Baese-Berk, M.M., 2022. Advancing the state of the art in L2
1253 speech perception-production research: Revisiting theoretical assumptions
1254 and methodological practices. *Studies in Second Language Acquisition* 44,
1255 1–26. doi:10.1017/S0272263121000371.
- 1256 Nance, C., Kirkham, S., 2022. Phonetic typology and articulatory con-
1257 straints: The realisation of secondary articulations in Scottish Gaelic
1258 rhotics. *Language* , 419–460.
- 1259 Nogita, A., 2016. *L2 Letter-Sound Correspondence: Mapping between En-*
1260 *glish Vowel Graphemes and Phonemes by Japanese EAL Learners*. Ph.D.
1261 thesis. University of Victoria. Victoria, British Columbia.
- 1262 Oakley, M., 2021. *Articulating Non-native Vowel Contrasts*. Ph.D. thesis.
1263 Georgetown University. District of Columbia, United States.

- 1264 Öhman, S.E.G., 1966. Coarticulation in VCV Utterances: Spectrographic
1265 Measurements. *The Journal of the Acoustical Society of America* 39, 151–
1266 168. doi:10.1121/1.1909864.
- 1267 Olsen, M.K., 2012. The L2 Acquisition of Spanish Rhotics by L1 English
1268 Speakers: The Effect of L1 Articulatory Routines and Phonetic Context
1269 for Allophonic Variation. *Hispania* 95, 65–82. arXiv:41440363.
- 1270 Palo, P., Schaeffler, S., Scobbie, J.M., 2014. Pre-speech tongue movements
1271 recorded with ultrasound, in: Fuchs, S., Grice, M., Hermes, A., Lancia,
1272 L., Mücke, D. (Eds.), *Proceedings of the 10th International Seminar on*
1273 *Speech Production (ISSP)*, Cologne, Germany, pp. 300-303.. pp. 300–303.
- 1274 Potts, A., Baker, P., 2012. Does semantic tagging identify cultural change in
1275 British and American English? *International Journal of Corpus Linguistics*
1276 17, 295–324. doi:10.1075/ijcl.17.3.01pot.
- 1277 Proctor, M., Walker, R., Smith, C., Szalay, T., Goldstein, L., Narayanan, S.,
1278 2019. Articulatory characterization of English liquid-final rimes. *Journal*
1279 *of Phonetics* 77, 100921. doi:10.1016/j.wocn.2019.100921.
- 1280 R Core Team, 2023. R: A Language and Environment for Statistical Com-
1281 puting. R Foundation for Statistical Computing.
- 1282 Recasens, D., 1991. On the production characteristics of apicoalveolar taps
1283 and trills. *Journal of Phonetics* 19, 267–280. doi:10.1016/S0095-4470(19)
1284 30344-4.
- 1285 Recasens, D., 2012. A cross-language acoustic study of initial and final al-
1286 lophones of /l/. *Speech Communication* 54, 368–383. doi:10.1016/j.
1287 *specom*.2011.10.001.
- 1288 Recasens, D., Rodríguez, C., 2016. A study on coarticulatory resistance and
1289 aggressiveness for front lingual consonants and vowels using ultrasound.
1290 *Journal of Phonetics* 59, 58–75. doi:10.1016/j.wocn.2016.09.002.
- 1291 Recasens, D., Rodríguez, C., 2017. Lingual Articulation and Coarticulation
1292 for Catalan Consonants and Vowels: An Ultrasound Study. *Phonetica* 74,
1293 125–156. doi:10.1159/000452475.

- 1294 Saito, K., Munro, M.J., 2014. The Early Phase of /ɹ/ Production Development in Adult Japanese Learners of English. *Language and Speech* 57, 1295 451–469. doi:10.1177/0023830913513206.
- 1297 Saito, K., van Poeteren, K., 2018. The perception–production link revisited: 1298 The case of Japanese learners’ English /ɹ/ performance. *International* 1299 *Journal of Applied Linguistics* 28, 3–17. doi:10.1111/ijal.12175.
- 1300 Schwartz, G., Kaźmierski, K., 2020. Vowel dynamics in the acquisition of L2 1301 English – an acoustic study of L1 Polish learners. *Language Acquisition* 1302 27, 227–254. doi:10.1080/10489223.2019.1707204.
- 1303 Scobbie, J., Lawson, E., Cowen, S., Cleland, J., Wrench, A., 2011. A common 1304 co-ordinate system for mid-sagittal articulatory measurement. *QMU CASL* 1305 *Working Papers* 20, 1–4.
- 1306 Setter, J., Jenkins, J., 2005. Pronunciation. *Language Teaching* 38, 1–17. 1307 doi:10.1017/S026144480500251X.
- 1308 Sheldon, A., Strange, W., 1982. The acquisition of /r/ and /l/ by 1309 Japanese learners of English: Evidence that speech production can pre- 1310 ceede speech perception. *Applied Psycholinguistics* 3, 243–261. doi:10. 1311 1017/S0142716400001417.
- 1312 Shinohara, Y., Iverson, P., 2018. High variability identification and discrimi- 1313 nation training for Japanese speakers learning English /r-/l/. *Journal of* 1314 *Phonetics* 66, 242–251. doi:10.1016/j.wocn.2017.11.002.
- 1315 Slud, E., Stone, M., Smith, P.J., Goldstein Jr., M., 2002. Principal Com- 1316 ponents Representation of the Two-Dimensional Coronal Tongue Surface. 1317 *Phonetica* 59, 108–133. doi:10.1159/000066066.
- 1318 Spreafico, L., Pucher, M., Matosova, A., 2018. UltraFit: A Speaker-friendly 1319 Headset for Ultrasound Recordings in Speech Science, in: *Interspeech 2018,* 1320 *ISCA*. pp. 1517–1520. doi:10.21437/Interspeech.2018-995.
- 1321 Sproat, R., Fujimura, O., 1993. Allophonic variation in English /l/ and its 1322 implications for phonetic implementation. *Journal of Phonetics* 21, 291– 1323 311. doi:10.1016/S0095-4470(19)31340-3.
- 1324 Stevens, K.N., 2000. *Acoustic Phonetics*. The MIT Press.

- 1325 Sudo, M.M., Kiritani, S., Yoshioka, H., 1982. An electro-palatographic study
1326 of Japanese intervocalic /r/. *Annual Bulletin of Research Institute of*
1327 *Logopedics and Phoniatics (RILP)* 16, 21–25.
- 1328 Turton, D., 2017. Categorical or gradient? An ultrasound investigation
1329 of /l/-darkening and vocalization in varieties of English. *Laboratory*
1330 *Phonology: Journal of the Association for Laboratory Phonology* 8, 1–
1331 13. doi:10.5334/labphon.35.
- 1332 Tyler, M.D., 2019. PAM-L2 and Phonological Category Acquisition in the
1333 Foreign Language Classroom, in: Nyvad, A.M., Hejná, M., Højen, A.,
1334 Jespersen, A.B., Sørensen, M.H. (Eds.), *A Sound Approach to Language*
1335 *Matters: In Honor of Ocke-Schwen Bohn*. Aarhus University, Denmark,
1336 pp. 607–630.
- 1337 Vance, T.J., 2008. *The Sounds of Japanese*. Cambridge University Press.
- 1338 Vasishth, S., Nicenboim, B., Beckman, M.E., Li, F., Kong, E.J., 2018.
1339 Bayesian data analysis in the phonetic sciences: A tutorial introduction.
1340 *Journal of Phonetics* 71, 147–161. doi:10.1016/j.wocn.2018.07.008.
- 1341 Wang, Y., Bundgaard-Nielsen, R.L., Baker, B.J., Maxwell, O., 2023. Diffi-
1342 culties in decoupling articulatory gestures in L2 phonemic sequences: The
1343 case of Mandarin listeners’ perceptual deletion of English post-vocalic lat-
1344 laterals. *Phonetica* 80, 79–115. doi:10.1515/phon-2022-0027.
- 1345 Watson, C.I., Harrington, J., 1999. Acoustic evidence for dynamic formant
1346 trajectories in Australian English vowels. *The Journal of the Acoustical*
1347 *Society of America* 106, 458–468. doi:10.1121/1.427069.
- 1348 Wells, J.C., 2008. *Longman Pronunciation Dictionary*. 3 ed., Pearson Edu-
1349 cation Ltd.
- 1350 Wilson, I., Kanada, S., 2014. Pre-speech Postures of Second-Language versus
1351 First-Language Speakers. *Journal of the Phonetic Society of Japan* 18,
1352 106–109. doi:10.24467/onseikenkyu.18.2_106.
- 1353 Winter, B., 2019. Correlations and clusters, in: *Sensory Linguistics*. John
1354 Benjamins Publishing Company, pp. 163–174.

- 1355 Wrench, A., Balch-Tomes, J., 2022. Beyond the Edge: Markerless Pose
1356 Estimation of Speech Articulators from Ultrasound and Camera Images
1357 Using DeepLabCut. *Sensors* 22, 1133. doi:10.3390/s22031133.
- 1358 Yamane, N., Howson, P., Po-Chun (Grace), W., 2015. An ultrasound ex-
1359 amination of taps in Japanese, in: *The Scottish Consortium for ICPHS*
1360 *2015 (Ed.)*, Proceedings of the 18th International Congress of Phonetic
1361 Sciences, The International Phonetic Association, Glasgow, UK. pp. 1–5.
- 1362 Ying, J., Shaw, J.A., Carignan, C., Proctor, M., Derrick, D., Best, C.T.,
1363 2021. Evidence for active control of tongue lateralization in Australian
1364 English /l/. *Journal of Phonetics* 86, 101039. doi:10.1016/j.wocn.2021.
1365 101039.
- 1366 Ying, J., Shaw, J.A., Kroos, C., Best, C.T., 2012. Relations Between Acous-
1367 tic and Articulatory Measurements of /l/, in: *Proceedings of the 14th*
1368 *Australasian International Conference on Speech Science and Technology*,
1369 Sydney. pp. 109–112.
- 1370 Zhou, X., Espy-Wilson, C.Y., Boyce, S., Tiede, M., Holland, C., Choe, A.,
1371 2008. A magnetic resonance imaging-based articulatory and acoustic study
1372 of “retroflex” and “bunched” American English /r/. *The Journal of the*
1373 *Acoustical Society of America* 123, 4466–4481. doi:10.1121/1.2902168.
- 1374 Zhou, Y., Bhattacharjee, S., Carroll, C., Chen, Y., Dai, X., Fan, J., Gajardo,
1375 A., Hadjipantelis, P.Z., Han, K., Ji, H., Zhu, C., Lin, S.C., Dubey, P.,
1376 Müller, H.G., Wang, J.L., 2022. *Fdapace: Functional Data Analysis and*
1377 *Empirical Dynamics*.

Chapter 9

Study 3: Intergestural timing of English liquids in L2 speech

This study extends the pilot study presented in Chapter 6, overcoming the limitations including a lack of control group (i.e., L1 English speakers) and consideration of the dynamic properties in English liquids. This study provides acoustic and articulatory analyses investigating the production of English liquids /l ɹ/ occurring word-initially and word-finally. The study compares L1 Japanese and L1 English speakers' production, in which articulatory analysis includes the coronal-dorsal timing lag measure, drawing directly on the articulatory data considering the displacement and velocity of tongue tip and dorsum. The findings suggest that L1 Japanese speakers produce target-like allophony in acoustics but not in articulation; coupled with the findings on coarticulation from the earlier chapters, this study supports the running hypothesis that L1 Japanese speakers produce English liquids with less active tongue dorsum gesture than L1 English speakers. More broadly, this study provides evidence that L2 English speakers can acquire the phonetic systems in their L2 involving new phonemes and new prosodic positions, but by resorting the articulatory strategies most available to them rather than acquiring new, difficult strategies. This manuscript is ready-for-submission to *Language and Speech*.

L1 Japanese speakers use a single articulatory strategy to produce onset-coda allophony in L2 English liquids

Takayuki Nagamine¹

¹Department of Linguistics and English Language, Lancaster University

t.nagamine@lancaster.ac.uk

Abstract

1
2 Second language (L2) learners are often challenged with a number of obstacles in
3 acquiring target-like speech production patterns. This study considers L1 Japanese
4 speakers' production of L2 English liquids /l/ and /ɹ/ to investigate how they im-
5 plement position-dependent allophonic variation while overcoming mismatches in (1)
6 the number of liquid phonemes, (2) their allophonic distributions, and (3) the syllable
7 structure between Japanese (L1) and English (L2). Acoustic and articulatory (ultra-
8 sound) analyses of word-initial and -final English /l/ and /ɹ/ produced by thirteen
9 L1 Japanese-L2 English speakers and nine L1 English speakers demonstrate that L1
10 Japanese-L2 English speakers implement the target-like lateral allophony in acoustics
11 but not in articulation. English /ɹ/ shows little effects of syllabic position in itself
12 but a tendency of the polarity effect in the whole liquid system. Overall, the results
13 presented in this study highlight the complex nature of L2 speech production in which
14 L2 speakers utilise a wider variety of phonetic cues to overcome the learning challenges
15 than commonly understood. It also highlights the importance of understanding the
16 English liquid system as a whole given the possible polarity effects between English /l/
17 and /ɹ/.

1 Introduction

19 This study investigates how second-language (L2) learners produce the phonetic systems
20 in their L2 speech production. It is widely-known that L2 learners must overcome various
21 structural differences between their first language (L1) and L2 to produce target-like L2 pro-
22 nunciation. The most obvious is the difference in phonological systems, in which L2 learners
23 need to make a meaningful contrast between phonemes in L2. In addition to this phonologi-
24 cal difference, L2 learners need to learn how to phonetically implement the phonemes in their
25 L2; the phonetic realisations of a phoneme, for instance, can be conditioned by differences
26 in prosodic position such as syllable-onset vs syllable-coda. In the acquisition of target-like
27 L2 segments, therefore, L2 learners are faced with a multitude of challenges, in which they
28 need to adjust articulatory and acoustic cues not only to make meaningful phonological
29 contrasts but also to match the phoneme with appropriate phonetic realisations in a given
30 prosodic condition. The importance of these phonetic details are highlighted in the theo-
31 retical frameworks in L2 speech learning, among which the Speech Learning Model (SLM)
32 explicitly hypothesises that the mapping between L1 and L2 segmental categories occur at
33 the positionally-conditioned allophonic level (Flege, 1995; Flege & Bohn, 2021). Previous
34 empirical work also demonstrates that the allophonic variation in the learner’s L1 could both
35 hinder or facilitate L2 speech perception and production (Colantoni et al., 2023; Llompart
36 et al., 2021; Olsen, 2012; Solon, 2017). Investigating L2 learners’ implementation of within-
37 category allophonic variation in their L2 speech production helps us better understand how
38 the phonetic systems interact between their L1 and L2 at acoustic and articulatory level.

39 This study considers allophonic variation of English liquids /l/ and /ɹ/ produced by L1
40 Japanese-L2 English speakers. This case provides an unique and interesting contribution to
41 the existing body of L2 speech learning research. English liquids provide good testing ground

42 in looking into the within-category allophonic variation given the well-known positional
43 allophony, especially for laterals /l/ (Campbell et al., 2010; Sproat & Fujimura, 1993). L2
44 acquisition of liquid allophony, however, has been mainly investigated with L1-L2 pairings
45 where both languages (1) have a common phoneme (e.g., /l/) while differing in allophonic
46 variation and (2) have similar syllable structure (e.g., Spanish, French and English Colantoni
47 et al., 2023). The case of L1 Japanese speakers' production of L2 English liquid allophony
48 differs from previous research in that (1) Japanese has only one liquid phoneme, /r/, which
49 is phonetically distinct from either English /l/ or /ɹ/ and (2) Japanese allows the liquid
50 consonant to appear only syllable-initially as opposed to English allowing it in both onset
51 and coda positions. This means that L1 Japanese speakers need to make adjustments in their
52 acoustic and articulatory strategies not only (1) to produce novel liquid consonants /l/ and
53 /ɹ/ but also (2) to associate phonetic realisations of English liquids with the appropriate
54 but novel syllabic position. The case of L1 Japanese speakers' acquisition of L2 English
55 allophony corresponds to the NEW scenario in the Second Language Linguistic Perception
56 model (L2LP), which is the "most difficult" scenario in L2 speech learning because of the
57 number of tasks involved in acquiring the segmental contrast (Escudero, 2005, p. 314).
58 Although L2LP is a model of perception, it provides a useful context to explain the patterns
59 in L2 speech production (Nance & Kirkham, 2023). The consideration of the most difficult
60 NEW scenario would uncover the specific mechanisms and challenges that may be involved
61 in the acquisition of L2 phonetic system.

62 The remaining sections are organised as follows. I first review some of the previous
63 research of (1) acquisition of phonetic systems in the target language, (2) allophonic variation
64 of English liquids and (3) L2 acquisition of allophonic variation of English liquids. I then
65 present a speech production study based on acoustic and articulatory (ultrasound) data
66 from 22 participants, including nine L1 English speakers and 13 L1 Japanese-L2 English
67 speakers. The data are analysed acoustically using the distance between the second (F2)
68 and first formants (F1) indexing liquid quality and articulatorily looking into onset-coda

69 differences in midsagittal tongue shape and intergestural timing between tongue tip (TT)
70 and tongue body (TB) gestures. Finally, the findings are discussed in light of the relationships
71 between acoustic and articulatory results, arguing that L2 learners may utilise a wider range
72 of phonetic cues than has been commonly understood to realise L2 phonetic variations.
73 The data and codes used in the analysis are publicly available in the OSF repository at
74 <https://osf.io/5sx7t/>.

1.1 Acquisition of phonetic systems in the target language

75 When learning a second language (L2), L2 learners are faced with many structural differences
76 between the two languages that they need to process. One factor includes segmental targets,
77 and L2 learners often need to acquire necessary acoustic and articulatory cues to make a
78 phonemic contrast in the target language (Chang, 2019). Depending on the L1-L2 pairings,
79 certain phonemic contrasts are harder for L2 learners to learn, including a tense-lax vowel
80 contrast in English for L1 Spanish speakers (e.g., Escudero, 2001) and a contrast between
81 liquids /l/ and /ɭ/ in L2 English for L1 Japanese speakers (e.g., Aoyama et al., 2004).
82 Theoretical frameworks agree that L2 speech learning occurs based on the learner's L1;
83 while in some cases they could simply reuse the phonemic categories that exist in their L1,
84 which would bear little difficulty in acquiring the L2 phonemic contrast, in other cases they
85 need to adjust the existing phonemic boundaries in their L1 to accommodate the phonemic
86 contrast in L2, which could involve some degrees of difficulty (Best & Tyler, 2007; Escudero,
87 2005).

88 Even though languages may have the same phoneme in its phonological inventory, each
89 language differs in the way the phoneme is realised phonetically. In English, for example,
90 voiceless stop consonants get aspirated when they occur at the onset of a stressed syllable
91 (e.g., as in *top*) whereas unaspirated when following a fricative in the onset cluster (e.g.,
92 as in *stop*). Spanish, on the other hand, also has a phoneme /t/ but its phonetic details
93 differ, such that a singleton [t] does not involve aspiration (Flege, 1991). The difference

94 in aspiration can be observed acoustically as voice onset time (VOT) and in intergestural
95 timing between laryngeal (voicing) and the constrictions in the oral cavity. The acquisition
96 of allophonic variation is an important aspect in L2 speech learning, which has implications
97 for the mechanisms behind ‘foreign accents’ in L2 speech (Kirkham & McCarthy, 2021;
98 Kochetov, 2022; Solon, 2017).

99 How a phoneme is phonetically realised is different across languages, and previous re-
100 search shows that L2 acquisition of allophonic variation is influenced by the L1-L2 differ-
101 ences in the phonemic status and allophonic distributions. Rhotics are well-known for a wide
102 range of differences in phonetic realisations across languages (Lindau, 1985). The rhotic con-
103 sonant /r/ in German shows free allophonic variation, meaning that the surface realisation
104 is not conditioned by phonological factors such as syllabic position, showing speaker-specific
105 and regional variations ranging from the alveolar trill [r], the uvular trill [R] to the uvular
106 fricative [ʁ] (Wiese, 2011). In acquiring L2 German rhotics, Llompart et al. (2021) demon-
107 strates through eye-tracking experiments that L1 Italian and French speakers’ perception of
108 word-initial /r/ in L2 German is influenced by the ‘canonical’ realisations of /r/ in their L1s;
109 L1 Italian/L2 German listeners, for example, were better at recognising the L2 words when
110 the /r/ is realised as an alveolar trill, one of the canonical realisations in Italian. Similarly,
111 an alveolar tap, for instance, is a phoneme in Spanish but an allophone of /t/ in Ameri-
112 can English, and L1 American English speakers produce alveolar taps in L2 Spanish more
113 accurately in syllables following the stressed syllable, the environment that an alveolar tap
114 appears as an allophonic variation of /t/ in American English (Olsen, 2012). These studies
115 suggest that an interaction of the phonetic structures between the learners’ L1 and L2 could
116 either hinder or facilitate L2 speech learning.

117 Theoretically, the Speech Learning Model and its revised version specifically posit that
118 L2 speech learning occurs at the level of positional allophones (Flege, 1995; Flege & Bohn,
119 2021). Under their view, L2 speech production accuracy depends on how L2 sounds are clas-
120 sified in relation to the closest L1 sound through equivalence classification. They hypothesise

121 that representations of both L1 and L2 sound categories co-exist in so-called ‘common pho-
122 netic space’. A new L2 category is likely to be formed, for instance, when the L2 learner
123 perceives it to be sufficiently auditorily distinct from the closest L1 counterpart, which pro-
124 vides a necessary condition for accurate production of the L2 sound (Nagle & Baese-Berk,
125 2022). In contrast, L2 categories are merged when L2 sounds are perceived to be similar
126 to the closest L1 category, resulting in both L1 and L2 sounds being perceptually linked
127 and becoming ‘diaphones’ (Chang, 2019). In this case, production of L1 and L2 sounds are
128 approximated with each other, resulting in ‘accented’ L2 production (Chang, 2019; Nagle
129 & Baese-Berk, 2022). The hypothesis that L2 categories are formulated at the positionally-
130 sensitive allophonic level is corroborated by previous findings that L2 learners’ perceptual
131 gains in one position does not necessarily generalise to other positions (Iverson et al., 2005)
132 and that L2 learners do not necessarily acquire all the acoustic cues at the same time to
133 produce allophonic variation of a given L2 phoneme (Colantoni & Steele, 2008).

1.2 Allophonic variation of English liquids

134 English liquid consonants /l/ and /ɭ/ exhibit relatively clear positionally-conditioned allo-
135 phonic variation, especially in the case of laterals. English /l/, for instance, is an alveolar
136 lateral approximant, whose articulation involves occlusions made in the alveolar or dental
137 region, accompanied by the airflow around one or both sides of the tongue (Ladefoged &
138 Maddieson, 1996, p. 183). In acoustics, the pre-vocalic syllable-onset /l/s exhibit a higher
139 second formant (F2) and a lower first formant (F1), thus a greater distance between the
140 F2 and F1 compared to the post-vocalic, syllable-final /l/s (Narayanan et al., 1997; Sproat
141 & Fujimura, 1993). The relationships between the F1 and F2 frequencies characterise the
142 ‘clear’ (/i/-like) and the ‘dark’ (/u/-like) percept of laterals occurring at different syllable
143 positions (Recasens, 2012). Note that the actual phonetic implementation of the lateral
144 darkness varies across dialects of English and some dialects have overall ‘dark’ realisations
145 as in Leeds (Carter & Local, 2007) and in American English (Recasens, 2012), as well as

146 overall ‘clear’ realisations in Newcastle (Carter & Local, 2007). Although the magnitude of
147 the overall lateral quality and the onset-coda distinction differ across languages and dialects,
148 there seems to be a cross-linguistic tendency that syllable-initial /l/s is ‘clearer’ than the
149 final /l/s (Gick et al., 2006; Kirkham et al., 2020; Recasens, 2012).

150 The clearness and darkness of /l/ can also lead to articulatory differences, specifically
151 in terms of spatiotemporal coordination of the tongue tip (coronal) and the tongue body
152 (dorsal) gestures. Laterals involve “the alveolar contact, inward-lateral compression, and
153 convex shaping of the middle and posterior tongue body” (Narayanan et al., 1997, p. 1074).
154 Syllable-final laterals exhibit a larger magnitude of dorsal gesture, and it reaches its max-
155 imum retraction prior to the tongue tip reaching the maximum displacement (Sproat &
156 Fujimura, 1993). In contrast, the syllable-initial, ‘clearer’ /l/s often involve either a syn-
157 chronous timing of the two gestures or the coronal gesture preceding the dorsal gesture
158 (Proctor et al., 2019; Sproat & Fujimura, 1993). The coordination pattern between the
159 coronal and dorsal gesture could signal the degree of syllable affinity of the consonant, given
160 that such a gestural coordination pattern of syllable-initial and syllable-final consonants can
161 also be observed for laterals across languages (Gick et al., 2006) as well as the other class of
162 sonorants (e.g., nasals; Krakow (1999)). Laterals in English are considered to form ‘extrin-
163 sic’ allophony, in which initial and final laterals have distinct articulatory targets (Recasens,
164 2012). According to this view, it is expected that L1 English speakers included in this study
165 would exhibit distinct midsagittal tongue shapes and gestural coordination patterns between
166 the initial and final laterals, leading to clear acoustic differences.

167 The discussion on the onset-coda allophony for English /ɹ/ is, on the other hand, mostly
168 centred on tongue shape and little research has compared acoustics of English /ɹ/ across
169 different syllabic positions. English /ɹ/ is well-known for its diversity in tongue shape, with
170 a tongue-tip-up, ‘retroflex’ to a tongue-tip-down, ‘bunched’ variant constituting a whole
171 spectrum (Alwan et al., 1997; Delattre & Freeman, 1968; King & Ferragne, 2020; Mielke
172 et al., 2016; Tiede et al., 2004; Zhou et al., 2008). In rhotic varieties of English, including

173 American English, retroflex tongue shape tends to be favoured pre-vocally whereas the
174 bunched post-vocally (Mielke et al., 2016). Unlike English /l/, however, the allophony of
175 English /ɹ/ seen in tongue is usually imperceptible, resulting in similarly low F3 frequency
176 (Mielke et al., 2016; Zhou et al., 2008). When looking at the English liquid system as a
177 whole, nevertheless, it has been claimed that the spectral properties of laterals and rhotics
178 exhibit the polarity effects; regardless of absolute realisations, Carter (2002) found that initial
179 laterals are ‘clearer’ than initial rhotics whereas final laterals ‘darker’ than final rhotics in
180 rhotic varieties of British English spoken in Northern Ireland and Scotland. These studies
181 suggest that positional allophony of English /ɹ/ might be exhibited in tongue shape and in
182 the liquid clearness relative to that of laterals.

183 Some previous studies also demonstrate that English /ɹ/ may also exhibit a similar onset-
184 coda distinction in terms of intergestural timing to that of English /l/ (Campbell et al., 2010;
185 Gick, 1999; Proctor et al., 2019). The two tongue configurations, retroflex and bunched,
186 commonly result in three constrictions along the vocal tract; palatal, pharyngeal and labial
187 (Alwan et al., 1997; Harper et al., 2020). Previous research claims that English /ɹ/ also
188 exhibits a similar gestural coordination pattern with that of English /l/; syllable-initial /ɹ/,
189 for instance, shows a “front-to-back” coordination, such that the labial constriction precedes
190 a coronal constriction at the palatal region, followed by or timed synchronously with the
191 tongue root constriction into the pharynx (Campbell et al., 2010). Also similarly to English
192 /l/, a reversed pattern of the initial /ɹ/ can be observed at the syllable-final position, such
193 that a posterior lingual gesture precedes an anterior lingual gesture (Campbell et al., 2010;
194 Proctor et al., 2019). The sequential nature of the coronal and dorsal gestures, however,
195 remains relatively unclear for /ɹ/ in contrast to /l/, with mixed findings of previous research
196 demonstrating that the coronal and dorsal gestures are achieved simultaneously for initial
197 (Proctor et al., 2019) or final rhotics (Gick & Campbell, 2003).

198 To summarise, English /l/ shows a relatively clear pattern of onset-coda allophony, corre-
199 lating with acoustic signals including the F2 frequency and articulatory patterns such as the

200 degree of tongue dorsum retraction and the intergestural coordination between the coronal
201 and dorsal gestures. Although English /ɹ/ may show similar patterns of positional allophony
202 to that of English /l/, previous research suggests that there is a wide range of individual
203 variation; some speakers may use different tongue shapes for English /ɹ/ occurring differ-
204 ent word positions whereas others may favour one tongue shape consistently throughout
205 (Delattre & Freeman, 1968; Mielke et al., 2016).

1.3 This study: L2 acquisition of allophonic variation of English liquids

206 In this study I investigate L1 Japanese-L2 English speakers' production of positional al-
207 lophony of English liquids. L2 acquisition of the allophonic variation of English liquids is a
208 well-researched case in L2 speech learning research, especially the lateral /l/. Previous re-
209 search shows that learners can make a contrast between onset and coda /l/s in a target-like
210 manner, but the specific phonetic implementations seem to be influenced by the lateral qual-
211 ity in their L1. This results in an overall clearer realisation across the syllabic positions in the
212 production of L2 English laterals by Spanish-English bilinguals (Barlow et al., 2013; Colan-
213 toni et al., 2023), Korean-English bilinguals (Chung & Kim, 2021; Hwang et al., 2019) and
214 Sylheti-English bilinguals (Kirkham & McCarthy, 2021). The opposite is also true, where L1
215 American English-speaking learners of Spanish produced Spanish /l/ with an overall darker
216 realisation than L1 Spanish speakers (Solon, 2017). However, the English-like positional
217 effect was only observed for learners who were less proficient in L2 Spanish, suggesting that
218 it is possible to inhibit the L1-like positional allophony to realise more Spanish-like (lack
219 of) positional allophony (Solon, 2017). These studies suggest that, while bilingual and L2
220 speakers classify the L2 English laterals as similar sounds to the lateral categories in their
221 L1, they could still establish two distinct categories between the onset and coda laterals
222 (Barlow et al., 2013; Hwang et al., 2019).

223 Previous research on L2 acquisition of English /ɹ/ shows that L2 learners employ a wide

224 range of tongue configurations to produce L2 English /ɹ/, analogous to the variation attested
225 in the L1 English speakers' production. L1 French learners of English, for example, used
226 tongue-tip-down tongue configurations more frequently for the coda /ɹ/ than the initial
227 /ɹ/, a pattern reported for the L1 English speakers (Léger et al., 2023). Similarly, L1
228 Japanese learners of English employ a wide variety of tongue shapes, from a single strategy
229 for both English /l/ and /ɹ/ to developing a stable tongue shapes for each /l/ and /ɹ/ with
230 a tendency of favouring tongue-tip-up tongue shape for English /ɹ/ (Moore et al., 2018).
231 In contrast, Mandarin-English bilingual speakers less favoured the retroflex tongue shape
232 than the bunched tongue shape, especially for the participants with a lower proficiency in
233 English who almost consistently used bunched tongue shapes across prevocalic, syllabic and
234 postvocalic positions (Chen et al., 2024). These studies overall highlight that the tongue
235 shape complexity attested in the L1 English varieties could also hold true in the context of
236 L2 speech learning.

237 As previously noted, the acquisition of L2 allophony can be challenging depending on the
238 differences in the allophonic variation between L1 and L2. Previous research, however, only
239 considers L1-L2 pairings in which both languages have a common phoneme (e.g., laterals
240 /l/) but differ in terms of allophonic distribution. What remains unclear is the acquisition
241 in which the learner's L2 has a more complex phonological and allophonic structures than
242 their L1. In other words, previous research has mostly investigated the SUBSET (i.e., a single
243 nonnative sound corresponds to more than one categories in the learner's L1) or the SIMILAR
244 scenarios (i.e., learners already have two separate categories in their L1 for a nonnative
245 phonemic contrast) in the Second Language Linguistic Perception (L2LP) model (Escudero,
246 2005; van Leussen & Escudero, 2015). Nance and Kirkham (2023) considered the SUBSET
247 scenario in L1 Gaelic speakers' production of L1 Gaelic and L2 English laterals, investigating
248 how L1 Gaelic speakers adjust the larger lateral system (involving a three-way contrast) in
249 Gaelic to produce L2 English laterals (a single phoneme with position-sensitive phonetic
250 implementation syllable-initially and syllable-finally). They found that L1 Gaelic speakers

251 developed separate acoustic and articulatory strategies in their production of L1 Gaelic and
252 L2 English laterals, with L2 English laterals showing onset-coda allophony, suggesting that
253 L2 speakers re-adjust the phonetic systems in their L1 to conform to the L2 phonetic system.

254 In contrast, this study considers L1 Japanese speakers' production of English liquid
255 allophony, one of the cases of the NEW scenario. In this scenario, L2 learners have to acquire
256 necessary acoustic and articulatory cues to (1) realise differences in equivalent phonemic
257 categories between L1 and L2 and (2) implement language-specific phonetic allophonic rules;
258 because of the number of tasks involved, it is predicted to be the "most difficult" scenario
259 in L2 speech learning (Escudero, 2005, p. 314). Although the L2LP model is a model
260 of speech perception and the majority of the previous research has accordingly focussed on
261 speech perception (e.g., Escudero, 2001; Yazawa et al., 2020). Other theoretical frameworks,
262 including both perception- and production-based models, also agree that this is one of the
263 most difficult scenarios in L2 speech learning (e.g., Best & Strange, 1992; Flege et al., 2021).
264 It is thus fruitful to extend this line of research to speech production to better understand
265 how L2 learners attempt to overcome these structural differences between the two phonetic
266 systems in their L1 and L2 (Nance & Kirkham, 2023). My study combines the acoustic and
267 articulatory data to seek to explain how L2 speakers overcome the substantial difficulty in
268 learning production of the novel L2 phonetic system in the NEW scenario.

269 In the context of the current study, L1 Japanese-L2 English speakers must overcome L1-
270 L2 differences in (1) phoneme inventory, (2) a subsequent allophonic realisation rule, and (3)
271 syllable structure and phonotactics. Japanese and English do not match in terms of both the
272 number and the identity of the liquid phonemes. Japanese has one liquid phoneme, usually
273 considered as a rhotic /r/, whereas English has two: a lateral /l/ and a rhotic /ɹ/. Japanese
274 /r/, however, is canonically realised as an alveolar tap or flap [ɾ], as opposed to approximant
275 realisations in English (Vance, 1987, 2008). Although the lateral and rhotic approximants [l]
276 and [ɹ] could still surface in Japanese, these are considered to be free allophones arising from
277 stylistic and idiosyncratic variation without any prosodic conditioning (Arai, 2013; Kawahara

278 & Matsui, 2017; Morimoto, 2020). This means that, compared to the L1-L2 pairings sharing
279 a common phoneme, learning liquid allophony in English would be a difficult task for L1
280 Japanese speakers, as they need to acquire the allophonic rules that do not exist in their
281 L1, in addition to the general difficulty associated with learning acoustic and articulatory
282 implementations of English /l/ and /ɾ/ (Kochetov, 2022; Tsui, 2012).

283 It has been shown that some of the L1 Japanese learners of English with intermediate
284 and advanced English proficiency could indeed show the onset-coda distinction in their pro-
285 duction of onset and coda laterals in terms of the F2-F1 measure in acoustics (Nagamine,
286 2022) and the anterior linguopalatal contact using the electropalatography (EPG; Kochetov,
287 2022). Nagamine (2022), however, did not find any articulatory correlates of the onset-coda
288 allophony in the ultrasound data, which could be due to the possibility that a mid-point
289 analysis might not have been able to capture the dynamic nature of lingual articulation
290 involved in laterals. In addition, the study only contains L1 Japanese speakers' population
291 and it is thus unclear how 'target-like' the participants' production was. The EPG data in
292 Kochetov (2022) shows a wide range of individual variation in the degree of anterior lin-
293 guopalatal contact, especially for the coda laterals. However, the EPG data provides only
294 a partial view in that the method registers the linguopalatal contact, meaning that the
295 tongue dorsum posture, an important articulatory correlate in understanding the positional
296 allophony of English liquids, remains unclear.

297 In addition, Japanese and English differ in syllable structure and as a consequence, L1
298 Japanese speakers need to overcome differences in phonotactics. English allows a wide range
299 of syllable structures, with a possibility of complex consonant clusters in both onset and coda
300 positions. Japanese, on the other hand, has a relatively restricted set of possibility of syllable
301 structures. The majority of Japanese syllables are CV structures and only allows nasals /N/,
302 a mora obstruent /Q/ (i.e., constituting a geminate obstruent) and a lengthening phoneme
303 /H/ (i.e., turning a short vowel into a long vowel) (Vance, 1987, 2008). Also, Japanese is
304 considered as a mora-timing language, in which a mora is a phonological unit distinguishing

305 the syllable weight between e.g., CV (one mora) and CVC (two morae) structures (Otake,
306 2015). This could result in re-syllabification of the word-final consonants; for instance, vowel
307 epenthesis after the word-final consonant is commonly attested in loanword adaptation from
308 English into Japanese in L2 English production (Kubozono, 2015; Li & Juffs, 2014).

309 Differences in L1 phonotactic knowledge and syllable structures could influence how ar-
310 ticulatory gestures are organised within a syllable. Previous research shows that L2 speakers
311 produce nonnative sequences of consonant clusters with different gestural timing; L1 Amer-
312 ican English speakers, for example, showed a trace of ‘transitional schwa’ in the word-initial
313 /zC-/ sequence that is not permitted in American English (Davidson, 2005, 2006). More
314 broadly, as mentioned earlier, the coordination of coronal and dorsal gestures in English
315 liquids /l/ and /ɹ/ may follow a principle of “gestural affinity”, seen across classes of sounds
316 such as nasals, stops and liquids, that a tighter constriction is attracted to the syllable margin
317 whereas a wider constriction to the syllable nuclei (Krakow, 1999; Sproat & Fujimura, 1993,
318 p. 306). A lack of syllable-final liquids in Japanese could, therefore, influence the production
319 strategies that L1 Japanese speakers employ in signalling the onset-coda allophony of English
320 liquids; a lack of syllable-final liquids in L1 Japanese would mean that L1 Japanese speakers
321 may resort to a single articulatory strategy to produce both syllable-initial and -final liquids.

322 In this study, I improve on the designs of previous work and our understanding of the
323 nature of L2 speech learning by adding new evidence as to how segmental (i.e., phonemic
324 status, allophonic realisations in L2) and prosodic (i.e., differences in syllable structure) fac-
325 tors may influence the acquisition of L2 allophony. Using L1 Japanese-L2 English speakers’
326 production of English liquid allophony as a test case, I aim to address the following questions;

- 327 1. Do L1 Japanese-L2 English speakers make a contrast between syllable-initial and -final
328 tokens of English liquids?
- 329 2. If so, what acoustics/articulatory strategies do they use? (Do they use acoustic/articulatory
330 cues in a target-like manner?)

331 In order to address these questions, I combine acoustic and articulatory data to assess the
332 production of English liquid allophony by L1 Japanese-L2 English speakers and L1 English
333 speakers. The acoustic measure is the static measurement of F2-F1 at the liquid mid-
334 point, which indexes the degree of liquid ‘darkness’. The articulatory measure derives from
335 the ultrasound tongue imaging technique that enables a holistic imaging of the midsagittal
336 tongue shape. Using ultrasound, I compare midsagittal tongue shape and intergestural tim-
337 ing between onset and coda liquid tokens to investigate how speakers realise the onset-coda
338 allophony. In addition, I also look into the tongue shape taxonomy for English /ɹ/ to inves-
339 tigate whether differences in tongue shape correlate with spatiotemporal patterns in English
340 /ɹ/ allophony.

2 Methods

2.1 Participants

341 The data set to be analysed in this study comprises of a simultaneous acoustic and high-
342 speed ultrasound recording from 22 speakers. This includes nine L1 English speakers (seven
343 female, two male) aged between 22 and 39 years ($M = 28.56$ years, $SD = 4.88$) and 13 L1
344 Japanese-L2 English speakers (six female, seven male) aged between 18 and 21 years (M
345 $= 19.69$, $SD = 0.95$) at the time of recording. This is a subset of a larger data collection,
346 in which originally 55 speakers (41 L1 Japanese-L2 English speakers and 14 L1 English
347 speakers) were recorded. The speakers included in this study are chosen on the basis of overall
348 clarity of ultrasound image and the confidence in tongue spline estimation; only speakers
349 whose ultrasound data shows the whole midsagittal tongue shape absolutely clearly from
350 the tongue tip to the tongue root have been chosen. Although the tongue spline estimation
351 programme, the DeepLabCut (DLC) plug-in in the Articulate Assistant Advanced (AAA)
352 software, is capable of estimating tongue tip location even if it is obscured by the mandible
353 shadow, intergestural timing analysis conducted in this analysis requires tongue tip to be

354 imaged very clearly.

355 In the participant population included in the current study, L1 English speakers were
356 born and raised at least until 13 years old either in the US ($n = 5$) or in Canada ($n = 4$)
357 and resided in the UK for their study or work when recording took place. They all identify
358 themselves as fluent L1 English speakers and rated their fluency being seven (1 = I do not
359 speak English at all., 7 = No problem in using English in daily life.) L1 Japanese-L2 English
360 speakers were all undergraduate students enrolled at a university in the central and western
361 Japan. They were born and raised in Japan using Japanese, and they studied English
362 mostly through school curriculum with mean overseas experience being approximately three
363 weeks (0.71 months, $\sigma = 1.47$). The wide SD for the overseas study results from two L1
364 Japanese students who have had overseas study experience for four months, whereas other
365 three participants have had one to two-week overseas experience. The rest has never studied
366 overseas. The mean subjective fluency rating was 3.85 ($\sigma = 1.21$). The L1 Japanese-speaking
367 participants in this study represent typical English-as-a-foreign-language (EFL) learners.

2.2 Materials

368 The materials in this study are 12 monosyllabic CVC words starting or ending with English
369 liquids /l r/. The vowel environment is kept consistent to /i/ in order to keep vowel quality
370 maximally similar between Japanese and English, even in the case of L1 substitution for
371 L1 Japanese-L2 English speakers. The target words are such that onset and coda words
372 are the mirror image of each other, sharing the same sets of phonemes and differing only
373 in their sequences (cf. Campbell et al., 2010; Gick et al., 2006). The other consonant is
374 always a labial or a labiodental consonant (i.e., /f/, /p/, /v/) in order to minimise lingual
375 coarticulation while facilitating acoustic segmentation. The word list is shown in Table 1

Table 1: Word list used in the production experiment

	Lateral /l/		Rhotic /ɹ/	
	onset	coda	onset	coda
/p/	peel	leap	peer	reap
/f/	feel	leaf	fear	reaf
/v/	veel	leave	veer	reeve

2.3 Procedure

376 Data collection was conducted in a quiet room at the universities in Japan for L1 Japanese
 377 speakers and in a sound-proof recording booth at the universities in the UK for L1 English
 378 speakers between October and December 2022. At the time of recording L1 Japanese speak-
 379 ers, Covid-19 measures were still in place mandating air ventilation at all times, so there was
 380 minor fan noise in the audio recording for some of the L1 Japanese speakers’ data.

381 Midsagittal ultrasound tongue imaging data were collected using a Teleded MicrUS
 382 system with a 64-element probe of 20 mm radius. The prompt presentation and recording
 383 was made on the Articulate Assistant Advanced (AAA) software version 220.5.1 (Articulate
 384 Instruments, 2022). The participants wore an UltraFit probe stabilisation headset to stabilise
 385 the probe relative to the head movement (Spreafico et al., 2018). Recording parameters for
 386 ultrasound tongue imaging were optimised per speaker at the start of the recording, with the
 387 field of view ranging between 80% and 100%, the depth between 80mm and 100mm, and the
 388 probe frequency between 2MHz and 4MHz. This results in the frame rate of approximately 80
 389 frames per second. The audio recording, also recorded on the AAA software, was made using
 390 an Opus 55 MK ii condenser microphone attached to the UltraFit headset, pre-amplified
 391 and digitised at 44.1 kHz with 16-bit quantisation using a Sound Device USB Pre-2 audio
 392 interface.

393 In the recording venue, participants sat in front of a laptop computer displaying the
 394 prompt on the AAA software. After participants were informed about the experiment and
 395 signed on the consent form, I fitted the UltraFit headset and explored an optimal recording

396 setting for each participant. I then recorded each participant’s bite plane to standardise
397 the coordinate system across participants in the analysis stage using a thin plastic plate
398 (Scobbie et al., 2011) as well as their palate shape by asking them to swallow some water.
399 These preparatory phase took approximately 20 minutes, although it took longer for some
400 participants due to difficulties in headset fitting and identifying recording parameters for
401 clear tongue images. Note that the participant’s side-profile lip video was also recorded
402 using a small camera mounted on the UltraFit headset extension but the lip data will not
403 be analysed or presented in this article.

404 During the recording, participants read the individual target words one by one. The
405 order of the target words was randomised but kept consistent across participants. The
406 prompt display was delayed by 1000 ms after the onset of ultrasound recording in order to
407 prevent participants from taking preparatory tongue posture for the initial consonant. The
408 instructions associated with recording were given entirely in English for both L1 Japanese
409 and L1 English speakers by the first author (an L1 Japanese speaker) in consideration of the
410 participant’s language mode (Grosjean, 2008). Participants produced each token between
411 two to five times. The ultrasound recording of the English words lasted for approximately
412 30 to 45 minutes. Due to a strict turn-around time restriction on the recording venues and
413 software processing time, some speakers only managed to produce a small number of tokens.

2.4 Analysis

414 This study combines acoustic and articulatory analyses to holistically assess the onset-coda
415 allophony in L2 speech production. Acoustic analysis is based on the F2–F1 measure taken
416 at the liquid midpoint, which has been used as a proxy for the degree of tongue dorsum re-
417 traction and thus an index of the overall liquid quality (Howson & Redford, 2021; Kirkham,
418 2017; Sproat & Fujimura, 1993). Articulatory analysis aims to investigate differences be-
419 tween onset and coda liquids in tongue shape and in intergestural timing. Note that the
420 acoustic analysis only considers the static F2–F1 data extracted at the liquid mid-point

421 because the current study only considers one vowel context as opposed to three in Nagamine
422 (2024a) which argued that dynamic analysis would highlight between-group liquid-vowel
423 coarticulatory differences as a function of vowel contexts. The static analysis also allows
424 me to compare the current results with previous research. The timing difference between
425 the two groups of speakers in Nagamine (2024a) would be highlighted in the intergestural
426 analysis.

2.4.1 Acoustic analysis

427 Prior to the analysis, the acoustic signals were downsampled to 22,050 Hz and high-pass
428 filtered with a cut-off frequency of 70 Hz. Automatic phone-level segmentation was then
429 performed using Montreal Forced Aligner (MFA) version 2.0.6 (McAuliffe et al., 2017). The
430 segmented boundaries were then visually inspected and adjusted wherever necessary using
431 Praat (Boersma & Weenink, 2022).

432 The liquid tokens were first classified into two categories of *approximant* and *non-*
433 *approximant* realisations. This is because it is highly likely that some L1 Japanese speakers
434 substitute English /l ɹ/ with Japanese /r/, phonetically realised as an alveolar tap or flap
435 [ɾ], for which spectral analysis may not be optimal due to its extremely short duration and
436 stop-like realisations. L1 Japanese speakers show a wider range of English liquid realisa-
437 tions, including allophonic variations associated with L1 Japanese /r/ (e.g., retroflex lateral
438 approximant [ɭ]; Arai (2013)). For these reasons, the spectral analysis only considers tokens
439 that were broadly classified as approximants, excluding 45 tokens that were classified as
440 non-approximants and 13 tokens that showed unclear formant structures.

441 In acoustic segmentation, English liquids are identified as a steady-state or an approxi-
442 mately steady-state of F2, guided by an abrupt change in amplitude (Lawson et al., 2019).
443 While this is inevitably only an approximation of liquid production that involves various
444 stages, including the transition into and out of the neighbouring vowels (Carter & Local,
445 2007; Nance, 2014), this is a common measure that has been used to characterise the onset-

446 coda allophony and thus facilitates comparison of the results from this study with that of
447 previous research (e.g., Aoyama et al., 2019; Flege et al., 1995). Tokens with poor recording
448 quality were excluded from the analysis when it was deemed difficult to implement reliable
449 acoustic segmentation. Further details of the acoustic segmentation procedure can be found
450 in Nagamine (2024a).

451 Formant frequencies were extracted using Fast Track, a Praat plug-in for automatic
452 formant estimation (Barreda, 2021). Fast Track performs up to 24 formant estimation
453 rounds per token, with varying ceiling frequencies, and autoselects the best-fit settings based
454 on a regression analysis. In this study, Fast Track returned estimations of three formants
455 (F1, F2 and F3) at 11 equi-distant time points during the liquid interval (point 1 = liquid
456 onset, point 6 = liquid midpoint, point 11 = liquid offset), with the range of upper formant
457 frequencies being between 5000-7000 Hz for female speakers and between 4500-6500 Hz for
458 male speakers as per the FastTrack recommendation. Inaccuracies in formant estimation
459 were either corrected by the author's visual inspection and nomination of better analysis or
460 by removing the token from the analysis when none of the analyses were looked reasonable.
461 Twenty tokens of two L1 English speakers (one female and one male) had to be analysed
462 separately due to extremely low F3 frequency for the initial /ɹ/. Also, formants were not
463 estimated for 105 tokens that were shorter than 30 ms as they were automatically removed
464 from FastTrack.

465 The procedure described here yields 962 tokens for analysis, with a detailed breakdown
466 in the number of tokens for each category shown in Table 2. Among the 11 time points,
467 this analysis selects the liquid midpoint (point 6) to characterise the spectral characteristics
468 between the two speaker groups. While spectral analysis at articulatorily-defined event (e.g.,
469 maximal TB displacement) would provide a more accurate picture given a correspondence
470 with the F2–F1 measure, the midpoint F2–F1 measurement is the only consistent data
471 point across the tokens because the articulatory data processing suggested that L1 Japanese
472 speakers do not necessarily show the maximal tongue body (TB) displacement during the

473 acoustic liquid interval.

2.4.2 Articulatory analysis

474 The articulatory analysis is based on the midsagittal tongue shape recorded with ultrasound.
475 Tongue splines were estimated using the DeepLabCut (DLC) plug-in on the AAA software
476 for the whole duration of each recording for each word (Mathis et al., 2018; Wrench &
477 Balch-Tomes, 2022). DLC estimates the tongue surface by tracking 11 key points along the
478 tongue based on the pre-trained neural network model, and the x/y coordinates of the key
479 points were then exported in millimetres. When exporting the data, the x/y coordinates
480 were standardised and rotated relative to each speaker’s bite plane measured prior to the
481 word list recording (Scobbie et al., 2011).

482 The precision in the tongue surface estimation was visually checked by the author, and
483 only the speakers whose images clearly captured the large area of the tongue surface from
484 tongue tip to tongue root were chosen, constituting the participant population in this study.
485 In addition, once the speakers have been chosen, nine further tokens were excluded due to
486 audio-ultrasound synchronisation issues and inaccurate tongue estimation. This results in
487 529 tokens for English /l/ and 527 tokens for /ɹ/, with the further breakdown shown in
488 Table 2.

Table 2: The number of tokens included in the analysis

	Acoustic analysis				Articulatory analysis			
	Lateral /l/ onset coda		Rhotic /ɹ/ onset coda		Lateral /l/ onset coda		Rhotic /ɹ/ onset coda	
L1 English	118	131	90	182	133	132	134	135
L1 Japanese	94	105	67	175	130	134	130	128

489 The analysis window was determined based primarily on the acoustic segmentation ex-
490 plained earlier. The articulatory analysis here considers the interval consisting of (1) the
491 liquid, (2) the vowel and (3) a 350 ms interval padded before the liquid-vowel interval for the
492 initial tokens or after the vowel + liquid interval for the final tokens. The onset and offset of

493 the vowel were identified in the acoustic signals referring to the F2 frequency and amplitude.
494 Given that the liquid interval, as explained earlier, only contains the F2 steady-state, the
495 transition into and out of the vowel is included in the vowel interval. The duration of the
496 350 ms padding was decided in order to capture the tongue movement from the tongue rest
497 position for the word-initial token based on previous research(Nagamine, 2024b). To make
498 the data compatible, a 350 ms padding was added to the final token so that the analysis
499 window captures the tongue movement from the acoustic onset of the vowel, through the
500 final liquid and then back to the speech rest position.

501 The main variable of interest in the articulatory analysis is the intergestural timing
502 between the coronal and dorsal gestures. In order to identify the functional and meaningful
503 tongue regions representing the coronal and dorsal gestures, correlation coefficients for the
504 11 DLC key points were calculated based on the hierarchical correlational clustering using
505 the *cor* function and the *corrplot* package on R. The labals for the 11 DLC key points are
506 shown in Figure 1 and the correlation analyses in Figure 2.

507 The correlation analysis here shows that, for both /l/ and /ɭ/, the DLC points 9-11 and
508 3-6 show higher correlations at the statistically significant level of $p < 0.05$. For /l/, DLC
509 point 1 (vallecula) also correlates with points 9 to 11 strongly ($r = 0.87$), but this is thought
510 to be an artifact given the location of the points. Based on these, my analysis considers the
511 DLC points 9-11 representing the tongue tip (TT) gesture and the points 3-6 the tongue
512 body (TB) gesture.

513 In order to reduce the 3 TT and 4 TB points to 1-dimensional TT and TB signals,
514 Principal Component Analysis (PCA) was computed on each triplet of points using R's
515 *princomp* function. In this sense, PCA is analogous to tangential velocity that captures
516 the tongue movement on both anterioposterior and superoinferior dimensions. The PCA
517 analysis suggests that PC1 and PC2 account for most of the variation present in the data;
518 PC1 70.13% and PC2 18.75% (overall 88.88%) for lateral TT; PC1 74.73% and PC2 17.71%
519 (overall 92.45%) for lateral TB; PC1 76.18% and PC2 19.74% (overall 95.91%) for rhotic

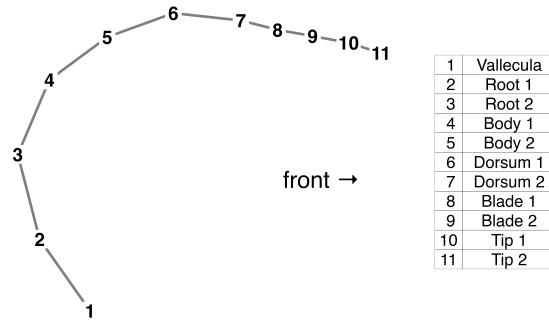


Figure 1: Illustration of the 11 DLC key points along the midsagittal tongue shape, adapted from Wrench and Balch-Tomes (2022, p. 7)

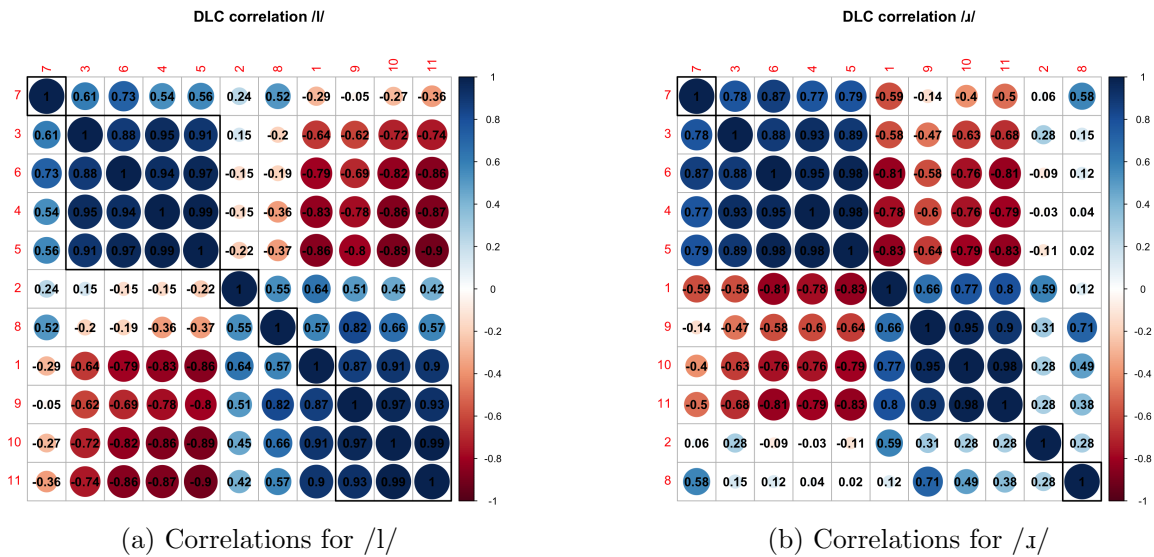


Figure 2: Correlation coefficients among the 11 DLC key points for /l/ (left) and /ɹ/ (right). The numbers in red outside the box indicate the point number estimated by DLC. Darker colours represent stronger correlations, with blue showing positive correlations and red negative. All correlation coefficients shown here are obtained at the statistically significant level ($p < .05$).

520 TT; and PC1 74.14% and PC2 16.31% (overall 90.45%) for rhotic TB.

521 TT and TB gestures were identified based on time-varying changes in the joint PC1+PC2
 522 scores, treated here as an approximation of the TT and TB positional trajectories. Articulation of English /l/, for example, involves tongue tip raising and tongue dorsum retraction,
 523

524 which should correspond to a local maximum peak in TT position and a local minimum
 525 peak in TB position. In order to facilitate gestural identification, the TT and TB positional
 526 trajectories were smoothed using a 5th-order Butterworth filter with a cut-off frequency of
 527 20 Hz. The cut-off frequency was determined after initial exploration; while a lower cut-off
 528 frequency would smooth out the trajectories more aggressively, resulting in smoother trajec-
 529 tories and thus easier identification of local peaks, the initial exploration suggested that a
 530 10-Hz cut-off frequency, as used in Sproat and Fujimura (1993), inevitably shifted the timing
 531 pattern of the local peaks, which is not desirable for the intergestural timing analysis here.
 532 The cut-off frequency was therefore determined on 20 Hz so that a shift in the peak tim-
 533 ing pattern could be minimised while achieving a smoothing of the trajectories to eliminate
 534 smaller noises, as illustrated in Figure 3.

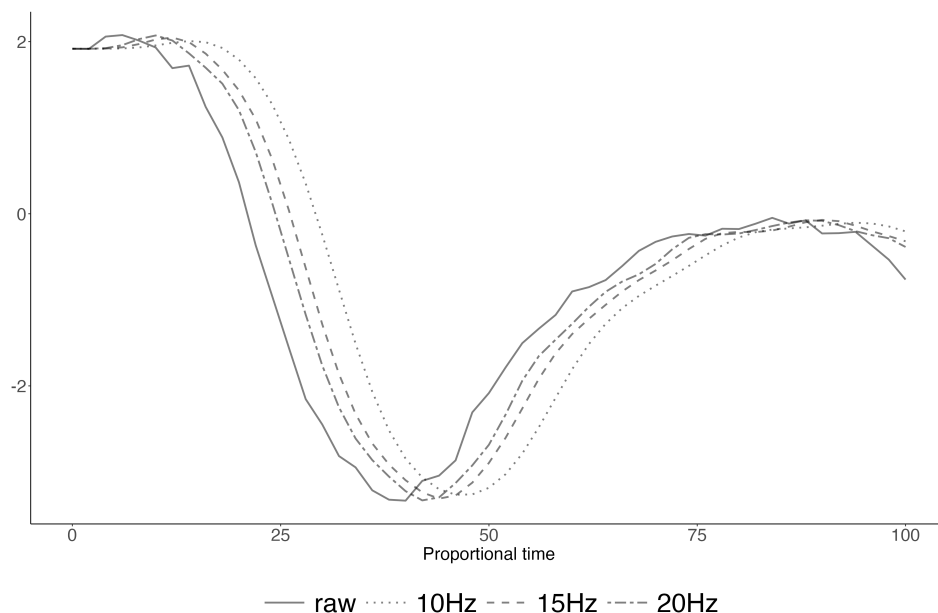


Figure 3: Raw TB displacement trajectory (solid) superimposed by smoothed trajectories with a cut-off frequency of 10Hz (dotted), 15Hz (dashed) and 20Hz (two-dashed) in the production of *feel* by an L1 English speaker.

535 The identification of TT and TB gestures was further facilitated by calculating the ve-
 536 locity profile. This is especially the case for English /ɹ/, where the positional peak (i.e.,
 537 the maximal TT raising) did not correspond well with the timing of maximal constriction

538 in raw ultrasound data, which could be because the maximal constriction for TT tended to
539 be achieved while the overall tongue configuration was low due to the TB retraction and
540 lowering. For this reason, the TT and TB gestures were identified based primarily on the
541 velocity profile, calculated as the first derivative from the positional trajectory using the
542 *central_difference* function in the *tadaR* package (Kirkham, 2024).

543 In identifying the TT and TB gestures, the positional and velocity peaks were first au-
544 tomatically identified using the *findpeaks* function in the *pracma* package (Borchers, 2023).
545 There was no *apriori* knowledge as to where the positional/velocity peaks would be found
546 given that (1) the analysis window was quite broad spanning over the combined interval of
547 liquid and vowel as well as a 350-ms padding and (2) L2 speakers may show different timing
548 patterns from L1 speakers. For this reason, the procedure could not be fully automated.
549 The parameter settings in the *findpeaks* function were decided so that as many peaks as
550 possible were found in the positional/velocity data, and I then visually compared all the
551 positional/velocity peaks against the raw ultrasound video to identify and verify which peak
552 would correspond to the articulatory events of interest. In the case of shoulder peaks, the
553 first frame that corresponds to the change in the trajectory dimension was chosen as a rep-
554 resentative of the maximal displacement unless later frames were deemed to be appropriate
555 based on the author’s visual inspection.

556 For /l/, the positional peaks usually corresponded well with the frames where the maximal
557 tongue tip raising and tongue dorsum retraction were achieved. Nine tokens from one L1
558 Japanese-L2 English speaker (2d57ke) that did not exhibit clear tongue tip raising were
559 considered to be instances of /l/-vocalisation and thus were excluded from the analysis. For
560 /ɹ/, as mentioned earlier, the positional peaks did not correlate well with the achievement of
561 TT maximal displacement, possibly due to the overall lower tongue configuration. Because
562 of this, the TT and TB gestures were identified based on the velocity minima. An example
563 illustration of the positional and velocity trajectories is shown in Figure 4 below.

564 Figure 4 shows that TT position is high during the vowel interval (with yellow back-

565 ground), then it is lowered to prepare for the rhotic articulation. There is a small local peak
 566 during the rhotic interval (with grey background) which corresponds to the TT raising for
 567 /ɹ/, which also corresponds to a local velocity peak in the velocity profile (bottom left). The
 568 tongue then returns back to the rest position in which TT position becomes higher again
 569 because of relatively higher tongue posture in the rest position compared to the rhotic articu-
 570 lation. The TB displacement is relatively clearer; the TB positional data (top right) remains
 571 high for the high vowel /i/ and then lowers and retracts towards the rhotic target, which is
 572 indicated as a local TB minimum in the positional data during the rhotic interval and also
 573 corresponds to a local velocity minimum. Note that the derived trajectories (e.g., Figure 4)
 574 are still quite jagged, in which the sudden discontinuities (especially in the velocity profile)
 575 might have resulted from the sampling rate, but the visual inspection of the raw ultrasound
 576 images ensured that the chosen positional and velocity peaks correspond to the articulatory
 577 events of interest by avoiding identifying an artifact as a meaningful velocity minimum.

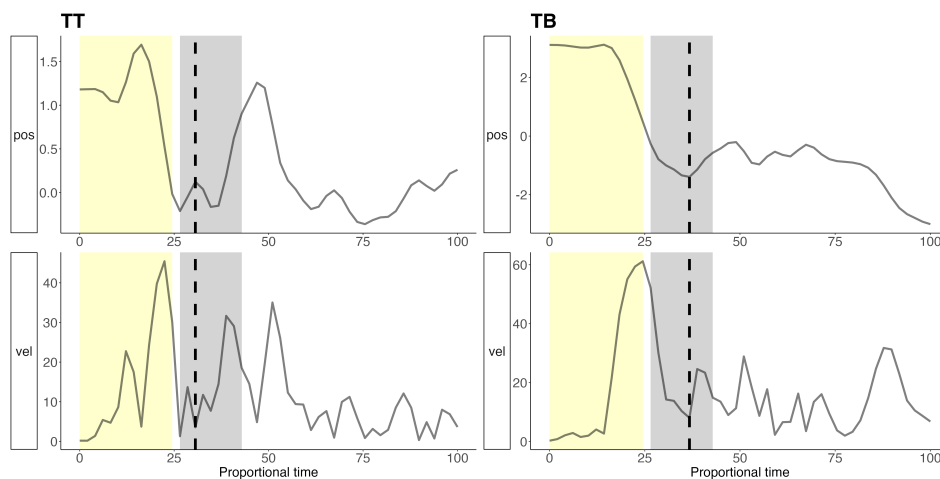


Figure 4: Example of the positional (top) and velocity (bottom) trajectories for TT (left) and TB (right) in the production of *peer* by an L1 English speaker. The interval spans from the onset of the vowel /i/ to the end of the 350-ms window added at the end. The yellow shadow represents acoustic interval for the vowel and the grey shadow for the liquid /ɹ/. The vertical dashed lines represent the frame containing the maximal TT and TB displacement.

578 Finally, the TB lag was calculated as the lag in time corresponding to the ultrasound
 579 frames for the TT and TB maximal displacement following previous studies (Sproat & Fu-

580 jimura, 1993; Ying et al., 2021), such that $TB\ lag = Time\ point\ of\ TT\ extremum - Time$
581 $point\ of\ TB\ extremum$ (Ying et al., 2021, p. 9). A negative lag indicates the activation of
582 tongue tip preceding that of tongue dorsum, whereas a positive lag vice versa. This mea-
583 sure, TB lag, has been shown to characterise the onset-coda distinction for laterals (Sproat
584 & Fujimura, 1993; Ying et al., 2021) and rhotics (Campbell et al., 2010; Gick & Campbell,
585 2003).

586 In addition to the gestural timing measure, tongue shape is classified for English /ɹ/
587 according to the decision tree proposed in King and Ferragne (2020). The tongue shape
588 analysis is based on the raw ultrasound image primarily at the maximal TT displacement,
589 although the participants in this study did not vary their tongue shape for the duration of
590 the rhotic interval. The five categories in King and Ferragne (2020); Curled Up (CU), Tip
591 Up (TU), Front Up (FU), Front Bunched (FB) and Mid Bunched (MB), however, were not
592 sufficient in the L1 Japanese speakers' data due to different articulatory strategies from that
593 of L1 English speakers. There were, for example, nine tokens in which the tongue shape was
594 similar to their production of laterals (four in the coda position, five in the onset position),
595 which could be due to their confusion between /l/ and /ɹ/ (cf. Moore et al., 2018). Similarly,
596 the tongue shape for 40 tokens were identical to that of the neighbouring vowel, making it
597 impossible to determine the tongue shape category. This results mostly from non-rhotic
598 production given that 39 tokens out of 40 occurred word-finally. These tokens were labelled
599 as 'lateral' and 'vowel' respectively but excluded from the statistical analysis.

2.4.3 Statistical analysis

600 Linear mixed-effect model was performed using the *lme4* package (Bates et al., 2015) in
601 order to investigate by-group differences in the way the onset-coda distinction is signalled.
602 Significance testing for the fixed effects was conducted through model comparison via a
603 likelihood ratio test, in which the full model is compared with the nested model excluding
604 the fixed effect of interest. The full model would be chosen as the best model when the model

605 comparison suggests an improvement in the degree of model fit at a statistically significant
606 level with a threshold of $p < .05$ (Winter, 2020). When the two models in model comparison
607 did not differ in the degree of model fit, then a more parsimonious model would be chosen
608 as the best-fit model.

609 For the spectral analysis, the model predicts the distance between F2 and F1 (F2–F1)
610 by the fixed effects of *L1* (two levels: Japanese and English), *position* (two levels: initial and
611 final), and *liquid* (two levels: /l/ and /ɾ/), as well as interactions between these. Random
612 effects include by-participant varying intercepts and the random slopes for *position* and *liquid*
613 for speakers. It also includes by-word varying intercepts and the random slopes for *L1* for
614 the word. The model specification in the lmer notation is:

```
615 lmer(f2f1_z ~ L1 + position + liquid + L1:position + position:liquid + L1:liquid + (1 +  
616 position + liquid|speaker) + (1 + L1|word))
```

617 The intergestural timing was analysed for /l/ and /ɾ/ separately to reduce the complexity
618 of the statistical analysis. The full model for laterals predict the TB lag in millisecond by the
619 fixed effects of L1, position, and the interaction between them. It also includes the by-speaker
620 varying intercept and by-speaker varying slope for position. The model specification for the
621 rhotic full model is the same for the lateral model, except for an addition of the fixed effect
622 of *tongue shape* to investigate how tongue shape differences influence the gestural timing.
623 The model specification is:

```
624 lateral model: lmer(TB_lag_ms ~ L1 + position + L1:position + (1 + position|speaker))  
625 rhotic model: lmer(TB_lag_ms ~ L1 + position + tongue shape + L1:position + (1 +  
626 position|speaker))
```

3 Results

3.1 Acoustic analysis

627 The first analysis concerns with acoustic differences between the onset and coda tokens of
628 English liquids /l/ and /ɭ/. As the index for the liquid darkness, F2–F1 values are extracted
629 at the liquid midpoint as summarised in Table 3 and visualised in Figure 5. Overall, both L1
630 English and L1 Japanese speakers seem to make a distinction between the onset and coda
631 laterals, in which initial tokens show higher F2–F1 values (L1 English speakers: $M = 817.61$
632 Hz, $SD = 272.80$; L1 Japanese speakers: $M = 1029.15$ Hz, $SD = 272.23$) than final tokens
633 (L1 English speakers: $M = 550.37$ Hz, $SD = 127.59$; L1 Japanese speakers: $M = 817.40$
634 Hz, $SD = 342.71$). For the rhotic /ɭ/, in contrast, both speaker groups show lower F2–F1
635 values for initial tokens (L1 English speakers: $M = 783.98$ Hz, $SD = 182.84$; L1 Japanese
636 speakers: $M = 965.28$ Hz, $SD = 289.26$) than for final tokens (L1 English speakers: $M =$
637 1040.52 Hz, $SD = 213.29$; L1 Japanese speakers: $M = 1042.66$ Hz, $SD = 240.19$).

638 Statistical analysis using linear mixed-effect modelling based on within-speaker z -normalised
639 F2–F1 values demonstrate that the interaction between *L1* and *position* is not statistically
640 significant ($\chi^2(1) = 1.75$, $p = 0.19$), suggesting that both the F2–F1 pattern is similar be-
641 tween L1 Japanese and L1 English speakers. The interaction between *position* and *liquid*
642 improves the model fit at the statistically significant level ($\chi^2(1) = 13.72$, $p < 0.001$), re-
643 flecting the opposite pattern in F2–F1 between /l/ and /ɭ/. Finally, the model comparison
644 suggests a significant interaction effect between *L1* and *liquid* ($\chi^2(1) = 4.36$, $p = 0.04$),
645 which could indicate an overall realisational difference of the liquid acoustics between L1
646 Japanese and L1 English speakers.

647 A post-hoc pairwise comparison using the *emmeans* package shows a statistically signif-
648 icant difference between onset and coda tokens of /l/ ($\beta = 1.11$, $SE = 0.28$, $t(25) = 3.98$, p
649 < 0.001) but not for /ɭ/ ($\beta = -0.41$, $SE = 0.29$, $t(23.8) = -1.42$, $p = 0.17$). The pairwise
650 comparison for the interaction between *L1* and *liquid*, however, does not show a statistically

651 significant difference between L1 Japanese and L1 English speakers for /l/ ($\beta = -0.45$, SE
652 $= 0.26$, $t(32.1) = -1.75$, $p = 0.09$) or /ɾ/ ($\beta = 0.37$, $SE = 0.19$, $t(26.6) = 1.81$, $p = 0.08$).

653 To summarise, the acoustic analysis indicates that both L1 English and L1 Japanese
654 speakers make a clear onset-coda contrast for laterals, in which the initial tokens exhibit
655 clearer realisations with higher F2–F1 values than the final tokens. Although English /ɾ/
656 shows a reversed trend, in which the initial tokens exhibiting higher F2–F1 than the final
657 tokens, statistical analysis did not yield any statistically significant difference. Finally, the
658 two groups of speakers, L1 Japanese and English speakers, seemed to behave similarly for
659 both English /l/ and /ɾ/.

Table 3: Mean F2–F1 (Hz) at the liquid midpoint for word-initial and -final tokens of English /l/ and /ɾ/

L1	liquid	position	Mean F2–F1 (Hz)	SD
English	/l/	initial	817.61	272.80
		final	550.37	127.59
	/ɾ/	initial	783.98	182.84
		final	1040.52	213.29
Japanese	/l/	initial	1029.15	272.73
		final	817.40	342.71
	/ɾ/	initial	965.28	289.26
		final	1042.66	240.19

3.2 Midsagittal tongue shape

660 We now turn to the articulatory data involving midsagittal tongue shape collected using
661 ultrasound tongue imaging. In order to draw an overall picture, here I present the midsagittal
662 tongue shape extracted at the liquid midpoint to compare acoustics (presented above) and
663 articulation. Tongue shapes are visualised in Figure 6 for English /l/ and Figure 7 for English
664 /ɾ/. Note that an additional comparison of tongue shape is included in Appendix C in which
665 tongue shape is extracted at maximal TB displacement as this time point may represent the
666 onset-coda allophony more clearly given that they differ in the degree of TB displacement

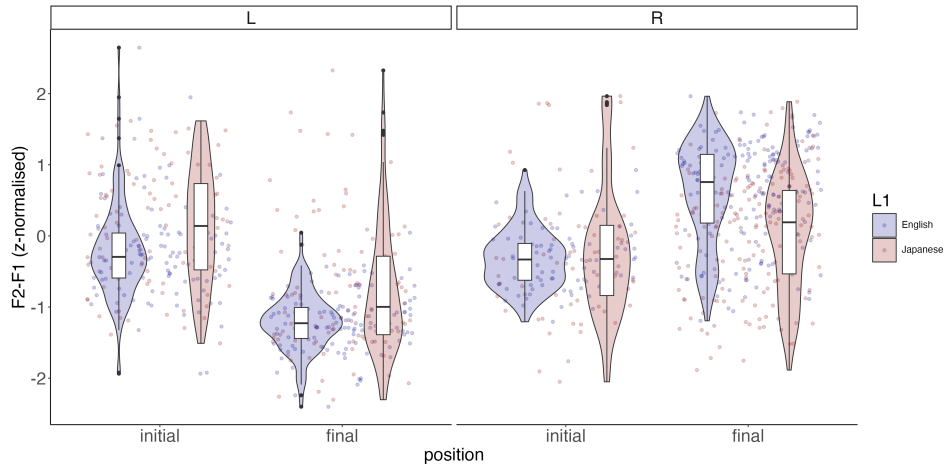


Figure 5: F2–F1 at the liquid midpoint for English /l/ (left) and /ɭ/ (right). Blue represents L1 English speakers and red indicates L1 Japanese speakers. F2–F1 values are within-speaker z -normalised.

667 (cf. Campbell et al., 2010; Lee-Kim et al., 2013)

668 L1 English speakers make a relatively clear onset-coda contrast for English /l/, in which
 669 their overall tongue shape is constantly lower for the final tokens than for the initial tokens.
 670 The difference is particularly pronounced around the tongue blade. The degree of tongue
 671 retraction varies across speakers, in which fivespeakers (4ps8zx, 5jzj2h, bwizh, jcy8xi, and
 672 xub9bc) show a greater degree of tongue body retraction in the final position than in the
 673 initial position. Although the rest of the speakers show a smaller difference between the
 674 onset and coda laterals, the tendency is similar to that of the first five speakers in that
 675 tongue blade is lowered for the final tokens compared to the initial tokens.

676 L1 Japanese speakers, on the other hand, exhibit somewhat a complicated pattern and
 677 the magnitude of the onset-coda difference is relatively small. Speakers birw55 and fdg95u
 678 exhibits a target-like pattern, in which the tongue blade region is lower for the final tokens
 679 than for the initial tokens. Speaker 3bcpyh shows a slight tongue body retraction for some
 680 of the final tokens. Further two speakers, 2zy9tf and cdsju7, show a reversed pattern such
 681 that their tongue is lower and more retracted for the initial tokens than for the final tokens.
 682 Finally, the rest of the speakers employ almost identical tongue shape for both onset and

683 coda /l/s, suggesting that they may not differentiate the onset and coda laterals by means
 684 of the tongue body retraction.

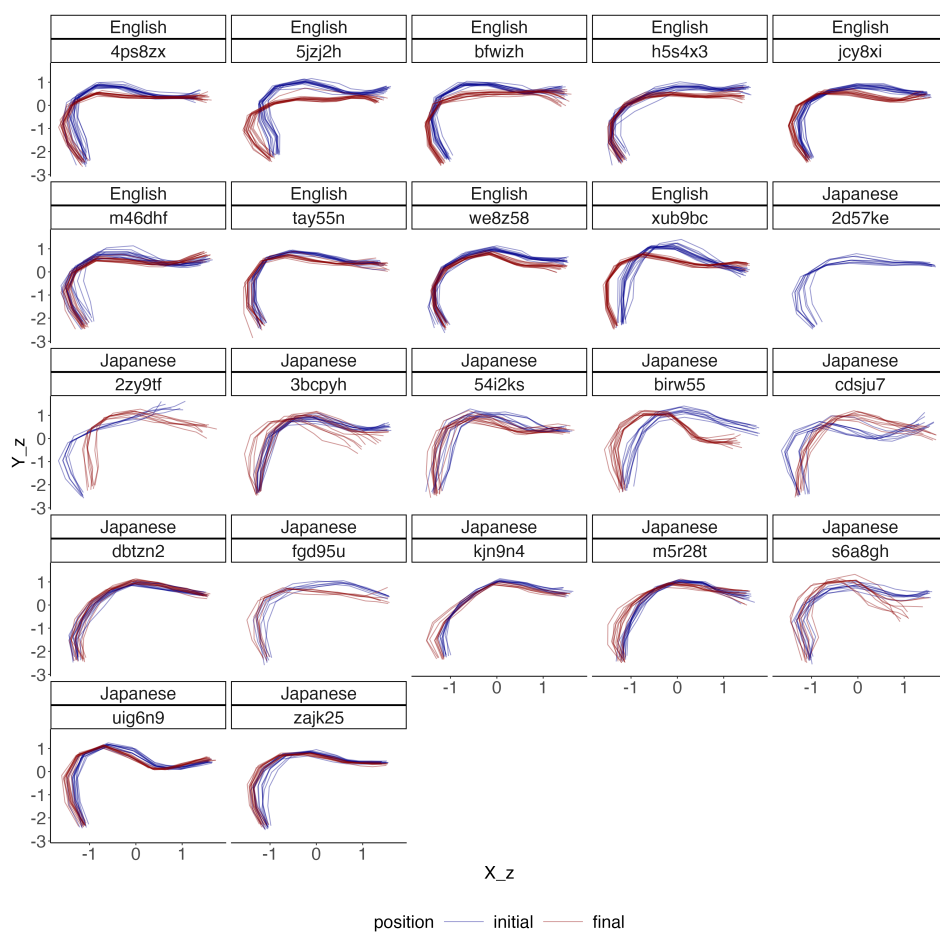


Figure 6: Midsagittal tongue shape extracted at the midpoint of the acoustically-defined liquid interval for English /l/. Tongue tip to the right. Blue splines represent initial tokens and red final tokens. The language label in each facet indicates each speaker's L1 with the anonymised speaker ID underneath. Note that the final tokens of the speaker 2d57ke are excluded from the analysis because they are heavily vocalised and thus the TD/TT displacement was not recorded.

685 A somewhat different pattern emerges for English /ɹ/. It is notable that L1 English speak-
 686 ers employ almost identical tongue shape for both onset and coda rhotics. An exception to
 687 this is two speakers, 4ps8zx and m46dhf, who use two different tongue-tip-up configurations
 688 for initial and final rhotics. Tongue tip difference is seen for two speakers 5jzj2h and we8z58
 689 but these speakers consistently use the front bunched tongue shape across positions; the

690 difference in tongue tip shown here could suggest a slight timing difference as the tongue
 691 shape here is extracted at the liquid midpoint instead of the TT maximal displacement.

692 Tongue shape for L1 Japanese speakers is more variable than that of L1 English speakers
 693 but the pattern is overall similar. Whereas some speakers employ a similar tongue shape
 694 for both onset and coda rhotics (2zy9tf, 3bcpyh, uig6n9), other speakers including cdsju7,
 695 knj9m4 and zajk25, employ different tongue shapes. There is also difference associated with
 696 tongue height and retraction for speakers 54i2ks, dbtzn2 and s6a8gh.

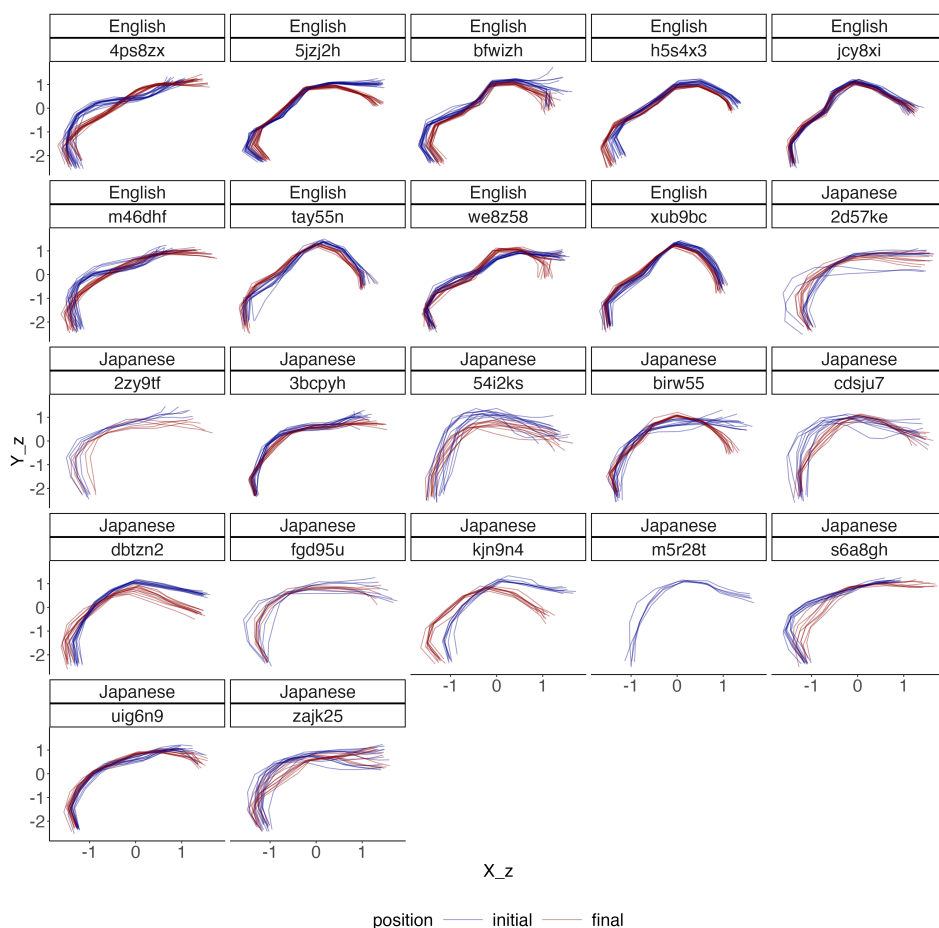


Figure 7: Midsagittal tongue shape extracted at the midpoint of the acoustically-defined liquid interval for English /ɹ/. Tongue tip to the right. Blue splines represent initial tokens and red final tokens. The language label in each facet indicates each speaker's L1 with the anonymised speaker ID underneath. Note that the final tokens of the speaker m5r28t are excluded from the analysis because the tongue shape for /ɹ/ is identical to that of the preceding vowel and thus is impossible to be distinguished.

697 The difference between L1 English and L1 Japanese speakers in tongue shape for /ɹ/ is
698 also reflected in the proportion of tongue shape speakers use. Figure 8 shows the proportion
699 of tongue shape categories that speakers employ to produce English /ɹ/, following King and
700 Ferragne (2020). L1 English speakers predominantly use front bunched (FB) tongue shape
701 for both initial and final /ɹ/s (76.9% for initial and 76.3% for final tokens). Also, there is a
702 greater proportion of the curled up (CU) tongue configuration in the initial position (20.9%)
703 than in the final position (3.7%). Finally, the front up configuration is exclusively used for
704 the final tokens, taking up 17.8% here.

705 In contrast, L1 Japanese speakers mostly employ tongue-tip-up configurations, with the
706 bunched tongue shape constituting only a minor proportion both for the initial (13.7% for
707 front bunched) and for the final tokens (1.1% for front bunched, 15.7% for mid bunched).
708 They are also more likely to use curled up configuration for the initial (61.3%) than for the
709 final rhotics (30.3%).

710 Overall, the qualitative tongue shape analysis highlights that L1 English speakers employ
711 consistent tongue shape strategy to make the onset-coda distinction for laterals and rhotics.
712 Coda /l/s exhibit a lower and more retracted tongue shape for L1 English speakers whereas
713 L1 Japanese speakers distinguish onset and coda laterals less clearly. For English /ɹ/, in
714 contrast, L1 Japanese speakers' tongue shape is more variable than that of L1 English
715 speakers who consistently use the bunched configuration. L1 Japanese speakers are more
716 likely to use curled-up tongue configuration for the initial than for the final /ɹ/s.

3.3 Intergestural timing

717 Finally, the lag between the maximal TT and TB displacement (TB lag) is summarised in
718 Table 4 and visualised in Figure 9. Negative TB lags indicate that the achievement of TT
719 displacement precedes that of TB, whereas positive TB lags vice versa. The TB lag of zero
720 indicate that the maximal displacement is achieved simultaneously between TT and TB.

721 For English /l/, L1 English speakers show a clear onset-coda distinction in TB lag, in

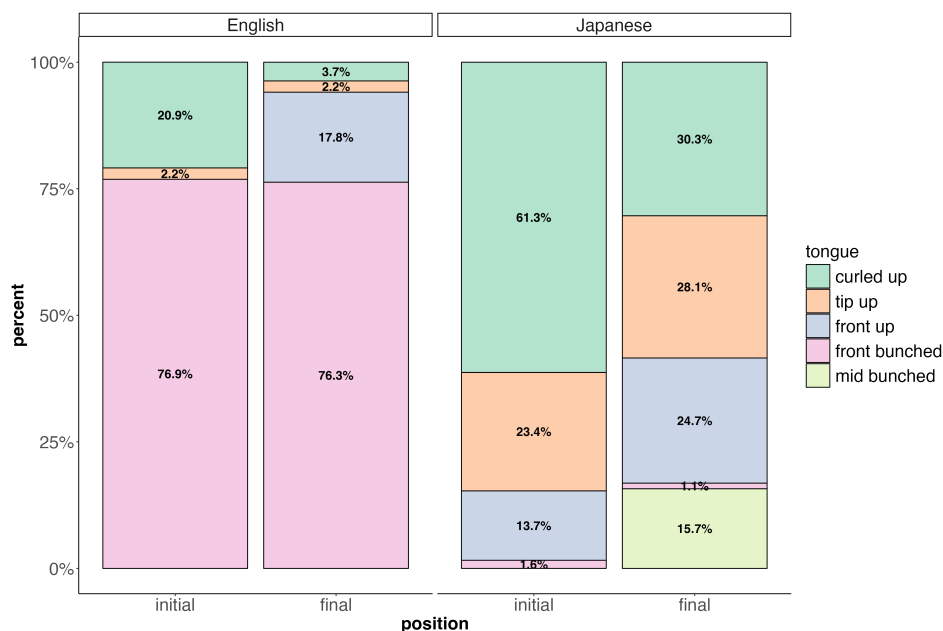


Figure 8: Proportion of each tongue shape category for English /ɪ/.

722 which TT leads TB by 112.47 ms in the initial position, whereas TB precedes TT by 36.60
 723 ms in the final position. L1 Japanese speakers, on the other hand, do not show such a
 724 clear onset-coda difference in TB lag; for both initial and final positions, the TB lag for L1
 725 Japanese speakers is negative (-0.78 ms for initial and -35.58 ms for final), suggesting that
 726 TT tends to precede TB in both positions. As can be seen in Figure 9, however, the time
 727 lag between TT and TB is close to zero, suggesting that L1 Japanese speakers may achieve
 728 the maximal TT and TB displacement almost simultaneously, especially for the initial /l/.

729 In contrast, English /ɪ/ does not exhibit a clear by-group difference in the timing pattern
 730 between TT and TB. The TB lag is close to zero in both positions for L1 English (-13.54
 731 ms for initial and -6.90 ms for final) and L1 Japanese speakers (25.64 ms for initial and
 732 -13.76 ms for final). Despite the TB lag being minimal, however, the overall pattern of TB
 733 lag for English /ɪ/ for each L1 English and L1 Japanese speakers replicates that of English
 734 /l/, in that initial tokens show smaller TB lag than final tokens for L1 English speakers,
 735 whereas vice versa for L1 Japanese speakers.

736 Statistical analysis was conducted using linear mixed-effect modelling separately for En-

Table 4: Mean and SD of the TT-TB lag (in millisecond) for English /l/ and /ɹ/ produced by L1 English and L1 Japanese speakers.

L1	liquid	position	Mean (ms)	SD
English	/l/	initial	-112.47	65.26
		final	36.60	67.18
	/ɹ/	initial	-13.54	56.31
		final	-6.90	49.29
Japanese	/l/	initial	-0.78	102.71
		final	-35.58	87.35
	/ɹ/	initial	25.64	82.60
		final	-13.76	59.71

737 glish /l/ and /ɹ/. The full model for English /l/ predicts TB lag in millisecond by fixed
738 effects of *L1* and *position* with the interaction between them. Random effects include by-
739 speaker varying intercept and by-speaker varying slope for position. Significance testing
740 through model comparison suggests that the interaction between *L1* and *position* improves
741 the degree of model fit at a statistically significant level ($\chi^2(1) = 12.60$, $p < 0.001$), sug-
742 gesting that the magnitude of positional effect on TD lag is different between L1 English
743 and L1 Japanese speakers. A post-hoc pairwise comparison indicates that this is due to L1
744 English speakers contrasting TB lag between the initial and final laterals at a statistically
745 significant level ($\beta = -148.7$, $SE = 34.3$, $t(23.1) = -4.34$, $p < 0.001$) as opposed to L1
746 Japanese speakers who show little evidence of statistically significant difference in TB lag
747 between onset and coda /l/s ($\beta = 34.4$, $SE = 29.3$, $t(24.9) = 1.174$, $p = 0.25$).

748 The specification of the English /ɹ/ models are identical to that of English /l/ model,
749 except that it also considers the fixed effect of *tongue shape*. It does not, however, include
750 the three-way interaction between *L1*, *position* and *tongue shape* due to a colinearity among
751 these leading to a unreliable model estimation. Inspection using the `caret::findLinearCombos`
752 function suggests that this may be due to unique occurrence of mid bunched (MB) and front
753 up (FU) tongue configuration for L1 Japanese speakers. Note that the ‘lateral’ and ‘vowel’
754 tongue shapes were excluded, so this analysis includes five levels for the fixed effect of *tongue*
755 *shape*: curled up (CU; $n = 127$), front up (FU; $n = 61$), tip up (TU; $n = 59$), front bunched

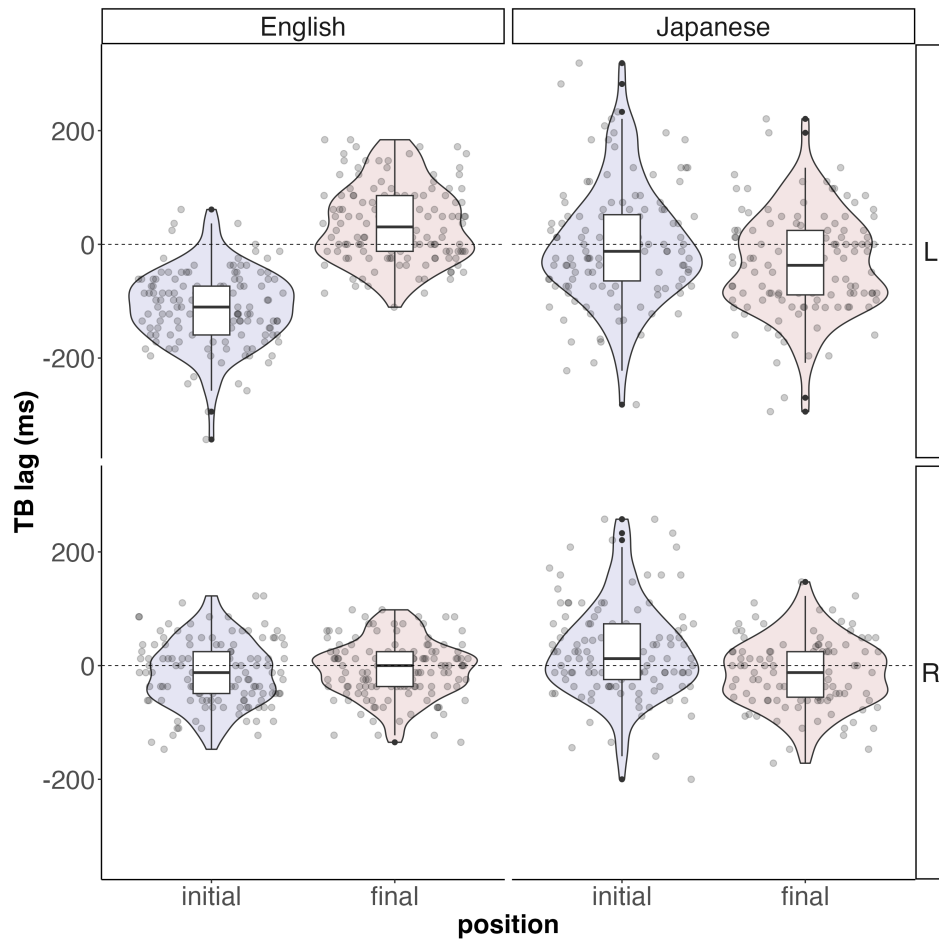


Figure 9: Time lag between TT and TB in ms for laterals (top row) and rhotics (bottom row) produced by L1 English (left column) and L1 Japanese speaker (right column).

756 (FB; $n = 209$) and mid bunched (MB; $n = 14$).

757 Model comparison between the full model and a nested model excluding the interaction
 758 term suggests that it improves the degree of model fit at a statistically significant level ($\chi^2(1)$
 759 $= 5.73$, $p = 0.02$), suggesting that the TB lag pattern could be different as a function of the
 760 speaker's L1. A post-hoc pair-wise comparison shows that this may be due to L1 Japanese
 761 speakers making a contrast in TB lag between onset and coda /ɹ/ ($\beta = 54.90$, $SE = 16.7$,
 762 $t(33.9) = 3.29$, $p = 0.002$) as opposed to L1 English speakers in which the magnitude of
 763 their onset-coda contrast is small ($\beta = -0.85$, $SE = 17.2$, $t(22.9) = -0.05$, $p = 0.96$).

764 Similarly, the full model and a nested model excluding the fixed effect of *tongue shape*

765 suggests that it improves the degree of model fit at a statistically significant level ($\chi^2(5)$
766 = 31.52, $p < 0.001$). A post-hoc pairwise comparison suggests a statistically significant
767 difference between CU and FU ($\beta = -39.72$, $SE = 13.3$, $t(227.7) = -2.98$, $p = 0.03$),
768 between CU and TU ($\beta = -35.64$, $SE = 11.3$, $t(354.0) = -3.15$, $p = 0.02$) and between
769 FB and FU ($\beta = -61.09$, $SE = 20.1$, $t(47.5) = -3.405$, $p = 0.03$). These results, however,
770 must be interpreted with caution due to inevitably a small number of tokens in each tongue
771 shape category.

772 To summarise, intergestural timing analysis indicates that L1 English speakers clearly
773 differentiate onset and coda laterals, with TT leading TB in the onset position and vice versa
774 in the final position. Such an onset-coda distinction is not found for L1 Japanese speak-
775 ers' production of laterals. For English /ɹ/, statistical analysis suggests that L1 Japanese
776 speakers make a contrast between onset and coda rhotics while L1 English speakers do not.
777 Finally, despite being inconclusive, tongue shape might influence the TB lag for English /ɹ/.

4 Discussion

4.1 Summary of the findings

778 My study explores how L1 Japanese-L2 English speakers distinguish the onset-coda allophony
779 in English liquids in their L2 English production. The analysis here suggests evidence that
780 L1 Japanese-L2 English speakers make a contrast between onset and coda /l/s in acoustics.
781 Articulatory strategies, however, are not target-like in that there is little difference in TB
782 lag between onset and coda laterals. In addition, neither speaker groups show evidence of
783 positionally-conditioned allophonic variation for rhotics, although there is a tendency for
784 both groups that the F2–F1 patterns are reversed to that of laterals.

4.2 Positional allophony in L1 and L2 English laterals

785 The analysis for laterals demonstrates that L1 English speakers show a clear contrast be-
786 tween the onset and coda laterals both in acoustics (i.e., higher F2–F1 word-initially than
787 word-finally) and in articulation (i.e., an overall lower/retracted tongue shape and the TB
788 gesture preceding TT). This replicates previous findings of lateral allophony in English with
789 EMA (Ying et al., 2021), X-ray microbeam (Sproat & Fujimura, 1993) and ultrasound (Gick
790 et al., 2006). The clear onset-coda distinction in articulation and in acoustics suggests that
791 L1 English speakers exhibit distinct articulatory targets that characterise the ‘extrinsic’ al-
792 lophonetic variation (Recasens, 2012). This study also adds further evidence of the clear
793 onset-coda allophony in intergestural timing, which can be reliably measured using ultra-
794 sound tongue imaging with a newly-developed DeepLabCut(DLC) implementation of tongue
795 surface estimation (Wrench & Balch-Tomes, 2022).

796 The TB lag pattern observed for L1 Japanese-L2 English speakers suggest that their
797 TT-TB coordination is nearly synchronous word-initially (see Figure 9). Word-finally, L1
798 Japanese-L2 English speakers show a negative TB lag, suggesting that tongue tip precedes
799 tongue dorsum. This overall indicates that they use a single articulatory strategy in both
800 positions, which resemble that of onset, ‘clear’ laterals(Sproat & Fujimura, 1993; Ying et al.,
801 2021). This suggests that that the target-like gestural coordination may be a challenging
802 aspect for L1 Japanese-L2 English speakers in signalling the onset-coda allophony. One
803 possible account for a lack of TT-TB timing relations is the influence of the canonical’
804 realisation of L1 Japanese liquid [r] (Riney et al., 2000). It has been shown that L1 Japanese
805 speakers tend to substitute English liquids with Japanese [r] (Riney et al., 2000), suggesting
806 that the articulatory strategy used for an alveolar tap or flap [r] could influence the way they
807 produce English laterals, and alveolar taps and flaps in Japanese do not involve active TB
808 gesture (Maekawa, 2023; Recasens, 1991). The difference in the intergestural timing pattern
809 for laterals could therefore be explained such that L1 Japanese speakers do not have as much
810 control over the TB gesture as is required for English laterals due to a carry-over effect the

811 L1 timing pattern, resulting in nearly synchronous timing for both initial and final laterals.

812 The individual speakers' data could further illuminate how it is challenging for L1
813 Japanese speakers to fully acquire the target-like intergestural timing pattern for laterals.
814 Looking into the individual variation in the TT-TB timing pattern (see Appendix D), some
815 of the L1 Japanese-L2 English speakers, specifically *kjn9n4* and *m5r28t*, show a reversed
816 pattern from what is expected for L1 English speakers. They show negative lags for the
817 coda position, meaning that their TT precedes TB in the coda position. Auditorily, these
818 speakers substitute /l/ with tap/flap [ɾ], and this may be reflected in their articulation in
819 which no clear TB retractions could be identified around the liquid interval during the anal-
820 ysis. In producing coda laterals, these speakers carry over the tongue body movement from
821 the preceding vowel /i/ without a distinctive TB lowering/retraction during the lateral; a
822 clear TB retraction occurs much later when the tongue goes back to the rest position. The
823 possibility that they use a single articulatory strategy for Japanese /r/ is also evidenced by
824 distinctively positive lags for the initial tokens, in which the tongue body movement precedes
825 tongue tip movement and the tongue body retracts/lowers prior to tongue tip movement,
826 (e.g., Recasens, 1991). These overall suggest that L1 Japanese-L2 English speakers produce
827 laterals with different timing patterns to that of L1 English speakers under the influence of
828 L1 Japanese alveolar taps or flaps [ɾ], especially with regard to the influence of vowels on
829 the TB movement (Maekawa, 2023).

830 The findings that L1 Japanese-L2 English speakers exhibit onset-like intergestural tim-
831 ing patterns for laterals in both onset and coda positions also seem to suggest that their
832 articulatory strategies are influenced by the prosodic position. This can be explained by the
833 difference in syllable structure between Japanese (L1) and English (L2); the prevalence of
834 CV syllable structures and the mora-timing in L1 Japanese often leads to vowel epenthesis
835 that can be attested in loanword adaptation from English to Japanese and in L1 Japanese
836 speakers' production of L2 English (Kubozono, 2015; Li & Juffs, 2014; Vance, 1987). Previ-
837 ous research also demonstrates that L1 Mandarin-L2 English speakers perceive the vocalic

838 component in coda laterals as a separate vowel (e.g., /u/) due to the influence of L1 phono-
839 tactic constraints in which laterals are not permitted in coda position (Wang et al., 2023).
840 Extending this, it is also possible that L1 Japanese speakers' production may be influenced
841 by their existing orthographic knowledge, in which coda laterals could be produced as a
842 sequence of a tap and a back vowel. Previous research shows that L1 Japanese speakers
843 tend to resort to vowel epenthesis as opposed to other modification strategies (e.g., deletion)
844 (Li & Juffs, 2014). Also, in loanwords from English to Japanese, word-final liquids tend to
845 accompany a back vowel /u/ and can be notated with a Japanese mora symbol 'ㇺ'. Overall,
846 the findings for English laterals suggest that, in addition to the influence of Japanese /r/ in
847 the segmental domain, the lack of onset-coda lateral allophony also seems to be influenced
848 by the difference in syllable structure between L1 Japanese and L2 English, articulating
849 the coda laterals similarly to the initial laterals as a result of re-syllabification and with a
850 possible effect of Japanese orthography.

4.3 Positional allophony in L1 and L2 English rhotics

851 For English /ɹ/, I found a lack of positional allophony for both L1 English and L1 Japanese-
852 L2 English speakers' data. There is little evidence to indicate that L1 English speakers
853 distinguish English /ɹ/ between word-initially and word-finally in terms of acoustics and
854 articulation. While this contradicts with findings of some studies claiming a lateral-like
855 onset-coda allophony for English /ɹ/ (Campbell et al., 2010; Proctor et al., 2019), there is
856 also evidence that TT and TB gestures may be achieved nearly simultaneously for English
857 /ɹ/ across different syllabic positions (Gick & Campbell, 2003). Furthermore, even the
858 studies demonstrating a similar TT-TB coordination pattern between laterals and rhotics,
859 the size of the lag is small compare to that of laterals. The TT-TB lag has been shown to
860 be smaller than that of laterals; e.g., initial ca. 10 ms, final below 10 ms (Gick & Campbell,
861 2003); initial 15.32 ms, final 33.89 ms (Campbell et al., 2010) and initial 0 ms, final 78 ms
862 (Proctor et al., 2019), in contrast to this study reporting -13.54 ms word-initially and -6.90

863 ms word-finally (see Table 4). These results agree with the hypothesis posited in Campbell
864 et al. (2010), which were eventually not supported in their study, that both TT and TB
865 gestures involved in articulation of English /ɹ/ could be considered to be vocalic and thus
866 could be timed synchronously both in initial and final positions.

867 It is possible that tongue shape may interact with the way TT and TB are coordinated.
868 A previous study suggests that English /ɹ/ may not show a clear ‘separation’ between the
869 coronal and dorsal gestures unless it is a tongue-tip-up strategy that is similar to laterals
870 (Lawson & Stuart-Smith, 2019). As shown in Figure 8, the majority of the L1 English
871 speakers included in this study employed the ‘front bunched’ configuration with the tongue
872 tip pointing down (cf. King & Ferragne, 2020). Given that previous research has not ad-
873 dressed the issue of the interaction between tongue shape and intergestural timing much,
874 further research is necessary in discussing whether and how English /ɹ/ shows a positional
875 allophony similarly to English /l/, and ultimately, whether the gestural coordination could
876 characterise the phonological class of liquids (Proctor, 2011).

877 Although this study does not find a clear pattern of position-dependent realisations for
878 English /ɹ/, both groups of speakers commonly produce final rhotics with higher F2–F1
879 (i.e., clearer realisations) than initial rhotics, which is an opposite pattern from the lateral
880 acoustics. This might indicate a polarity effect in the liquid system (Carter, 2002); in rhotic
881 varieties (like North American English used in the study), Carter (2002) argues that the
882 liquid darkness show a relative relationship, such that initial laterals tend to be clearer than
883 initial rhotics, and final laterals tend to be darker than final rhotics, which can be seen in
884 Figure 5. It might therefore be the case that rhotic allophony may not just depend on the
885 syllable position that the consonant occurs but could also be best understood in relation
886 to the whole liquid system (Carter, 2002, p. 266). Kirkham (2017) finds a similar polarity
887 effect in Sheffield English and British Asian English. My study could therefore extend the
888 previous claim of the polarity effect to L2 speech production; it is possible that L1 Japanese
889 speakers acquire the liquid system in L2 English, given a similar polarity effect observed

890 with that of L1 English speakers.

4.4 Theoretical implications: Acoustic-articulatory relations and gestural affinity in L2 speech production

891 This study considers the NEW scenario in the Second Language Linguistic Perception (L2LP)
892 model, which presents a substantial difficulty to L2 learners in acquiring the new L2 cate-
893 gories as they need to learn how to make a novel phonological contrast as well as how to
894 phonetically implement each phoneme in a target-like manner (Escudero, 2005). Although
895 previous theoretical frameworks diverge in their postulates on whether L2 speech learning
896 occurs at the phonetic or phonological level, all models agree that L2 learners take phonetic
897 details into account when learning novel L2 segments (Best & Tyler, 2007; Escudero, 2005;
898 Flege, 1995). The current results may contribute to our understanding of L2 speech learn-
899 ing by raising questions as to what phonetic details L2 learners make use of in acquiring
900 target-like production of L2 segments.

901 The biggest question emerging from the current results is: why could L1 Japanese-L2
902 English speakers realise onset-coda lateral allophony in acoustics but not in articulation?
903 It should be noted here that the current data sets do not have a 100% correspondence of
904 tokens included in acoustic and articulatory analysis; acoustic analysis focussed on English
905 liquid tokens that were considered to be realised as an approximant, excluding tokens that
906 were substituted by alveolar taps or flaps [ɾ] or showed unclear formant structures. It is
907 therefore possible that the acoustic analysis for L1 Japanese-L2 English speakers may only
908 contain relatively ‘target-like’ tokens compared to the articulatory analysis, with a total of
909 115 tokens excluded from the acoustic analysis (i.e., 70 tokens among the 105 ‘too short’
910 tokens automatically eliminated by FastTrack and an additional 45 tokens that were not
911 considered approximants). This, in itself, raises a question as to whether all tokens of English
912 liquids produced by L1 Japanese-L2 English speakers could be analysed in a uniform manner;
913 specifically, it needs to be discussed whether a single spectral analysis could be applied to

914 both alveolar taps and liquid approximants, apparently different classes of sounds, which has
915 not been addressed explicitly in the previous research.

916 Despite the methodological considerations, the mismatch between acoustics and artic-
917 ulation seem to be common among studies investigating L2 learners' production of L2 al-
918 lophony (Colantoni et al., 2023; Kochetov, 2022; Nagamine, 2022). In this study, I found a
919 statistically significant difference in F2–F1 but not in the TB lag between word-initial and
920 word-final laterals. The qualitative tongue shape analysis also suggests that L1 Japanese-L2
921 English speakers tend to employ a similar articulatory strategy between the onset and coda
922 laterals. The rhotics result do not demonstrate clear onset-coda allophony, but there is a
923 tendency that initial rhotics are overall darker whereas the final rhotics are overall clearer
924 than the lateral counterparts, suggesting a polarity effect in the phonetic system of English
925 liquids (Carter, 2002).

926 The current results speak for a possibility that L2 learners optimise articulatory strategies
927 in order to achieve target-like acoustic targets (Tourville & Guenther, 2011). The acoustic
928 results are in accordance with the postulations of the Speech Learning Model (SLM) that
929 L2 speech learning occurs based on acoustics at the phonetic, position-dependent allophonic
930 level (Flege, 1995; Flege & Bohn, 2021). L1 Japanese-L2 English speakers in this study make
931 a clear contrast between onset and coda laterals along the F2–F1 dimension, which suggests
932 that they develop separate acoustic categories for the initial and final /l/s. Previous research
933 has shown that, despite an overall difficulty, L1 Japanese speakers learn to make use of the
934 F2 dimension when learning the phonological contrast between English /l/ and /ɹ/ (Iverson
935 et al., 2003; Saito & Munro, 2014). Considering this, it makes sense that L1 Japanese
936 speakers can make adjustments in acoustics to realise the onset-coda lateral allophony along
937 the F2–F1 dimension. Note that much of our understanding of L1 Japanese-L2 English
938 speakers' production of English liquids is based on research focussing on word-initial tokens,
939 and my study here adds new evidence that L1 Japanese speakers can learn English liquids
940 not just word-initially but also word-finally as far as acoustics is concerned (Flege et al.,

941 1995, 2021).

942 In contrast to the acoustic findings, the articulatory results seem to suggest a possibility
943 that L1 Japanese-L2 English speakers classify /l/ as a single category regardless of syllabic
944 position that it occurs, judging from similar patterns in tongue shape and in timing between
945 lingual gestures, supporting an overall claim by PAM-L2 that L2 sounds are assimilated into
946 L1 categories at the phonological level (Best & Tyler, 2007). At the same time, Best and
947 Tyler (2007) also claims that perceptual assimilation of L2 phonemes into L1 phonology
948 occurs when L2 learners identify both L1 and L2 phones sharing the same gestural constella-
949 tion at the phonological level. The current results broadly agree with this claim, except that
950 I did not find a clear evidence that L1 Japanese - L2 learners phonetically distinguish the
951 onset and coda lateral allophones at the articulatory level, suggested by the tongue shape
952 and intergestual timing analysis.

953 The articulatory findings open up a possibility that L2 speech production may not always
954 proceed in the way that is expected from L1 speech production. Previous research claims
955 that the gestural coordination for laterals and rhotics could be understood in a broader
956 framework of gestural affinity to syllables in English (Krakow, 1999; Sproat & Fujimura,
957 1993). My results here suggest that L2 learners of English whose L1 may show different
958 syllable structures may be influenced by their L1 phonotactics to signal gestural affinity in
959 articulation. Previous research demonstrates that L1 Japanese speakers hear a /w/- or /u/-
960 like percept when hearing English liquids in perception (Best & Strange, 1992; Guion et al.,
961 2000) and that they struggle to differentiate word-initial lateral from the following vowel in
962 the /u/ environment (Nagamine, 2024a). In perception, it was shown that L1 Mandarin-
963 speaking listeners recognised the vocalic gesture in the coda /l/s as part of the preceding
964 vowel possibly because the vowel-lateral sequence is not permitted in the phonotactics of
965 L1 Mandarin (Wang et al., 2023). Overall, these indicate a possibility that L1 Japanese-L2
966 English speakers may also recognise the vocalic component in laterals somewhat differently
967 from L1 English speakers due to the influence from the L1 phonotactics, which surface as a

968 different timing pattern in their L2 speech production (Davidson, 2005).

969 Although L1 Japanese-L2 English speakers seem to develop distinct acoustic strategies
970 for onset and coda laterals, in accordance with the predictions made by the SLM, a lack of
971 clear articulatory differences might suggest that these categories are somehow linked at the
972 articulatory level. In other words, they might develop similar articulatory commands in their
973 representations for the two lateral allophones and fine-tune the acoustic output using a wide
974 range of articulatory gestures available to them. Particularly, the possible confusion between
975 /l/ and /u/ for L1 Japanese speakers also opens up a possibility that they might actively use
976 lips to achieve acoustic target for laterals. It is known that lip rounding influences acoustic
977 outputs by lowering the lower three formants (Ladefoged & Ferrari Disner, 2012, p. 179).
978 Although lips are not usually considered an active articulator for laterals, an additional lip
979 rounding could be attested in the case of /l/-vocalisation, where the coda laterals lose tongue
980 tip contact against the alveolar ridge and are therefore produced similarly with back vowels
981 such as [ʊ] or [ɤ] (Hardcastle & Barry, 1989; Strycharczuk et al., 2020; Wells, 1982). Coda
982 laterals share a similar acoustic output with a semi-vowel /w/ (Recasens, 1996), and given
983 the likelihood of confusion for L1 Japanese speakers between English liquids and /w/ or
984 /u/, it is possible that L1 Japanese speakers actively utilise lip movement, instead of lingual
985 articulation, to produce target-like lateral quality and to make a contrast between onset and
986 coda laterals.

987 The data presented here also support this possibility; the intergestural timing analysis
988 suggests that L1 Japanese speakers coordinate both onset and coda laterals with nearly
989 synchronous coordination between TT and TB, a pattern that can be observed for onset
990 laterals (Sproat & Fujimura, 1993; Ying et al., 2021). The tongue shape analysis also
991 demonstrates that L1 Japanese speakers tend to either use a similar tongue shape for both
992 onset and coda laterals or non-target-like differentiation. Furthermore, although not included
993 in the main analysis, L1 Japanese-L2 English speakers in this study produce word-final
994 laterals with lower F3 than L1 English speakers at a statistically significant level, suggesting

995 the possibility of the formant lowering effects of the lips (see Appendix E). Overall, the lateral
996 results suggest that L1 Japanese speakers use a single articulatory strategy for both onset
997 and coda laterals due to cross-linguistic differences in syllable structure and phonotactics
998 between Japanese (L1) and English (L2). The articulatory strategy resembles that of coda
999 laterals in that both TT and TB are timed synchronously. Because of this, they need to
1000 make an adjustment to the acoustic output in order to conform the phonetic realisations of
1001 the lateral allophony, in which one possibility is through the use of lip movement. The lip
1002 data were not analysed here due to its strong focus on lingual articulation, but the possibility
1003 of the labial gesture for laterals will be addressed in future research.

1004 Overall, in light of these theoretical claims and the current results, it could be argued
1005 that L2 learners acquire novel L2 phonetic systems using both phonological and phonetic
1006 information, but they use language specific articulatory strategies to make their production
1007 conform to the target-like acoustic characteristics. My study here provides evidence that L1
1008 Japanese-L2 English speakers acquire target-like allophonic variation of laterals as well as the
1009 English liquid system as a whole given the polarity effect between laterals and rhotics. This
1010 raises a question of functional equivalence in the gestural structure for the onset and coda
1011 liquids (especially laterals) in L2 English production by L1 Japanese speakers (cf. Best &
1012 Tyler, 2007). Nagle and Baese-Berk (2022, p. 16) argues that “L2 speakers may implement
1013 L2 contrasts in nonnativelike ways, in which case they might use phonetic cues that acoustic
1014 analyses of canonical features (i.e., the features that monolingual speakers use to perceive
1015 and produce the target sound) would not detect.” In addition, my study demonstrates that
1016 L2 segmental learning may be more complicated than has been assumed with an addition
1017 of prosodic factors such as differences in syllable structure in this study. It could be argued,
1018 therefore, that L2 speakers may resort to different articulatory strategies to achieve the
1019 target-like acoustic output in L2 speech production in order to overcome differences in both
1020 segmental and prosodic structures. This suggests that they may develop distinct phonetic
1021 categories at the level of position-dependent allophones in acoustics, the categories may still

1022 maintain some link at the articulatory level.

5 Conclusion

1023 This study explores the way L2 speakers realise allophonic variation in their L2 speech
1024 production. Thirteen L1 Japanese-L2 English speakers' production of English liquids /l/ and
1025 /ɹ/ in the initial and final position embedded in monosyllabic words are compared against L1
1026 North American English speakers' production. The results demonstrate that L1 Japanese
1027 speakers clearly implement the lateral allophony in acoustics in a target-like manner but
1028 little evidence is found in articulation (tongue shape and intergestural timing). This could
1029 be due to a combination of factors, including the possible influence of L1 articulation and L1
1030 syllabic structure. Also, little evidence of positional allophony is found for English /ɹ/ but
1031 there was an indication of the polarity effect, highlighting the importance of understanding
1032 the liquid phonetic system as a whole.

1033 Overall, this study provides more questions than answers, especially with regard to the
1034 nature of L2 speech learning. My results seem to suggest that L2 learners indeed produce
1035 position-dependent allophonic variation in acoustics, the phonetic categories might still be
1036 linked at the fundamental, articulatory level. While the mismatch in acoustics and artic-
1037 ulation may just be a consequence of between-speaker variability, it is also possible that
1038 L2 learners fundamentally differ from L1 speakers in the way they implement L2 segments,
1039 especially in articulation in which L2 learners may resort to different articulatory strategies
1040 in order to overcome L1-L2 differences in segments and prosody that may not always be
1041 transparent from the acoustic signals. Overall, L2 speech production may be more compli-
1042 cated than just an L1-L2 mapping between available sounds in both languages, and future
1043 research should therefore look into phonetic dimensions that may not always be conventional
1044 in L1 speech production.

Acknowledgements

1045 This is part of the author's PhD research at Lancaster University in the UK. I thank Professor
1046 Claire Nance and Dr Sam Kirkham for their comments on earlier drafts and overall support
1047 and encouragement. The data collection was made possible thanks to great help by Professor
1048 Noriko Nakanishi (Kobe Gakuin University, Japan), Professor Yuri Nishio (Meijo University,
1049 Japan) and Dr Bronwen Evans (University College London, UK). This work is financially
1050 supported by the Graduate Scholarship for Degree-Seeking Students by the Japan Student
1051 Services Organization (JASSO) [ID: 20SD10500601] and the 2022 Research Grant by the
1052 Murata Science Foundation [grant number: M22助人027] awarded to the author. Codes and
1053 data supporting the findings in this article are publicly available at <https://osf.io/5sx7t/>.
1054 The author has no conflicts to declare. This research is approved by ethics committees at
1055 Lancaster University, Kobe Gakuin University, and Meijo University. Informed consent was
1056 obtained from all participants.

References

- 1057 Alwan, A., Narayanan, S., & Haker, K. (1997). Toward articulatory-acoustic models for liquid
1058 approximants based on MRI and EPG data. Part II. The rhotics. *The Journal of the*
1059 *Acoustical Society of America*, *101*(2), 1078–1089. <https://doi.org/10.1121/1.417972>
- 1060 Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2019). An acoustic analysis
1061 of American English liquids by adults and children: Native English speakers and
1062 native Japanese speakers of English. *The Journal of the Acoustical Society of America*,
1063 *146*(4), 2671–2681. <https://doi.org/10.1121/1.5130574>
- 1064 Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Per-
1065 ceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and
1066 English /l/ and /r/. *Journal of Phonetics*, *32*(2), 233–250. [https://doi.org/10.1016/](https://doi.org/10.1016/S0095-4470(03)00036-6)
1067 [S0095-4470\(03\)00036-6](https://doi.org/10.1016/S0095-4470(03)00036-6)

- 1068 Arai, T. (2013). On Why Japanese /r/ Sounds are Difficult for Children to Acquire. *Inter-*
1069 *speech 2013*, 2445–2449.
- 1070 Articulate Instruments. (2022). Articulate Assistant Advanced version 220.
- 1071 Barlow, J. A., Branson, P. E., & Nip, I. S. B. (2013). Phonetic equivalence in the acquisition
1072 of /l/ by Spanish–English bilingual children*. *Bilingualism: Language and Cognition*,
1073 *16*(1), 68–85. <https://doi.org/10.1017/S1366728912000235>
- 1074 Barreda, S. (2021). Fast Track: Fast (nearly) automatic formant-tracking using Praat. *Lin-*
1075 *guistics Vanguard*, *7*(1). <https://doi.org/10.1515/lingvan-2020-0051>
- 1076 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models
1077 Using lme4. *Journal of Statistical Software*, *67*, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- 1078
- 1079 Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-
1080 language perception of approximants. *Journal of Phonetics*, *20*(3), 305–330. [https://doi.org/10.1016/S0095-4470\(19\)30637-0](https://doi.org/10.1016/S0095-4470(19)30637-0)
- 1081
- 1082 Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Com-
1083 monalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language*
1084 *experience in second language speech learning: In honor of James Emil Flege* (pp. 13–
1085 34). John Benjamins Publishing Company. <https://doi.org/10.1075/lllt.17.07bes>
- 1086 Boersma, P., & Weenink, D. (2022, May). Praat: Doing Phonetics by Computer.
- 1087 Borchers, H. W. (2023, November). Pracma: Practical Numerical Math Functions.
- 1088 Campbell, F., Gick, B., Wilson, I., & Vatikiotis-Bateson, E. (2010). Spatial and Temporal
1089 Properties of Gestures in North American English /r/. *Language and Speech*, *53*(1),
1090 49–69. <https://doi.org/10.1177/0023830909351209>
- 1091 Carter, P. (2002). *Structured variation in British English liquids : The role of resonance*
1092 [Doctoral dissertation, University of York].

- 1093 Carter, P., & Local, J. (2007). F2 variation in Newcastle and Leeds English liquid systems.
1094 *Journal of the International Phonetic Association*, 37(2), 183–199. [https://doi.org/](https://doi.org/10.1017/S0025100307002939)
1095 10.1017/S0025100307002939
- 1096 Chang, C. B. (2019, March). The phonetics of second language learning and bilingualism. In
1097 W. F. Katz & P. F. Assmann (Eds.), *The Routledge Handbook of Phonetics* (1st ed.,
1098 pp. 427–447). Routledge. <https://doi.org/10.4324/9780429056253-16>
- 1099 Chen, S., Whalen, D. H., & Mok, P. P. K. (2024). Production of the English /ɹ/ by Mandarin–
1100 English Bilingual Speakers. *Language and Speech*, 00238309241230895. [https://doi.](https://doi.org/10.1177/00238309241230895)
1101 [org/10.1177/00238309241230895](https://doi.org/10.1177/00238309241230895)
- 1102 Chung, H., & Kim, Y. (2021). Acoustic characteristics of Korean-English bilingual speakers’
1103 /l/ and the relationship to their foreign accent ratings. *Journal of Communication*
1104 *Disorders*, 94, 106157. <https://doi.org/10.1016/j.jcomdis.2021.106157>
- 1105 Colantoni, L., Kochetov, A., & Steele, J. (2023). Articulatory Insights into the L2 Acquisition
1106 of English-/l/ Allophony. *Language and Speech*, 00238309231200629. [https://doi.org/](https://doi.org/10.1177/00238309231200629)
1107 10.1177/00238309231200629
- 1108 Colantoni, L., & Steele, J. (2008). Integrating articulatory constraints into models of second
1109 language phonological acquisition. *Applied Psycholinguistics*, 29(3), 489–534. [https://](https://doi.org/10.1017/S0142716408080223)
1110 doi.org/10.1017/S0142716408080223
- 1111 Davidson, L. (2005). Addressing phonological questions with ultrasound. *Clinical Linguistics*
1112 *& Phonetics*, 19(6-7), 619–633. <https://doi.org/10.1080/02699200500114077>
- 1113 Davidson, L. (2006). Phonotactics and Articulatory Coordination Interact in Phonology:
1114 Evidence from Nonnative Production. *Cognitive Science*, 30(5), 837–862. [https://](https://doi.org/10.1207/s15516709cog0000_73)
1115 doi.org/10.1207/s15516709cog0000_73
- 1116 Delattre, P., & Freeman, D. C. (1968). A Dialect Study of American R’s by X-ray motion
1117 picture. *Linguistics*, 6(44), 29–68. <https://doi.org/10.1515/ling.1968.6.44.29>

- 1118 Escudero, P. (2001). The perception of English vowel contrasts: Acoustic cue reliance in the
1119 development of new contrasts. *Proceedings of the 4th International Symposium on the*
1120 *Acquisition of Second-Language Speech, New Sounds*.
- 1121 Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the*
1122 *attainment of optimal phonological categorization* [Doctoral dissertation, LOT].
- 1123 Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop
1124 consonants produced in a second language. *The Journal of the Acoustical Society of*
1125 *America*, 89(1), 395–411. <https://doi.org/10.1121/1.400473>
- 1126 Flege, J. E. (1995). Second Language Speech Learning Theory, Findings and Problems. In W.
1127 Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language*
1128 *Research* (pp. 233–277). York Press.
- 1129 Flege, J. E., Aoyama, K., & Bohn, O.-S. (2021, February). The Revised Speech Learn-
1130 ing Model (SLM-r) Applied. In R. Wayland (Ed.), *Second Language Speech Learn-*
1131 *ing* (1st ed., pp. 84–118). Cambridge University Press. [https://doi.org/10.1017/](https://doi.org/10.1017/9781108886901.003)
1132 [9781108886901.003](https://doi.org/10.1017/9781108886901.003)
- 1133 Flege, J. E., & Bohn, O.-S. (2021, February). The Revised Speech Learning Model (SLM-r).
1134 In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical*
1135 *Progress* (1st ed., pp. 3–83). Cambridge University Press. [https://doi.org/10.1017/](https://doi.org/10.1017/9781108886901.002)
1136 [9781108886901.002](https://doi.org/10.1017/9781108886901.002)
- 1137 Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese Adults can Learn to Produce English
1138 /I/ and /l/ Accurately. *Language and Speech*, 38(1), 25–55. [https://doi.org/10.1177/](https://doi.org/10.1177/002383099503800102)
1139 [002383099503800102](https://doi.org/10.1177/002383099503800102)
- 1140 Gick, B. (1999). A gesture-based account of intrusive consonants in English. *Phonology*,
1141 16(1), 29–54. <https://doi.org/10.1017/S0952675799003693>
- 1142 Gick, B., & Campbell, F. (2003). Intergestural Timing in English /r/. In M.-J. Solé, D.
1143 Recasens, & J. G. Romero (Eds.), *Proceedings of the XVth international congress of*
1144 *phonetic sciences* (pp. 1911–1914).

- 1145 Gick, B., Campbell, F., Oh, S., & Tamburri-Watt, L. (2006). Toward universals in the gestural
1146 organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*,
1147 *34*(1), 49–72. <https://doi.org/10.1016/j.wocn.2005.03.005>
- 1148 Grosjean, F. (2008). The bilingual's language mode. In *Studying bilinguals* (pp. 36–66).
1149 Oxford University Press.
- 1150 Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of
1151 current models of second language speech perception: The case of Japanese adults'
1152 perception of English consonants. *The Journal of the Acoustical Society of America*,
1153 *107*(5), 2711–2724. <https://doi.org/10.1121/1.428657>
- 1154 Hardcastle, W., & Barry, W. (1989). Articulatory and perceptual factors in /l/ vocalisations
1155 in English. *Journal of the International Phonetic Association*, *15*(2), 3–17. <https://doi.org/10.1017/S0025100300002930>
- 1157 Harper, S., Goldstein, L., & Narayanan, S. (2020). Variability in individual constriction
1158 contributions to third formant values in American English /ɹ/. *The Journal of the*
1159 *Acoustical Society of America*, *147*(6), 3905–3916. [https://doi.org/10.1121/10.](https://doi.org/10.1121/10.0001413)
1160 [0001413](https://doi.org/10.1121/10.0001413)
- 1161 Howson, P. J., & Redford, M. A. (2021). The Acquisition of Articulatory Timing for Liquids:
1162 Evidence From Child and Adult Speech. *Journal of Speech, Language, and Hearing*
1163 *Research*, *64*(3), 734–753. https://doi.org/10.1044/2020_JSLHR-20-00391
- 1164 Hwang, Y., Lulich, S. M., & de Jong, K. J. (2019). Articulatory and acoustic characteristics
1165 of the Korean and English word-final laterals produced by Korean female learners of
1166 American English. *The Journal of the Acoustical Society of America*, *146*(5), EL444–
1167 EL450. <https://doi.org/10.1121/1.5134656>
- 1168 Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manip-
1169 ulations: A comparison of methods for teaching English /r-/l/ to Japanese adults.
1170 *The Journal of the Acoustical Society of America*, *118*(5), 3267–3278. [https://doi.](https://doi.org/10.1121/1.2062307)
1171 [org/10.1121/1.2062307](https://doi.org/10.1121/1.2062307)

- 1172 Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A.,
1173 & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for
1174 non-native phonemes. *Cognition*, *87*(1), B47–B57. [https://doi.org/10.1016/S0010-](https://doi.org/10.1016/S0010-0277(02)00198-1)
1175 [0277\(02\)00198-1](https://doi.org/10.1016/S0010-0277(02)00198-1)
- 1176 Kawahara, S., & Matsui, M. F. (2017). Some aspects of Japanese consonant articulation: A
1177 preliminary EPG study. *ICU Working Papers in Linguistics (ICUWPL)*, *2*, 9–20.
- 1178 King, H., & Ferragne, E. (2020). Loose lips and tongue tips: The central role of the /r/-
1179 typical labial gesture in Anglo-English. *Journal of Phonetics*, *80*, 100978. [https://](https://doi.org/10.1016/j.wocn.2020.100978)
1180 doi.org/10.1016/j.wocn.2020.100978
- 1181 Kirkham, S. (2017). Ethnicity and phonetic variation in Sheffield English liquids. *Journal*
1182 *of the International Phonetic Association*, *47*(1), 17–35. [https://doi.org/10.1017/](https://doi.org/10.1017/S0025100316000268)
1183 [S0025100316000268](https://doi.org/10.1017/S0025100316000268)
- 1184 Kirkham, S. (2024). TadaR: R interface to Task Dynamic Application (v 1.0.0). [https://doi.](https://doi.org/10.5281/zenodo.13329512)
1185 [org/10.5281/zenodo.13329512](https://doi.org/10.5281/zenodo.13329512)
- 1186 Kirkham, S., & McCarthy, K. M. (2021). Acquiring allophonic structure and phonetic detail
1187 in a bilingual community: The production of laterals by Sylheti-English bilingual
1188 children. *International Journal of Bilingualism*, *25*(3), 531–547. [https://doi.org/10.](https://doi.org/10.1177/1367006920947180)
1189 [1177/1367006920947180](https://doi.org/10.1177/1367006920947180)
- 1190 Kirkham, S., Turton, D., & Leemann, A. (2020). A typology of laterals in twelve English
1191 dialects. *The Journal of the Acoustical Society of America*, *148*(1), EL72–EL76. [https:](https://doi.org/10.1121/10.0001587)
1192 [//doi.org/10.1121/10.0001587](https://doi.org/10.1121/10.0001587)
- 1193 Kochetov, A. (2022). Production of English phonemic contrasts and allophony by Japanese
1194 learners: Electropalatographic evidence. *Phonology Festa 17*.
- 1195 Krakow, R. A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics*,
1196 *27*(1), 23–54. <https://doi.org/10.1006/jpho.1999.0089>

- 1197 Kubozono, H. (2015, December). Loanword phonology. In H. Kubozono (Ed.), *Handbook of*
1198 *Japanese Phonetics and Phonology* (pp. 313–362). DE GRUYTER. [https://doi.org/](https://doi.org/10.1515/9781614511984.313)
1199 [10.1515/9781614511984.313](https://doi.org/10.1515/9781614511984.313)
- 1200 Ladefoged, P., & Ferrari Disner, S. (2012). *Vowels and Consonants* (3rd ed.). Wiley-Blackwell.
- 1201 Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell Pub-
1202 lishers.
- 1203 Lawson, E., & Stuart-Smith, J. (2019). The effects of syllable and sentential position on
1204 the timing of lingual gestures in /l/ and /r/. *Proceedings of the 19th International*
1205 *Congress of Phonetic Sciences*, 547–551.
- 1206 Lawson, E., Stuart-Smith, J., & Rodger, L. (2019). A comparison of acoustic and articulatory
1207 parameters for the GOOSE vowel across British Isles Englishes. *The Journal of the*
1208 *Acoustical Society of America*, 146(6), 4363–4381. <https://doi.org/10.1121/1.5139215>
- 1209 Lee-Kim, S.-I., Davidson, L., & Hwang, S. (2013). Morphological effects on the darkness of
1210 English intervocalic /l/. *Laboratory Phonology*, 4(2). [https://doi.org/10.1515/lp-](https://doi.org/10.1515/lp-2013-0015)
1211 [2013-0015](https://doi.org/10.1515/lp-2013-0015)
- 1212 Léger, A., King, H., & Ferragne, E. (2023). Is rhoticity on the tip of your tongue? Tongue
1213 shapes for English /r/ in French learners with ultrasound. In Radek Skarnitzl & Jan
1214 Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*
1215 (pp. 2741–2745). Guarant International.
- 1216 Li, N. H., & Juffs, A. (2014). The Influence of Moraic Structure on L2 English Syllable-Final
1217 Consonants. *Proceedings of the Annual Meetings on Phonology*. [https://doi.org/10.](https://doi.org/10.3765/amp.v2i0.3767)
1218 [3765/amp.v2i0.3767](https://doi.org/10.3765/amp.v2i0.3767)
- 1219 Lindau, M. (1985). The story of /r/. In V. A. Fromkin (Ed.), *Phonetic Linguistics: Essays*
1220 *in Honor of Peter Ladefoged* (pp. 157–68). Academic Press.
- 1221 Llompart, M., Eger, N. A., & Reinisch, E. (2021). Free Allophonic Variation in Native and
1222 Second Language Spoken Word Recognition: The Case of the German Rhotic. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.711230>
1223

- 1224 Maekawa, K. (2023). Articulatory characteristics of the Japanese /r/: A real-time MRI study.
1225 In Radek Skarnitzl & Jan Volín (Eds.), *Proceedings of the 20th International Congress*
1226 *of Phonetic Sciences* (pp. 992–996). Guarant International.
- 1227 Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge,
1228 M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with
1229 deep learning. *Nature Neuroscience*, *21*(9), 1281–1289. [https://doi.org/10.1038/
1230 s41593-018-0209-y](https://doi.org/10.1038/s41593-018-0209-y)
- 1231 McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal
1232 Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech 2017*, 498–
1233 502. <https://doi.org/10.21437/Interspeech.2017-1386>
- 1234 Mielke, J., Baker, A., & Archangeli, D. (2016). Individual-level contact limits phonologi-
1235 cal complexity: Evidence from bunched and retroflex /ɾ/. *Language*, *92*(1), 101–140.
1236 <https://doi.org/10.1353/lan.2016.0019>
- 1237 Moore, J., Shaw, J., Kawahara, S., & Arai, T. (2018). Articulation strategies for English
1238 liquids used by Japanese speakers. *Acoustical Science and Technology*, *39*(2), 75–83.
1239 <https://doi.org/10.1250/ast.39.75>
- 1240 Morimoto, M. (2020, March). *Geminated Liquids in Japanese: A Production Study* [Doctoral
1241 dissertation, University of California Santa Cruz].
- 1242 Nagamine, T. (2022). Acquisition of allophonic variation in second language speech: An
1243 acoustic and articulatory study of English laterals by Japanese speakers. *Interspeech*
1244 *2022*, 644–648. <https://doi.org/10.21437/Interspeech.2022-11020>
- 1245 Nagamine, T. (2024a). Formant dynamics in second language speech: Japanese speakers’ pro-
1246 duction of English liquids. *The Journal of the Acoustical Society of America*, *155*(1),
1247 479–495. <https://doi.org/10.1121/10.0024351>
- 1248 Nagamine, T. (2024b). Acquisition of articulatory dynamics in second language speech:
1249 Japanese speakers’ production of English and Japanese liquids. *The 13th Interna-*
1250 *tional Seminar of Speech Production*.

- 1251 Nagle, C. L., & Baese-Berk, M. M. (2022). Advancing the state of the art in L2 speech
1252 perception-production research: Revisiting theoretical assumptions and methodolog-
1253 ical practices. *Studies in Second Language Acquisition*, 44(2), 1–26. [https://doi.org/
1254 10.1017/S0272263121000371](https://doi.org/10.1017/S0272263121000371)
- 1255 Nance, C. (2014). Phonetic variation in Scottish Gaelic laterals. *Journal of Phonetics*, 47,
1256 1–17. <https://doi.org/10.1016/j.wocn.2014.07.005>
- 1257 Nance, C., & Kirkham, S. (2023). Producing a smaller sound system: Acoustics and articu-
1258 lation of the subset scenario in Gaelic–English bilinguals. *Bilingualism: Language and
1259 Cognition*, 1–13. <https://doi.org/10.1017/S1366728923000688>
- 1260 Narayanan, S. S., Alwan, A. A., & Haker, K. (1997). Toward articulatory-acoustic models for
1261 liquid approximants based on MRI and EPG data. Part I. The laterals. *The Journal
1262 of the Acoustical Society of America*, 101(2), 1064–1077. [https://doi.org/10.1121/1.
1263 418030](https://doi.org/10.1121/1.418030)
- 1264 Olsen, M. K. (2012). The L2 Acquisition of Spanish Rhotics by L1 English Speakers: The
1265 Effect of L1 Articulatory Routines and Phonetic Context for Allophonic Variation.
1266 *Hispania*, 95(1), 65–82.
- 1267 Otake, T. (2015, December). Mora and mora-timing. In H. Kubozono (Ed.), *Handbook of
1268 Japanese Phonetics and Phonology* (pp. 493–524). DE GRUYTER. [https://doi.org/
1269 10.1515/9781614511984.493](https://doi.org/10.1515/9781614511984.493)
- 1270 Proctor, M. (2011). Towards a gestural characterization of liquids: Evidence from Spanish
1271 and Russian. *Laboratory Phonology*, 2(2), 451–485. [https://doi.org/10.1515/labphon.
1272 2011.017](https://doi.org/10.1515/labphon.2011.017)
- 1273 Proctor, M., Walker, R., Smith, C., Szalay, T., Goldstein, L., & Narayanan, S. (2019).
1274 Articulatory characterization of English liquid-final rimes. *Journal of Phonetics*, 77,
1275 100921. <https://doi.org/10.1016/j.wocn.2019.100921>
- 1276 Recasens, D. (1991). On the production characteristics of apicoalveolar taps and trills. *Jour-
1277 nal of Phonetics*, 19(3-4), 267–280. [https://doi.org/10.1016/S0095-4470\(19\)30344-4](https://doi.org/10.1016/S0095-4470(19)30344-4)

- 1278 Recasens, D. (1996). An Articulatory-Perceptual Account of Vocalization and Elision of
1279 Dark /l/ in the Romance Languages. *Language and Speech*, 39(1), 63–89. <https://doi.org/10.1177/002383099603900104>
1280
- 1281 Recasens, D. (2012). A cross-language acoustic study of initial and final allophones of /l/.
1282 *Speech Communication*, 54(3), 368–383. [https://doi.org/10.1016/j.specom.2011.10.](https://doi.org/10.1016/j.specom.2011.10.001)
1283 001
- 1284 Riney, T. J., Takada, M., & Ota, M. (2000). Segmentals and Global Foreign Accent: The
1285 Japanese Flap in EFL. *TESOL Quarterly*, 34(4), 711–737. [https://doi.org/10.2307/](https://doi.org/10.2307/3587782)
1286 3587782
- 1287 Saito, K., & Munro, M. J. (2014). The Early Phase of /ɹ/ Production Development in
1288 Adult Japanese Learners of English. *Language and Speech*, 57(4), 451–469. <https://doi.org/10.1177/0023830913513206>
1289
- 1290 Scobbie, J., Lawson, E., Cowen, S., Cleland, J., & Wrench, A. (2011). A common co-ordinate
1291 system for mid-sagittal articulatory measurement. *QMU CASL Working Papers*, 20,
1292 1–4.
- 1293 Solon, M. (2017). Do learners lighten up? Phonetic and Allophonic Acquisition of Spanish
1294 /l/ by English-Speaking Learners. *Studies in Second Language Acquisition*, 39(4),
1295 801–832. <https://doi.org/10.1017/S0272263116000279>
- 1296 Spreafico, L., Pucher, M., & Matosova, A. (2018). UltraFit: A Speaker-friendly Headset
1297 for Ultrasound Recordings in Speech Science. *Interspeech 2018*, 1517–1520. <https://doi.org/10.21437/Interspeech.2018-995>
1298
- 1299 Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications
1300 for phonetic implementation. *Journal of Phonetics*, 21(3), 291–311. [https://doi.org/](https://doi.org/10.1016/S0095-4470(19)31340-3)
1301 10.1016/S0095-4470(19)31340-3
- 1302 Strycharczuk, P., Derrick, D., & Shaw, J. (2020). Locating de-lateralization in the pathway
1303 of sound changes affecting coda /l/. *Laboratory Phonology: Journal of the Association*
1304 *for Laboratory Phonology*, 11(1), 21. <https://doi.org/10.5334/labphon.236>

- 1305 Tiede, M. K., Boyce, S. E., Holland, C. K., & Choe, K. A. (2004). A new taxonomy of Amer-
1306 ican English /r/ using MRI and ultrasound. *The Journal of the Acoustical Society of*
1307 *America*, 115(5), 2633–2634. <https://doi.org/10.1121/1.4784878>
- 1308 Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech
1309 acquisition and production. *Language and Cognitive Processes*, 26(7), 952–981. <https://doi.org/10.1080/01690960903498424>
- 1310
1311 Tsui, H. M.-L. (2012). *Ultrasound speech training for Japanese adults learning English as a*
1312 *second language* [Doctoral dissertation, University of British Columbia]. [https://doi.](https://doi.org/10.14288/1.0073242)
1313 [org/10.14288/1.0073242](https://doi.org/10.14288/1.0073242)
- 1314 van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second
1315 language: The L2LP model revised. *Frontiers in Psychology*, 6, Article1000. <https://doi.org/10.3389/fpsyg.2015.01000>
- 1316
1317 Vance, T. J. (1987). *An introduction to Japanese phonology*. State University of New York
1318 Press.
- 1319 Vance, T. J. (2008). *The Sounds of Japanese*. Cambridge University Press.
- 1320 Wang, Y., Bundgaard-Nielsen, R. L., Baker, B. J., & Maxwell, O. (2023). Difficulties in
1321 decoupling articulatory gestures in L2 phonemic sequences: The case of Mandarin
1322 listeners' perceptual deletion of English post-vocalic laterals. *Phonetica*, 80(1-2), 79–
1323 115. <https://doi.org/10.1515/phon-2022-0027>
- 1324 Wells, J. C. (1982, April). Accents of English. <https://doi.org/10.1017/CBO9780511611759>
- 1325 Wiese, R. (2011, April). The Representation of Rhotics: The Representation of Rhotics. In
1326 M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Com-*
1327 *panion to Phonology* (pp. 1–19). John Wiley & Sons, Ltd. [https://doi.org/10.1002/](https://doi.org/10.1002/9781444335262.wbctp0030)
1328 [9781444335262.wbctp0030](https://doi.org/10.1002/9781444335262.wbctp0030)
- 1329 Winter, B. (2020). *Statistics for Linguists: An Introduction Using R*. Routledge.

- 1330 Wrench, A., & Balch-Tomes, J. (2022). Beyond the Edge: Markerless Pose Estimation of
1331 Speech Articulators from Ultrasound and Camera Images Using DeepLabCut. *Sen-*
1332 *sors*, 22(3), 1133. <https://doi.org/10.3390/s22031133>
- 1333 Yazawa, K., Whang, J., Kondo, M., & Escudero, P. (2020). Language-dependent cue weight-
1334 ing: An investigation of perception modes in L2 learning. *Second Language Research*,
1335 36(4), 557–581. <https://doi.org/10.1177/0267658319832645>
- 1336 Ying, J., Shaw, J. A., Carignan, C., Proctor, M., Derrick, D., & Best, C. T. (2021). Evi-
1337 dence for active control of tongue lateralization in Australian English /l/. *Journal of*
1338 *Phonetics*, 86, 101039. <https://doi.org/10.1016/j.wocn.2021.101039>
- 1339 Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A
1340 magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and
1341 “bunched” American English /r/. *The Journal of the Acoustical Society of America*,
1342 123(6), 4466–4481. <https://doi.org/10.1121/1.2902168>

Appendix A Statistical results

A.1 Acoustic analysis F2–F1

1343 The details of the full model and model comparisons for the acoustic analysis is shown below
 1344 in Table A.5. All fixed effects are treatment coded, such that the baseline level for *L1* is L1
 1345 English speakers, for *Position* the initial position, and for *Liquid* English /l/. The outcome
 1346 variable is within-speaker *z*-normalised F2–F1.

Table A.5: Summary of the linear mixed-effect modelling for F2–F1 (within-speaker *z*-normalised).

Full model					
Variable	β	<i>SE</i>	<i>df</i>	<i>t</i>	$p(\chi^2)$
Intercept	−0.29	0.29	19.34	−0.99	
L1					
Japanese	0.74	0.32	13.82	2.31	
Position					
Final	−0.83	0.32	23.87	−2.61	
Liquid					
/r/	0.12	0.35	27.60	0.34	
L1:Position					0.19
Japanese:Final	−0.42	0.31	18.38	−1.34	
Position:Liquid					< 0.001
Final:/r/	1.52	0.27	11.50	5.56	
L1:Liquid					0.04
Japanese:/r/	−0.86	0.36	20.22	−2.41	
Post-hoc pairwise comparison					
Contrast	β	<i>SE</i>	<i>df</i>	<i>t</i>	$p(t)$
Position:Liquid, Liquid = /l/					
Initial - final	1.11	0.28	25.0	3.98	< 0.001
Position:Liquid, Liquid = /r/					
Initial - final	−0.41	0.29	23.8	−1.42	0.17
L1:Liquid, Liquid = /l/					
English - Japanese	−0.45	0.26	32.1	−1.75	0.09
L1:Liquid, Liquid = /r/					
English - Japanese	0.35	0.19	26.6	1.81	0.08

A.2 Intergestural timing

1347 The details of the full models and the post-hoc pairwise comparisons for the TB lag analysis
 1348 are shown below in Table A.6 for laterals and Table A.7 for rhotics. The outcome variable
 1349 is TB lag in millisecond. All fixed effects are treatment coded, such that the baseline for *L1*
 1350 is L1 English speakers and for *position* the initial position.

Table A.6: Summary of the linear mixed-effect modelling for laterals /l/ for TB lag (ms).

Full model					
Variable	β	<i>SE</i>	<i>df</i>	<i>t</i>	$p(\chi^2)$
Intercept	-112.46	20.05	20.90	-5.61	
L1					
Japanese	111.49	26.29	21.54	4.24	
Position					
Final	148.70	32.61	18.59	4.56	
Interaction					< 0.001
Japanese:Final	-183.12	42.91	19.29	-4.27	
Post-hoc pairwise comparison					
Contrast	β	<i>SE</i>	<i>df</i>	<i>t</i>	$p(t)$
Interaction: L1 = English					
Initial - final	-148.7	34.3	23.1	-4.34	< 0.001
Interaction: L1 = Japanese					
Initial - final	34.4	29.3	24.9	1.17	0.25

1351 For the rhotic model shown in Table A.7, the fixed effect of *tongue shape* is also added,
 1352 in which the baseline level is the Curled Up (CU) tongue shape.

Table A.7: Summary of the linear mixed-effect modelling for English /ɹ/ for TB lag (ms)

Full model					
Variable	β	SE	df	t	$p(\chi^2)$
Intercept	1.17	19.80	33.05	0.06	
L1					
Japanese	17.44	22.82	28.88	0.77	
Position					
Final	0.85	15.81	15.79	0.05	
Tongue shape					< 0.001
Front Bunched (FB)	-21.36	17.26	29.72	-1.24	
Front Up (FU)	39.72	12.13	143.27	3.28	
Mid Bunched (MB)	-7.46	24.23	24.40	-0.31	
Tip Up (TU)	35.64	10.63	276.76	3.35	
Interaction					0.02
Japanese:Final	-55.75	21.68	18.59	-2.57	
Post-hoc pairwise comparison					
Contrast	β	SE	df	t	$p(t)$
Interaction: L1 = English					
Initial-final	-0.85	17.2	22.9	-0.05	0.96
Interaction: L1 = Japanese					
Initial-final	54.90	16.7	33.9	3.29	0.002
Tongue shape					
CU - FB	21.36	20.3	55.7	1.05	0.83
CU - FU	-39.72	13.3	227.7	-2.98	0.03
CU - MB	7.46	29.0	43.1	0.26	1.00
CU - TU	-35.64	11.3	354.0	-3.15	0.02
FB - FU	-61.09	20.1	47.5	-3.05	0.03
FB - MB	-13.91	32.9	57.1	-0.42	0.99
FB - TU	-57.00	21.5	69.7	-2.65	0.07
FU - MB	47.18	29.5	44.1	1.60	0.51
FU - TU	4.08	14.9	187.4	0.27	1.00
MB - TU	-43.09	29.2	42.4	-1.48	0.59

Appendix B Individual variation in F2–F1

1353 Individual variation of F2–F1 is visualised in Figure B.1 for laterals and Figure B.2 for
 1354 rhotics. All F2–F1 values are within-speaker z -normalised. Each facet indicates at the top
 1355 the L1 of each speaker and the anonymised speaker ID underneath.

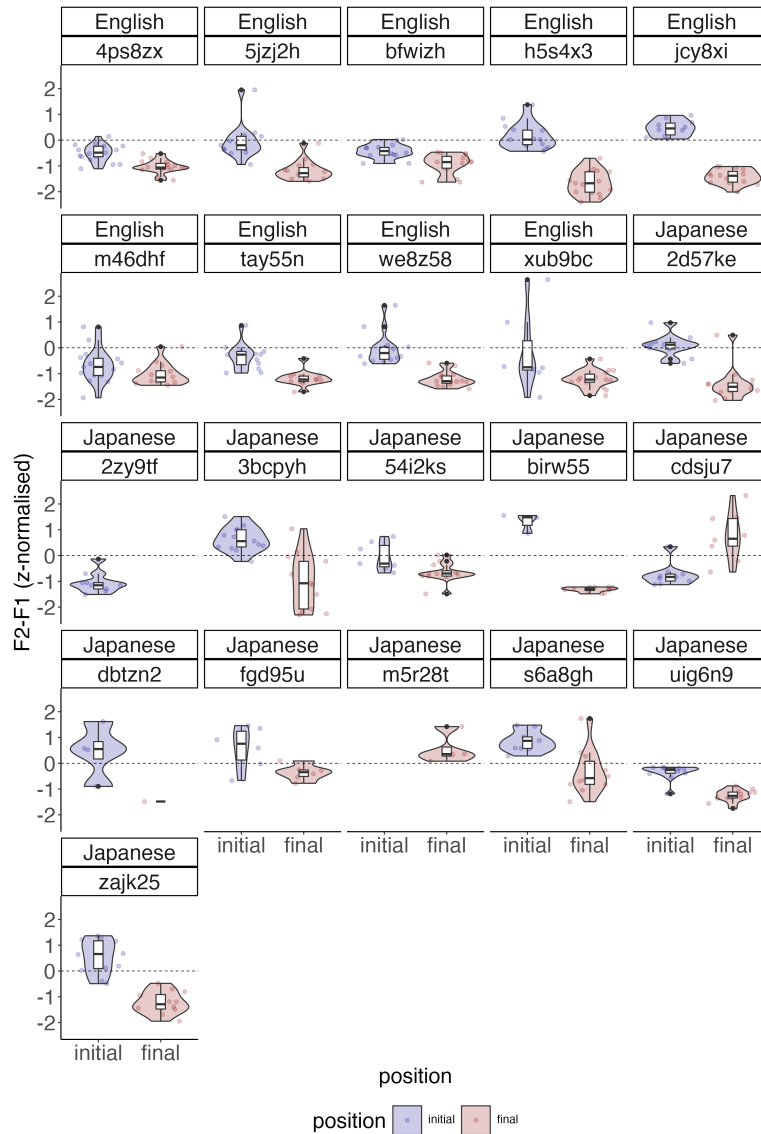


Figure B.1: F2–F1 at the liquid midpoint for laterals. F2–F1 values are within-speaker z -normalised. The language label in each facet indicates each speaker’s L1 with the anonymised speaker ID underneath.

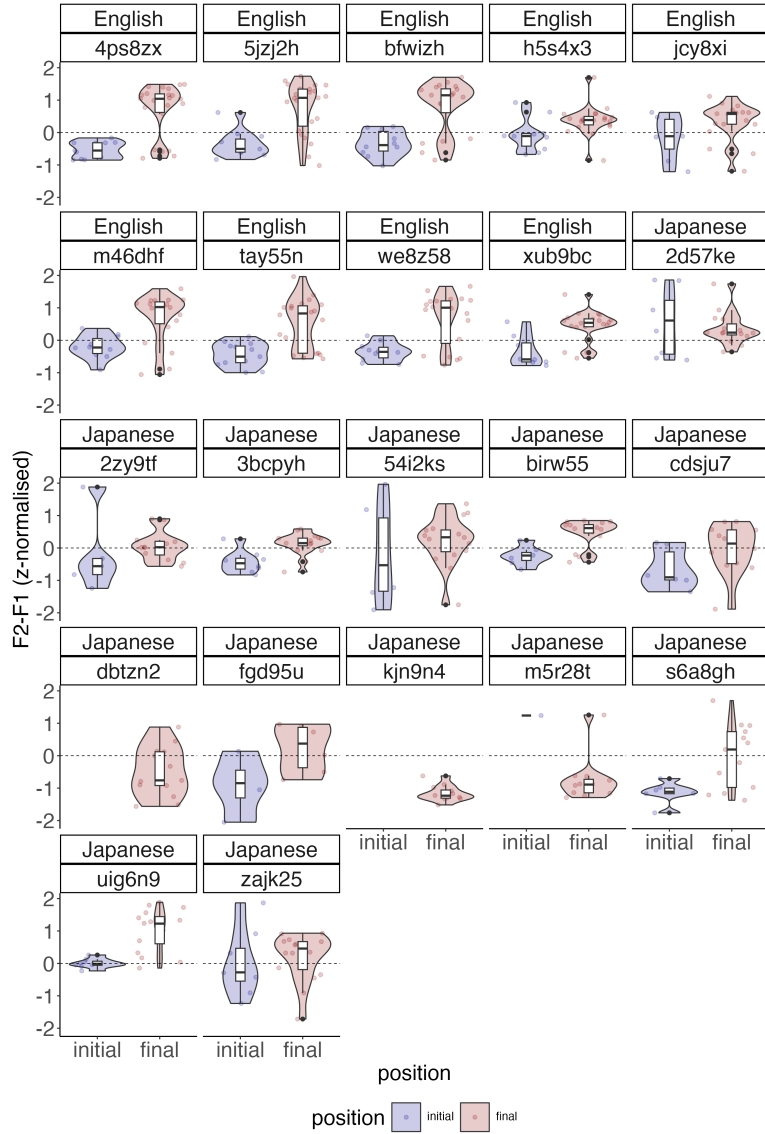


Figure B.2: F2–F1 at the liquid midpoint for rhotics. F2–F1 values are within-speaker z -normalised. The language label in each facet indicates each speaker’s L1 with the anonymised speaker ID underneath.

Appendix C Tongue shape comparison at maximal TB displacement

C.1 Population-level comparison

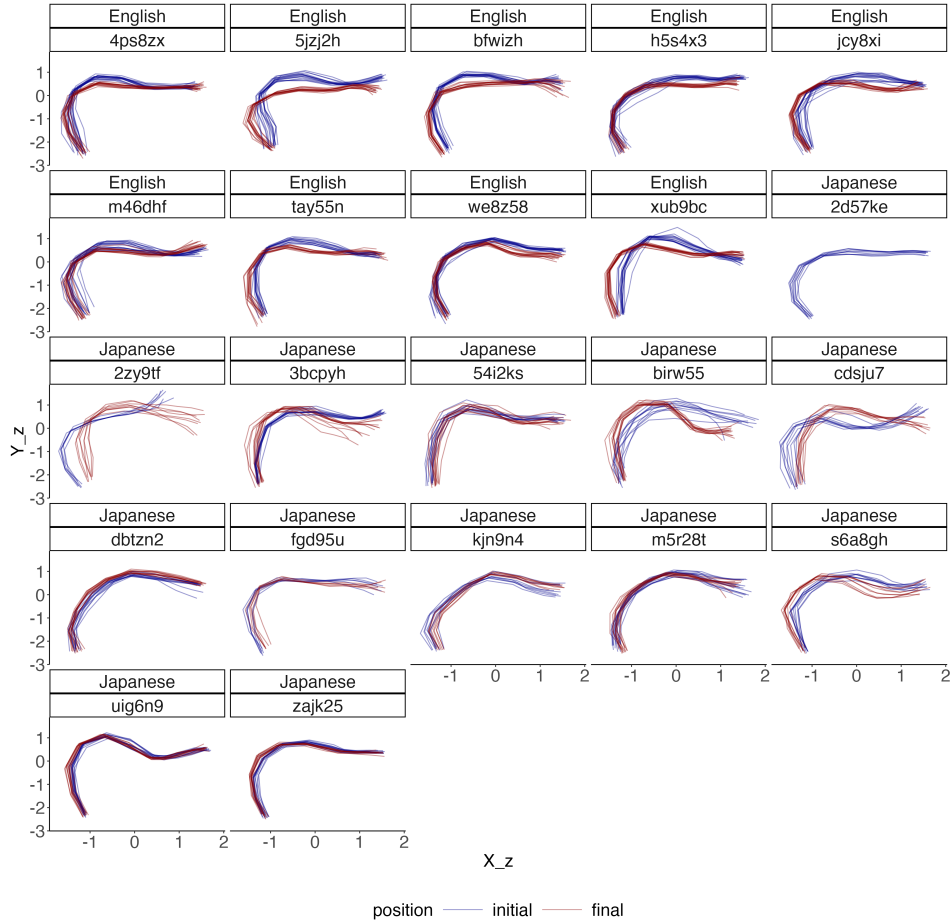


Figure C.1: Midsagittal tongue shape extracted at the maximal TB displacement for English /l/. Tongue tip to the right. Blue splines represent initial tokens and red final tokens. The language label in each facet indicates each speaker's L1 with the anonymised speaker ID underneath. The rotation and origin of the tongue splines are standardised relative to each speaker's bite plane. Note that the final tokens of the speaker 2d57ke are excluded from the analysis because they are heavily vocalised and thus the TB/TT displacement was not recorded.

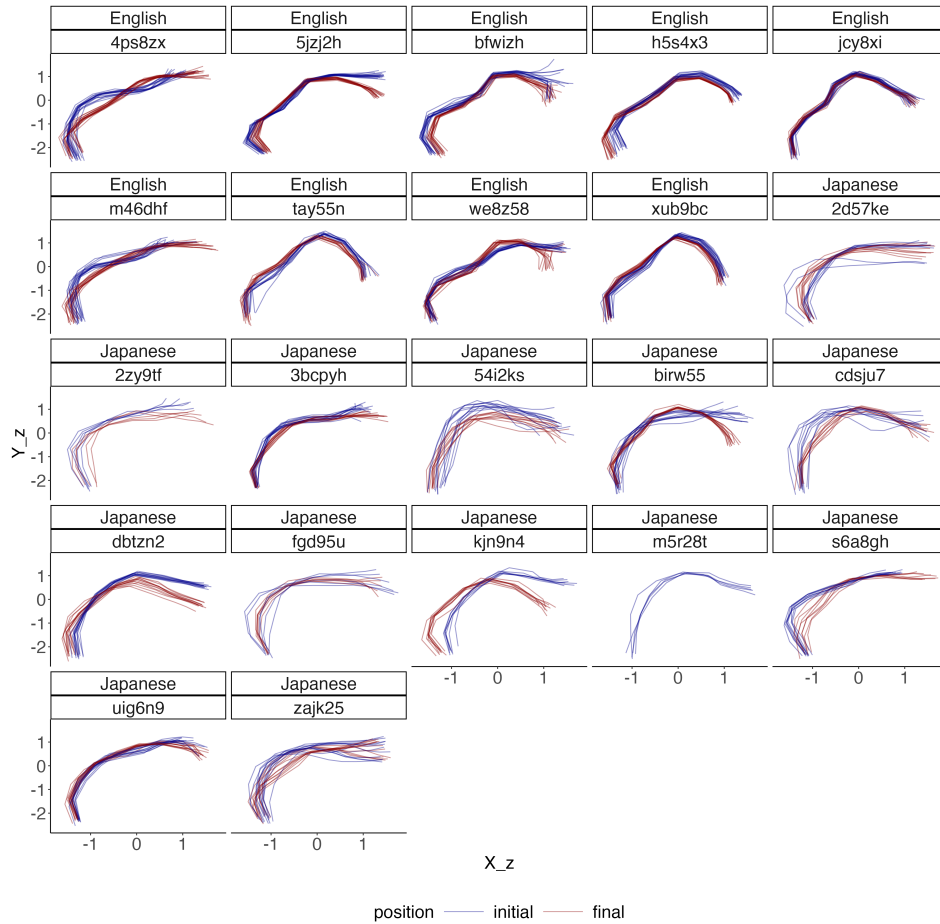


Figure C.2: Midsagittal tongue shape extracted at the maximal TB displacement for English /ɪ/. Tongue tip to the right. Blue splines represent initial tokens and red final tokens. The language label in each facet indicates each speaker's L1 with the anonymised speaker ID underneath. The rotation and origin of the tongue splines are standardised relative to each speaker's bite plane. Note that the final tokens of the speaker m5r28t are excluded from the analysis because the tongue shape for /ɪ/ is identical to that of the preceding vowel and thus is impossible to be distinguished.

Appendix D Individual variation in TB lag

1356 Individual variation of the TB lag measure is visualised in Figure D.1 for laterals and Fig-
 1357 ure D.2 for rhotics. The TB lag measure is shown in millisecond. Each facet indicates at
 1358 the top the L1 of each speaker and the anonymised speaker ID underneath.

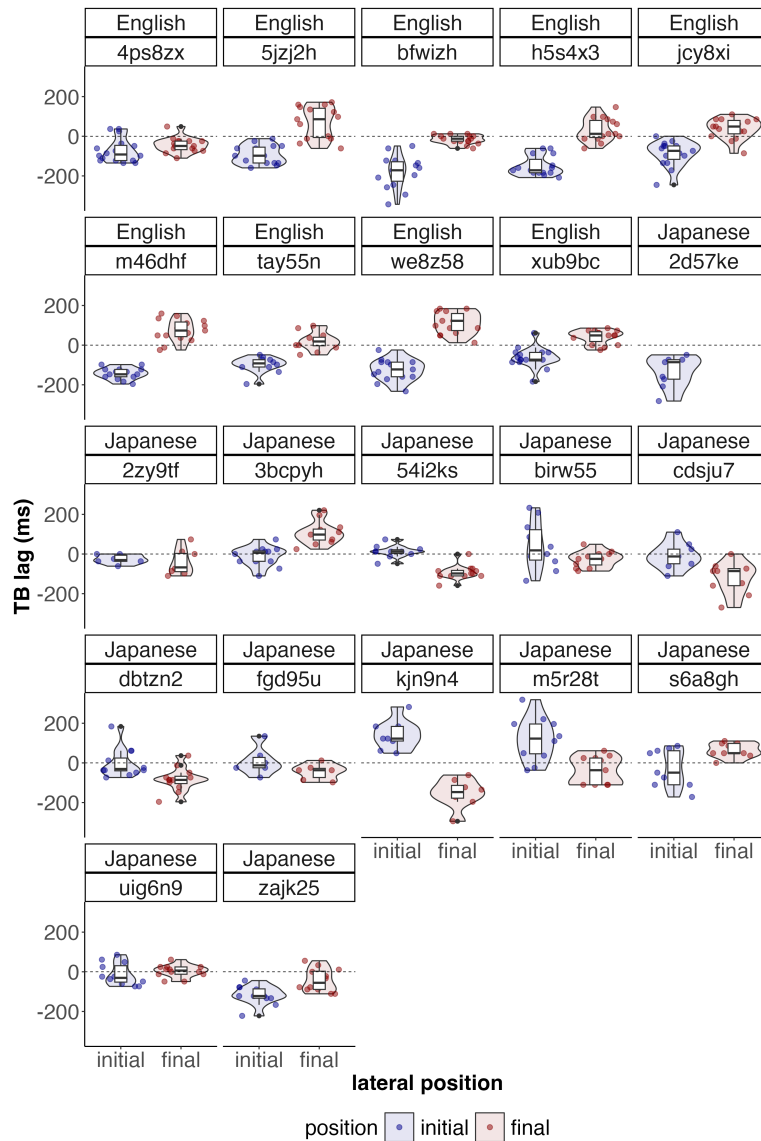


Figure D.1: Time lag between TT and TB in ms for laterals for each speaker. The language label in each facet indicates each speaker's L1 with the anonymised speaker ID underneath.

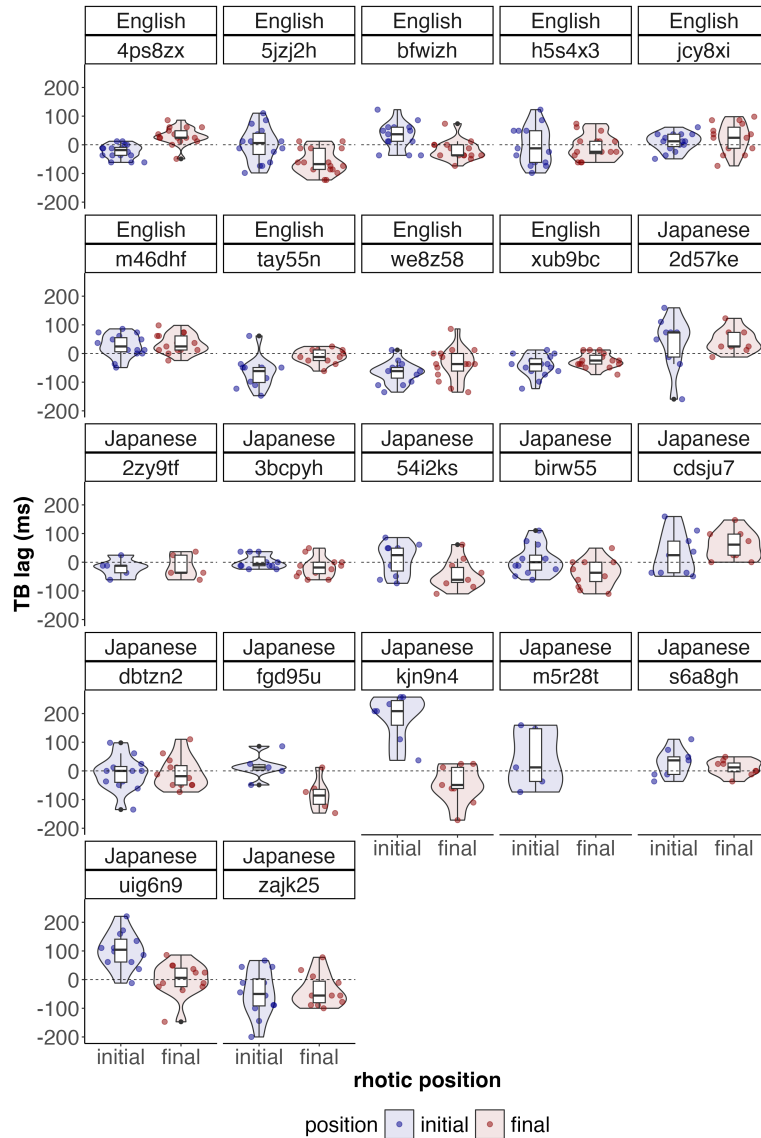


Figure D.2: Time lag between TT and TB in ms for rhotics for each speaker. The language label in each facet indicates each speaker's L1 with the anonymised speaker ID underneath.

Appendix E F3 analysis

E.1 Population-level comparison

1359 The acoustic and articulatory analyses presented in this study suggests a possibility of lip
1360 activity for L1 Japanese-L2 English speakers to implement the onset-coda allophony for
1361 laterals. Since lip rounding has an lowering effect on all formants (Ladefoged & Ferrari
1362 Disner, 2012), I conducted an additional analysis on the F3 frequeneis to see if there is
1363 any by-group difference. The following analysis shows that L1 Japanese-L2 English speakers
1364 produce word-final /l/s with lower F3 than L1 English speakers do. Together with the
1365 acoustic and articulatory analyses in the main text, this points to a possibility that L1
1366 Japanese-L2 English speakers may resort to non-lingual articulatory strategies, such as lip
1367 rounding, to lower the formant frequencies at the word-final position for laterals.

Table E.1: Mean F3 (Hz) at the liquid midpoint for word-initial and -final tokens of English /l/ and /ɭ/

L1	liquid	position	Mean F3 (Hz)	SD
English	/l/	initial	2977.45	522.80
		final	2944.88	544.90
	/ɭ/	initial	1786.47	448.99
		final	1962.19	309.67
Japanese	/l/	initial	2631.26	388.90
		final	2580.19	284.80
	/ɭ/	initial	2302.84	279.23
		final	2244.07	247.03

1368 For F3, separate linear mixed-effect models were fit for English /l/ and /ɭ/ due to the
1369 singular fit warning. For laterals, the best model predicts z -normalised F3 values by fixed
1370 effects of $L1$, $position$ and the interaction between them. The random effects include by-
1371 speaker and by-word varying intercepts. Model comparisons suggest that the full model is
1372 favoured over the nested model excluding the interaction term ($\chi^2(1) = 13.79$, $p < 0.01$).
1373 A post-hoc pairwise comparison suggests that this results from a statistically significant
1374 difference in F3 values for the coda laterals between L1 Japanese and L1 English speakers

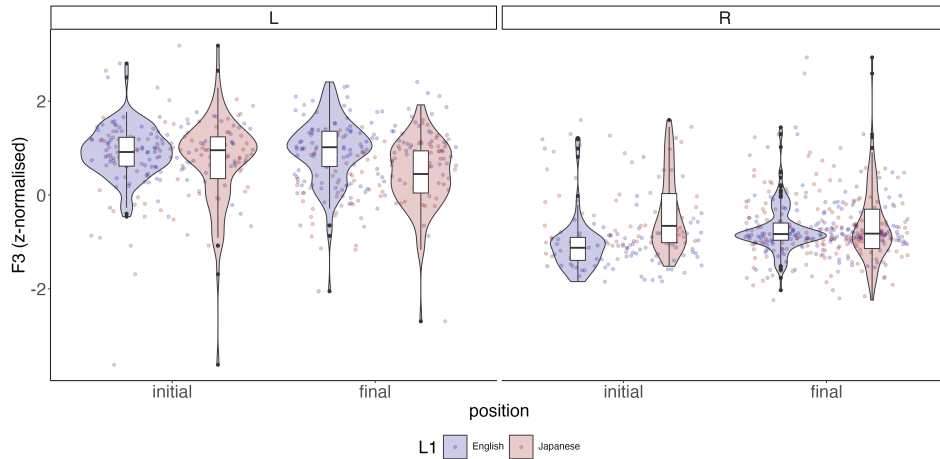


Figure E.1: F3 at the liquid midpoint for English /l/ (left) and /r/ (right). Blue represents L1 English speakers and red indicates L1 Japanese speakers. F3 values are within-speaker z -normalised.

1375 ($\beta = 0.70$, $SE = 0.21$, $t(28.9) = 3.39$, $p = 0.002$). The initial laterals do not suggest such
 1376 a between-group difference at a statistically significant level ($\beta = 0.22$, $SE = 0.21$, $t(30.1)$
 1377 $= 1.05$, $p = 0.30$). These results indicate that L1 Japanese-L2 English speakers produce
 1378 word-final tokens with lower F3 than L1 English speakers.

1379 The full model for rhotics share the same fixed effect structure, but it only includes
 1380 the by-speaker varying slope as the by-word varying slopes do not correctly predict the
 1381 variance. Model comparisons demonstrate that the full model improves the degree of model
 1382 fit over the nested model excluding the interaction between *L1* and *position*. A post-hoc
 1383 pairwise comparisons suggests that F3 values differ between the two speaker groups at a
 1384 statistically significant level in the initial position ($\beta = -0.78$, $SE = 0.16$, $t(49.5) = -4.81$,
 1385 $p < 0.001$) but not in the final position ($\beta = 0.07$, $SE = 0.14$, $t(28.1) = -0.50$, $p = 0.62$).
 1386 This suggests that L1 English speakers produce word-initial rhotics with lower F3 than L1
 1387 Japanese speakers, whereas the F3 values for word-final rhotics are comparable between the
 1388 two groups of speakers.

Table E.2: Summary of the linear mixed-effect modelling for laterals /l/ for F3

Full model					
Variable	β	SE	df	t	$p(\chi^2)$
Intercept	0.93	0.15	18.89	6.19	
L1					
Japanese	-0.22	0.20	20.27	-1.10	
Position					
Final	0.02	0.10	11.66	0.16	
Interaction					< 0.001
Japanese:Final	-0.48	0.13	430.41	-3.78	
Post-hoc pairwise comparison					
Contrast	β	SE	df	t	$p(t)$
Interaction: position = initial					
L1 English - L1 Japanese	0.22	0.21	30.1	1.05	0.30
Interaction: position = final					
L1 English - L1 Japanese	0.70	0.21	28.9	3.39	0.002

Table E.3: Summary of the linear mixed-effect modelling for rhotics /r/ for F3

Full model					
Variable	β	SE	df	t	$p(\chi^2)$
Intercept	-1.07	0.11	29.40	-9.63	
L1					
Japanese	0.78	0.16	38.47	5.01	
Position					
Final	0.34	0.08	489.83	4.17	
Interaction					< 0.001
Japanese:Final	-0.71	0.12	498.77	-5.81	
Post-hoc pairwise comparison					
Contrast	β	SE	df	t	$p(t)$
Interaction: position = initial					
L1 English - L1 Japanese	-0.78	0.16	49.5	-4.81	< 0.001
Interaction: position = final					
L1 English - L1 Japanese	-0.07	0.14	28.1	-0.50	0.62

E.2 Individual variation in F3

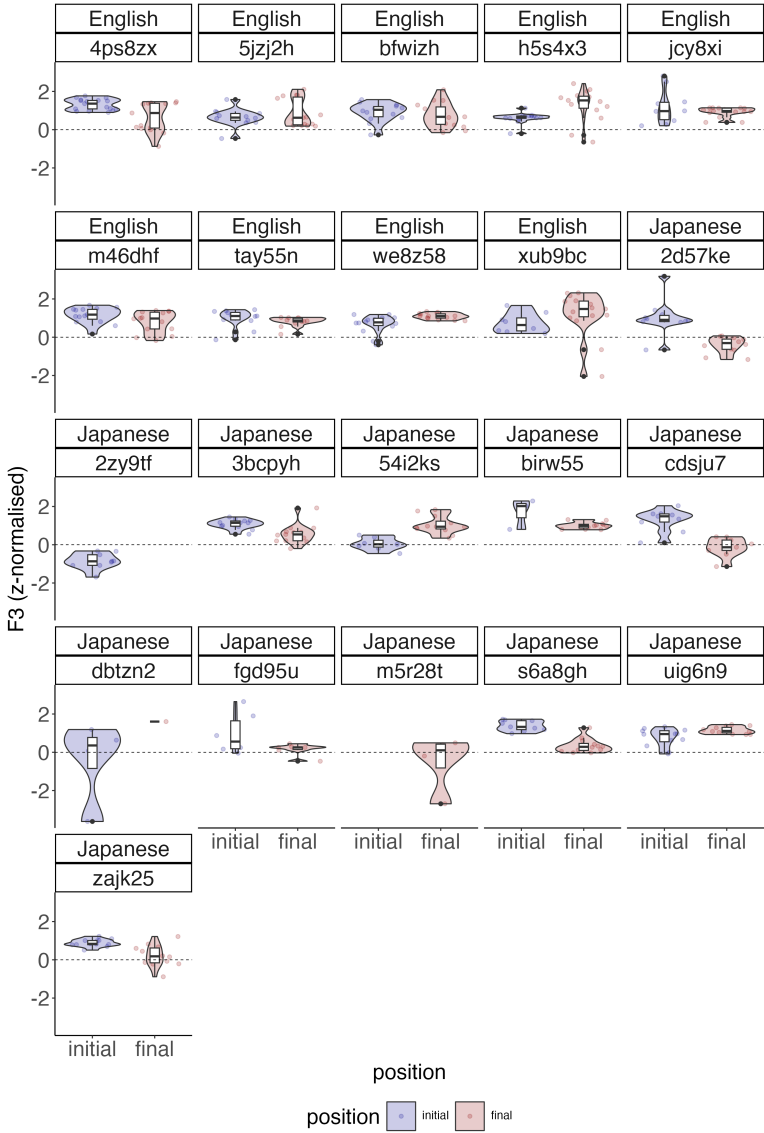


Figure E.2: F3 at the liquid midpoint for laterals. F2–F1 values are within-speaker z -normalised. The language label in each facet indicates each speaker’s L1 with the anonymised speaker ID underneath.

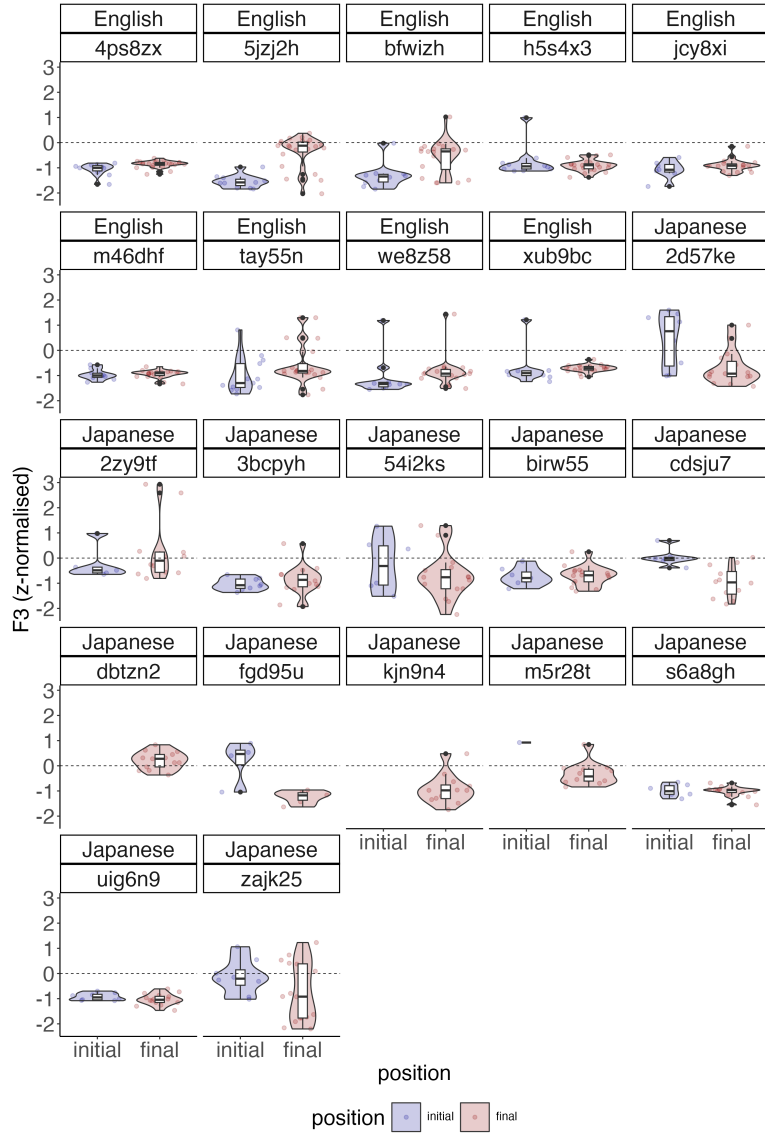


Figure E.3: F3 at the liquid midpoint for rhotics. F2–F1 values are within-speaker z -normalised. The language label in each facet indicates each speaker’s L1 with the anonymised speaker ID underneath.

Chapter 10

Summary and conclusions

10.1 Summary of findings

This PhD research sets out to better understand how L2 speakers make use of dynamic, time-varying phonetic cues to produce L2 segments. I use the L1 Japanese speakers' acquisition of L2 English liquid production as the testing ground, as this allows me to make specific hypotheses regarding the use of time-varying information in articulation. The particular focus in this PhD research is the degree of tongue dorsum involvement; English liquids /l ɹ/ involve an active movement of tongue tip and dorsum gestures that are patterned according to the syllabic position. This results in a certain degree of coarticulatory resistance from the neighbouring vowels and position-dependent gestural coordination patterns. The Japanese liquid, canonically an alveolar taps or flaps [ɾ], on the other hand, does not employ active participation of tongue dorsum in the production, resulting in a rather strong vocalic coarticulation. Given this background and the persistent difficulty that L1 Japanese speakers have in producing L2 English liquids, I hypothesise that the difference between the L1 Japanese and L2 English liquid productions in dynamic, time-varying characteristics may be a particular challenge that hinders them from producing target-like English liquids. And the review of the previous research yields two major themes running through the thesis: *articulation* and *dynamics*.

The thesis consists of two pilot studies and three main empirical studies to test the hypothesis both from the acoustic and articulatory viewpoints in terms of (1) liquid-vowel coarticulation and (2) position-sensitive allophony. In the first strand, an acoustic study has been presented in Chapter 7 in which I compare time-varying changes in F2–F1 and F3 in the production of word-initial liquid-vowel sequences in L1 and L2 English. The study demonstrates a clear between-group difference in terms of height and shape of the formant trajectories; along the F2–F1 dimension, whereas L1 English speakers show a consistent trajectory pattern, the trajectory shape and height in the L1 Japanese speakers’ production varies across the vowel contexts, with the GAMMs analysis showing a statistically significant difference for the main effect of the speakers’ L1. Particularly, in the /u/ context (e.g., in the production of words like *room* and *loom*), L1 Japanese speakers showed almost an monotonic decrease in F2–F1 as opposed to L1 English speakers showing a clear peak for the vowel target. This suggests that L1 Japanese speakers may differentiate the liquid and vowel less clearly than L1 English speakers. Overall, the study shows that formant dynamics uncovers finer-grained, specific between-speaker differences between L1 and L2 productions of English liquids that the static analysis, as commonly employed in the previous research, does not offer.

Informed by the acoustic differences just discussed above, I have then presented two studies that compared time-varying changes in midsagittal tongue shape in the intervocalic vowel-liquid-vowel sequence (Chapter 5) and the word-initial liquid-vowel sequence (Chapter 8) based on ultrasound tongue imaging data. In Chapter 5, I compared the articulation of intervocalic liquid consonants in Japanese and in English flanked by high front vowels produced by L1 Japanese and L1 English speakers. I identified major dimensions in the midsagittal ultrasound tongue imaging data using the Principal Component Analysis (PCA), in which the largest proportion of variance was explained by the first principal component (PC1) capturing the front-back movement of the tongue. By tracking time-varying changes in PC1 over the vowel-liquid-vowel interval, an approximation of the front-back movement of the

tongue, the results show that L1 Japanese speakers differ from L1 English speakers in the magnitude (for /l/) and timing (for /ɹ/) for maximal tongue retraction in the production of English words *believe* and *bereave*. In addition, the PC1 trajectory for the Japanese production of was almost linear for the Japanese word *biribiri*, suggesting that little tongue retraction is involved in the production of the Japanese liquid consonant. Although no statistical tests were conducted to formally test these observations, the study presents a clear between-group difference in tongue retraction.

In Chapter 8, I compare articulation of word-initial English liquids produced by L1 Japanese and L1 English speakers more formally than in Chapter 5. Similar to Chapter 5, I identified the major lingual dimensions involved in the English liquid articulation using PCA, suggesting that the largest proportion of variance is explained by tongue dorsum raising, which I focussed on in the subsequent statistical analysis. The derived PC1 trajectories over the liquid-vowel intervals were then further submitted to the Functional PCA (FPCA) in order to identify the major trends in the dynamic trajectory patterns, showing that the first functional principal component (FPC1) explains the largest variability in the data. Finally, I modelled the FPC1 values across three vowel contexts (i.e., /i/, /u/ and /æ/) using Bayesian linear-mixed effect models. Here, I focus on the variability in the FPC1 scores across the vowel contexts, as it is hypothesised that L1 Japanese speakers would exhibit a stronger vocalic coarticulation in the production of word-initial English liquids than L1 English speakers if they re-use the articulatory strategy from their L1. The results support this hypothesis for English /ɹ/, showing that the difference in tongue dorsum height between vowel contexts is larger for L1 Japanese speakers than for L1 English speakers. On the other hand, the two groups of speakers showed a similar distribution in FPC1 scores for English /l/, but this could be because coarticulatory resistance is weaker for English /l/ than for English /ɹ/. The results suggest that tongue dorsum height in English liquids is influenced by the following vowel to a greater extent for L1 Japanese speakers than for L1 English speakers, which could

be due to a carry-over effect in tongue dorsum movement from L1 Japanese.

The remaining two chapters, Chapters 6 and 9, addressed the second strand in this PhD thesis: the acquisition of position-sensitive allophony in L2 English liquids. Specifically, I investigated how L1 Japanese speakers signal the onset-coda distinction in laterals (Chapter 6) and in both laterals and rhotics (Chapter 9). In Chapter 6, I compared acoustics and articulation in advanced L1 Japanese - L2 English learners who produced intervocalic syllable-initial and syllable-final laterals in a carrier phrase. Their production was evaluated acoustically with the F2–F1 measure and midsagittal tongue shape, both extracted at the midpoint of the acoustically-defined lateral interval. The acoustic analysis shows overall that they produce target-like lateral allophony with a higher F2–F1 for onset than for coda with a statistically significant difference. The articulatory analysis, on the other hand, does not suggest a clear onset-coda distinction, in which the GAMMs models show a statistically significant effect of vowel context but not of syllable position. This study thus raises a possibility that L1 Japanese speakers may employ articulatory strategies that are different from the ones conventionally understood (e.g., lower and more retracted tongue dorsum for the final /l/s) that may not necessarily be captured by the midpoint analysis, calling for the need to take the dynamic information into account. In addition, this study only considers the production of English laterals by L1 Japanese speakers, so it remained inconclusive as to the degree of ‘target-likeness’ in the acoustic results.

The study presented in Chapter 9 aims to overcome these limitations and provide a holistic picture of the production of L2 liquid allophony by combining various types of analysis including the acoustic measure of F2–F1, midsagittal tongue shapes, and intergestural timing between tongue tip and tongue dorsum. In this study, I compared productions of syllable-initial and syllable-final English /l/ and /ɹ/ in the /i/ context between L1 Japanese and L1 English speakers. Replicating the results of the pilot study in Chapter 6, it was shown that L1 Japanese speakers do make a contrast between onset and coda laterals along the F2–F1 dimension, and this

can be considered to be target-like production given a lack of statistically significant effect of the speaker's L1. Agreeing again with the pilot study, however, such a clear onset-coda contrast for laterals was not found in L1 Japanese speakers' articulation; they seemed to employ a similar midsagittal tongue shape in contrast to L1 English speakers using a lower and more retracted tongue shape syllable-finally than for syllable-initially. In addition, the intergestural timing between tongue tip and dorsum was measured by obtaining the lag between the timing of these two gestures, which showed little effects of syllable position for L1 Japanese speakers as opposed to L1 English speakers who showed a clear onset-coda contrast in an expected manner (i.e., tongue tip preceding tongue dorsum for initial laterals). Although no clear onset-coda distinction was found for English /ɹ/ for both speaker groups in terms of both acoustic and articulatory measures, it was shown that the rhotic acoustics tends to exhibit the opposite pattern to the lateral acoustics, which could result from the polarity effect of the liquid acoustic system as a whole. This study overall raises a possibility that L2 learners could produce target-like allophonic variation in their L2 speech acoustics, but they might stick to the articulatory strategies that are available to them. In the context of the current research, L1 Japanese speakers might compensate a lack of tongue dorsum control by using other articulatory strategies that could not be captured by the midsagittal ultrasound tongue imaging, such as the labial gestures, to achieve an overall lower F2–F1 for the final laterals and eventually to make a clear onset-coda distinction in the lateral allophony.

To summarise all these, the findings from the five studies included in this thesis indicate that L1 Japanese speakers:

1. exhibit a greater variability in the time-varying acoustics and articulation than L1 English speakers as a function of neighbouring vowels (Chapters 7 and 8);
2. use the tongue dorsum in producing L2 English liquids less actively than L1 English speakers, which could be due to the influence of the articulatory strategies from that of L1 Japanese liquid (Chapters 5 and 8), and;

3. show a target-like liquid allophony in acoustics but use different articulatory strategies from that of L1 English speakers (Chapters 6 and 9).

More importantly, this PhD thesis demonstrates clear differences in *dynamic* characteristics in the English liquid production between L1 Japanese and L1 English speakers, in addition to the *static* property as has been understood previously. Specifically, this points to a specific challenge that hinders L1 Japanese speakers from producing L2 English liquids in a target-like manner, which is the tongue *movement*, especially the tongue dorsum. In the section below, I discuss how each of these findings contribute to advancing our understanding of the nature of L2 speech production.

10.2 Contribution of the thesis

The most significant contribution of this thesis to the existing body of knowledge is that it offers an explanation as to how articulations in two systems may interact with each other. This has been made possible in this PhD research because of (1) the study's focus on dynamic properties whose scope may span beyond an individual segment and (2) the addition of articulatory data to the existing body of L2 speech production research. I will discuss how the thesis makes a contribution to the research field in light of these two aspects.

10.2.1 Speech dynamics provides an important language-specific phonetic detail in L2 speech learning.

Broadly, this thesis demonstrates that, at least in some cases, it is necessary to look beyond the scope of an individual segment in order to gain a better understanding of how a given L2 segment can be acquired in L2 speech production. The main contribution of this thesis is that it demonstrates that speech dynamics could explain the variability attested in L2 speech production. The thesis, in particular, shows

that it is the degree of tongue dorsum involvement that L1 Japanese speakers need to adjust in producing the target-like L2 English liquid system. The possibility of the difference in tongue dorsum movement has been implicated in one of the previous articulatory studies of L1 Japanese speakers' production (Zimmermann et al., 1984) as well as in acoustic studies in which the F2 signal was seen as a proxy of overall front-back movement of the tongue (Saito & van Poeteren, 2018). The static analyses employed in these studies, however, did not account for specific production mechanisms that make it difficult for L1 Japanese speakers to produce target-like L2 English liquids; in particular, it remained unclear why it is the degree of tongue retraction or front-back tongue movement that was related to L1 Japanese speakers' production of L2 English liquids.

This PhD thesis agrees with previous research that L2 speech learning involves acquisition of *both* segmental targets and coarticulation (Beristain, 2022; Oh, 2008). In the acoustic study presented in Chapter 7, the dynamic analysis suggests that the two groups show clear differences in word-initial liquid-vowel coarticulation, shown by the difference in trajectory shape and height along F2–F1. While the static analysis showed differences in liquid acoustic target for F2–F1 between L1 Japanese speakers and L1 English speakers, the two groups did not differ in the static vowel targets for /i/ and /u/ (see the supplementary materials included in the study). This suggests that, if coarticulation is a universal, automatic process as a result of transition between targets, then we would expect that L1 Japanese speakers would exhibit similar trajectory patterns in both /i/ and /u/ contexts despite differences in liquid targets. The dynamic analysis shows, however, a clear difference in trajectory shape and height between the two groups in both vowel contexts, with even clearer between-group differences in overall trajectory pattern in the /u/ context. These findings agree with the claim that coarticulation is a language-specific process that needs to be acquired as part of L2 speech learning (Beristain, 2022; Keating, 1985; Oh, 2008). Specifically, Oh (2008) shows that some of the L1 English-L2 French speakers achieve target-like vowel targets but differ in

the shape of formant trajectories in their production of the back vowel /u/ in coronal and non-coronal contexts in French words ‘ou’ and ‘tou’ compared to English words ‘who’ and ‘two’.

My study further adds articulatory evidence to the acquisition of L2 coarticulation. In Chapter 8, I demonstrate that L1 Japanese speakers and L1 English speakers differ in the articulation of word-initial liquid-vowel sequences in English along the tongue dorsum dimension and that L1 Japanese speakers show a greater variability as a function of vowel contexts than L1 English speakers. Although I did not conduct static analysis in Chapter 8, which would have augmented the target versus coarticulation discussion better, the results could identify a specific mechanism as to why such a between-group coarticulatory difference has been attested. The variability according to the phonetic context, by definition, shows the degree of *coarticulatory resistance*, which inversely correlates with the degree of active tongue dorsum involvement (Recasens & Espinosa, 2009). Given the L1 influence assumed in L2 speech learning, L1 Japanese speakers’ production of English liquids would be under an influence of L1 Japanese liquid. Also, the liquid consonants in Japanese and English are expected to differ in the degree of coarticulatory resistance; tongue dorsum is actively involved in the articulation of English liquids, especially greater for English /ɹ/ than for English /l/ (Proctor et al., 2019) and greater for dark /l/ than for clear /l/ (Recasens & Rodríguez, 2016), whereas the coarticulatory resistance in alveolar taps and flaps is small (Maekawa, 2023; Recasens, 1991; Recasens & Rodríguez, 2016). Taken together, differences in coarticulation shown here have an articulatory basis, such that L1 Japanese speakers transfer the pattern of tongue dorsum movement from L1 Japanese to L2 English, resulting in a greater variability attested in their production of L2 English liquids. This explains why L1 Japanese speakers’ production of L2 English liquids is more variable than that of L1 English speakers, and also, the differences in trajectory patterns, not just in static segmental targets, suggest that coarticulation is one kind of phonetic detail that needs to be learnt in L2 speech learning (Flege, 1995; Oh, 2008).

One limitation of the coarticulatory studies presented in this thesis is that it only considers coarticulation in one direction (i.e., vowel-to-liquid coarticulation in word-initial liquid-vowel sequences) when coarticulation in fact shows a bidirectional relationship; it is indeed possible that the vowel realisation is influenced by the realisation of liquids. Coarticulatory resistance and aggressiveness is a mirror image of each other, such that a given segment that exerts stronger coarticulatory resistance to the coarticulatory influence from the neighbouring segments would also influence the realisations of the neighbouring segments. This is particularly the case of English liquids; previous research suggests that English /ɹ/ typically exhibits a stronger coarticulatory aggressiveness than English /l/ does (Proctor et al., 2019) and that dark /l/ involves a greater degree of coarticulatory aggressiveness than clear /l/s do (Recasens & Espinosa, 2005). This suggests that vowel qualities also vary depending on the degree of coarticulatory aggressiveness for the preceding consonants, as much as the liquid quality changing as a function of the neighbouring vowels.

Nevertheless, this consideration further reinforces the importance of dynamic perspectives in understanding speech production, as the dynamic analysis necessarily encompasses the liquid-vowel relationship as a whole. Specifically, the dynamic analysis in the production of English liquids opens up a possibility that L1 Japanese speakers' vowel productions are variable because their English liquid articulation does not exert much coarticulatory aggressiveness. It has been argued in the previous research that it is often challenging to disentangle the effect of liquids from that of vowels in understanding the dynamic properties in the liquid-vowel sequence (Kirkham & Nance, 2017; Macdonald & Stuart-Smith, 2024). It is because of this difficulty in separating between liquid and vowel that has motivated me to pursue the dynamic approach in the analysis of English liquid production (e.g., Plug & Ogden, 2003), and the dynamic analysis would allow us to provide a more complete picture on how liquid and vowel interacts with each other and how such interaction patterns might differ between L1 and L2 speech production than static analyses

alone.

Although dynamics have not fully been addressed in current perception-based theoretical frameworks in L2 speech learning, their assumptions show the possibility of capturing the role of time-varying properties in L2 speech learning, including coarticulation. The assumption of L2 speech perception at the positional allophonic level, for example, not only explains a varying degree of perceptual accuracy of English /l/ and /ɾ/ for L1 Japanese speakers across syllabic positions, but also could potentially account for different phonetic realisations of segments due to context-specific coarticulatory influence (Bradlow et al., 1999; Colantoni et al., 2015). In addition, the Perceptual Assimilation Model for L2 Learning states that coarticulation could be one of the factors that influence L2 learners' accurate perception of L2 sounds (Best & Tyler, 2007).

10.2.2 L2 speakers compensate L2 speech production with existing L1 articulatory strategies.

Another contribution of this PhD thesis is that the articulatory data offers a new hypothesis regarding the specific mechanisms to explain variability attested in L2 speech production. In particular, it demonstrates that L2 learners may use various, sometimes non-target-like, articulatory strategies to produce target-like acoustic contrast (Song & Eckman, 2021). This is shown in the studies presented in Chapters 6 and 9, in which the combination of acoustic and articulatory analyses of L1 Japanese speakers' production of L2 English liquid allophony suggests that they produce target-like liquid allophony in L2 English in acoustics but not in articulation.

Through the discrepancy between acoustics and articulation, these studies demonstrate the utility of ultrasound tongue imaging in L2 speech production research as it is suggested that acoustic analysis alone does not uncover the whole picture of L2 speech production. The acoustic-articulatory discrepancy has been documented in the previous research on the L2 acquisition of allophonic variation (e.g., Colantoni et al., 2023b; Kochetov, 2022). Also, Song and Eckman (2021)

shows similar findings that some of their L1 Korean/L1 Spanish speakers of L2 English made a distinction between L2 English tense-lax vowel phonemic contrast (e.g., /i/-/ɪ/ and /ɛ/-/æ/) in acoustics but not in midsagittal tongue shape captured by ultrasound tongue imaging. They contextualise the findings along the line of *covert contrast*, an imperceptible distinction that a given speaker makes in acoustics and/or in articulation (e.g., Scobbie et al., 1996). They explain the findings by arguing that L2 learners may use less typical articulatory strategies to produce target-like acoustic outputs because of a lack of knowledge in relevant articulatory dimensions (Song & Eckman, 2021). The findings presented in this PhD thesis support this view and add further evidence of such acoustic-articulatory discrepancy not just in phonemic contrasts but also in phonetic (i.e., allophonic) contrasts in L2 speech production.

The analysis provided in Chapter 9 draws a more complete picture of the acquisition of L2 speech production by combining the acoustic and articulatory data. Much of the knowledge that we have on L2 speech production is based on the theoretical frameworks that have been developed around the acoustic findings (Escudero, 2000; Flege, 1995; Flege & Bohn, 2021), and the number of studies incorporating articulatory data is still relatively small (e.g., Wieling, 2018). This study demonstrates that articulatory data would offer explanations of the specific mechanisms in L2 speech production that acoustic and/or perception-based evaluation of the production data may not uncover. In the context of my research, L1 Japanese speakers' production has been commonly analysed by means of perceptual evaluation by L1 English-speaking listeners (e.g., Aoyama et al., 2004) and/or by acoustic analysis (e.g., Aoyama et al., 2019; Flege, 1995; Saito & Munro, 2014). The articulatory mechanisms producing the acoustic output have therefore been inferred from the acoustic findings (Aoyama et al., 2023; Saito & van Poeteren, 2018). Such articulatory descriptions, however, remain abstract as they are based on existing phonological descriptions such as the correspondence between F2 and vowel backness, and they have received little empirical support (cf. Scobbie et al., 2012).

Furthermore, previous research has also claimed that perceptual evaluation may not always detect subtle physical characteristics in L1 Japanese speakers' production of L2 English liquids, a similar claim made in the research of covert contrast (Aoyama et al., 2019; Scobbie et al., 1996; Song & Eckman, 2021).

In Chapter 9, the acoustic data along F2–F1, the acoustic dimension known to correlate with liquid quality (Carter & Local, 2007; Sproat & Fujimura, 1993), would suggest that L1 Japanese speakers have acquired target-like allophonic variation for L2 English liquid production. Articulatory data, however, suggest that they seemed to use a single articulatory strategy. Specifically, L1 Japanese speakers exhibit non-target-like distinction in tongue shape between word-initial and word-final laterals, including those who do not make a clear distinction and those who show a opposite tongue shape pattern to that of L1 English speakers. Also, the timing analysis shows no difference in the coronal-dorsal coordination pattern between the onset and coda laterals, meaning that they coordinate tongue tip and tongue dorsum similarly between the two positions, as opposed to L1 English speakers whose initial laterals involve the coronal-dorsal sequence as predicted (Browman & Goldstein, 1995; Proctor et al., 2019; Sproat & Fujimura, 1993). This opens up a possibility that L1 Japanese speakers may utilise different articulatory strategies to produce target-like acoustic contrast between the initial and final laterals, and an additional analysis of F3 suggests that L1 Japanese speakers might use lips to differentiate onset and coda laterals.

This articulatory observation can further be elaborated by incorporating insights from the dynamic approach discussed earlier. In particular, the lack of active involvement of tongue dorsum gesture in L1 Japanese speakers' production of L2 English liquids could explain why they do not exhibit clear onset-coda lateral allophony in articulation. The lateral allophony in American English can be classified as 'extrinsic', meaning that the onset and coda laterals have distinct articulatory targets (Recasens, 2012). This further supports the observation that the L1 English-speaking participants in the study, who are speakers of either American

English or Canadian English, exhibits distinct articulatory patterns between onset-coda laterals.

Taken together, this PhD thesis offers a new and finer-grained account of the mechanisms underlying L1 Japanese speakers' production of L2 English liquids. L1 Japanese speakers generally have less active control on tongue dorsum because of L1 influence given the findings from the dynamic analysis (cf. Maekawa, 2023; Recasens & Rodríguez, 2016), which hinders them from coordinating the tongue dorsum gesture relative to the tongue tip gesture as a function of syllabic position (cf. Proctor et al., 2019; Sproat & Fujimura, 1993). Nevertheless, they can perceive the clear-dark contrast in acoustics given their perceptual sensitivity to F2 (cf. Iverson et al., 2003; Recasens, 2012; Saito & van Poeteren, 2018), but they have not acquired the target-like articulatory cues to make the position-based allophony (Song & Eckman, 2021). For this reason, they might use non-typical articulatory cues that are readily available to them, which, in the case of the current research, could be the labial gestures that have effects of lowering formants (Ladefoged & Ferrari Disner, 2012; Song & Eckman, 2021). While this broadly agrees with previous postulations that L1 Japanese speakers re-deploy articulatory dimensions that are available in L1 Japanese (i.e., the front-back dimension associated with the F2 frequency), I would argue that this explanation offers a more finer-grained description of factors involved in L1 Japanese speakers' production of L2 English liquids, compared to existing descriptions based on the associations between acoustics and abstract articulatory properties (e.g., Aoyama et al., 2023; Bradlow, 2008; Saito & van Poeteren, 2018).

10.3 Limitations, future research and concluding remarks

Overall, this PhD thesis addresses possible L1 influence seen in the case of L1 Japanese speakers' production of L2 English liquids through the lens of *articulation* and *dynamics* in speech production. One limitation of this thesis is a lack of formal

direct comparison between the Japanese and English consonants. Whereas a pilot study presented in Chapter 5 did involve comparison of the midsagittal tongue shape between Japanese and English liquids produced by L1 Japanese speakers, this remained an impressionistic, informal description. Directly comparing the Japanese and English liquid production needs further considerations in the statistical design due to data sampling from different populations (i.e., the Japanese liquid consonant produced only by L1 Japanese speakers); this necessarily involves within-participant design focussing only on L1 Japanese speakers' data. Despite these limitations, the results presented in Chapter 5 might be the first in directly comparing between Japanese and English liquid articulations among a handful of existing articulatory studies on this topic (Masaki et al., 1996; Moore et al., 2018; Zimmermann et al., 1984), and the findings from the thesis overall provide a more specific account of a possible mechanism underlying difficulty L1 Japanese speakers have in producing English liquids. Developing a statistical design that allows for such within-speaker comparison would further provide a more direct evidence as to the nature of L1 influence in the articulatory domain than the evidence provided in this PhD research.

In considering ways of direct comparison between Japanese and English liquids, one possible way to go forward is to classify L1 Japanese speakers' production of English liquids according to the listeners' judgement of intelligibility. This would allow me to extend the findings of previous research; it has been shown that L1 English-speaking listeners' judgement of production accuracy correlate with the F2 frequency and the F2–F1 distance in L1 Japanese speakers' production of English liquids, suggesting that L1 Japanese speakers may be able to produce English liquids that are perceived as 'accurate' without acquiring 'target-like' production strategies (Aoyama et al., 2019; Saito & van Poeteren, 2018). This supports the possibility raised in Chapter 9 regarding the use of lips to realise lateral allophony. While the relationships between vocal tract shape and the listener's judgement of intelligibility were discussed in a previous articulatory study in the context of L1 Japanese speakers' production of L2 English liquids, there were some methodological

challenges including the qualitative nature of analysis and the fact that articulatory and acoustic data were not fully aligned with each other as they were collected at different occasions (Masaki et al., 1996). With the quantification of tongue shape and simultaneous acoustic-articulation synchronisation enabled by ultrasound, conducting an additional perceptual study would provide a more complete picture of L1 Japanese speakers' production of English liquids. This will also provide another dimension to classify the English liquid tokens produced by L1 Japanese speakers, enabling an investigation of the relationships between the L1-L2 similarity in tongue shape and speech intelligibility. This would bear an important pedagogical implication in English pronunciation teaching given the recent priority on intelligible and comprehensible pronunciation over native-like pronunciation (e.g., Munro & Derwing, 2015).

This PhD thesis is focussed on the specific context of L1 Japanese speakers' production of L2 English liquids. In order to claim the importance of speech dynamics involved in L2 speech production, it would be ideal to consider a wider range of L1-L2 pairings. One possible way to go forward is to test how L1 English speakers would acquire alveolar taps or flaps; previous research has suggested a similar interaction between the tongue shapes used for L1 and L2 in L1 American English speakers' production of L2 Spanish liquids, but the study only considers acoustic characteristics (Olsen, 2012). If the L1 articulatory routine include dynamic properties as suggested in this thesis, it could be predicted that L1 English speakers would have to either suppress the tongue dorsum gesture that is inherently specified for English /l/ and /ɹ/ or show different substitution patterns such as using /t/ or /d/ to produce alveolar taps/flaps. Olsen (2012) shows that L1 American English speakers are better at producing Spanish taps in the phonetic environment in which alveolar taps are observed in American English, suggesting that L1 English speakers' substitution pattern would vary as a function of vowel context, where a dynamic analysis of articulation would be useful. Overall, this PhD thesis demonstrates the utility of dynamic analysis in uncovering the subtle phonetic details involved in L2

speech production, but further research is clearly needed that considers a wide range of contexts.

Methodologically, this thesis demonstrates that ultrasound tongue imaging is a useful tool in investigating articulation in L2 speech production (e.g., Gick et al., 2008). Ultrasound data in this thesis has uncovered the spatiotemporal properties of the tongue dorsum that is not captured clearly by other existing methods such as electropalatography (EPG) due to a lack of linguopalatal contact and electromagnetic articulography (EMA) due to the nature of sensor placement often constrained to the front part of the tongue. One obvious limitation of the ultrasound data included in this research, however, is a challenge involved in the collection of lip movement data. Previous research has collected and analysed the side-view and frontal-view of the speaker's lips using a lip camera (e.g., King & Ferragne, 2020; Lawson et al., 2019). Existing descriptions suggest that labial gesture is important to understand English /ɹ/ in which interactions are hypothesised between tongue shape and the degree of labial gesture (e.g., Delattre & Freeman, 1968; King & Ferragne, 2020; Mielke et al., 2016; Proctor et al., 2019). King and Ferragne (2020) shows that bunched tongue shapes, particularly the front bunched (FB) and mid bunched (MB) configurations, could induce lip rounding more than other configurations. In addition, the results from Chapter 9 suggest a possibility that L1 Japanese speakers might use their lips to distinguish onset-coda laterals given lower F2–F1 and F3 values. It will therefore be interesting if future research could address the role of lip rounding involved in the articulation of L2 English liquids, given especially that (1) L1 Japanese speakers use different tongue shape strategies from that of L1 English speakers for English /ɹ/ and that (2) L1 Japanese speakers may employ labial gestures in realising positional contrast for English /ɹ/.

Finally, it needs to be emphasised that the findings regarding L1 Japanese speakers' articulation of L2 English liquids in this PhD research are restricted to the midsagittal dimension, whereas English liquids can also be characterised by a lateral movement. The formation of lateral channel is a vital part in understanding

the articulation of English laterals (Browman & Goldstein, 1995; Sproat & Fujimura, 1993; Ying et al., 2021). Although previous research shows opposing views as to whether lateral channel formation is an active process (Browman & Goldstein, 1995; Sproat & Fujimura, 1993; Strycharczuk et al., 2020), Ying et al. (2021) demonstrates using EMA that the lateral movement of the tongue could be considered as part of the active gestural coordination pattern. English /ɹ/, on the other hand, may involve lateral bracing, in which the back part of the tongue is braced against the upper molar (Collins et al., 2013; Gick et al., 2017). Almost all sounds in English may involve lateral bracing, except for the lateral and back vowels, which agrees with the observation of articulatory settings in English (Gick et al., 2017; Honikman, 1964). This contrasts with the descriptions of articulatory settings in Japanese in which lateral bracing might not be a salient feature (Somedá, 1966). Although lateral /l/ could surface as a free allophone of Japanese /r/, the speaker's lateralisation strategy in L1 Japanese may not facilitate accurate production of L2 English /ɹ/ (Kawahara & Matsui, 2017; Morimoto, 2021). Despite these observations, there is little empirical research that formally addresses tongue lateralisation in the context of L1 Japanese speakers' production of L2 English liquids. With a possibility of EMA-ultrasound co-registration design (e.g., Kirkham et al., 2023), it would be a fruitful path to investigate the role of tongue lateralisation.

With all these limitations and future directions in mind, nevertheless, I believe that this PhD thesis builds a solid foundation of articulatory and dynamic approaches to the “Japanese /r/-/l/ problem” (Flege et al., 2021, p. 84). The collection of five empirical studies included in this study, based on the acoustic and articulatory data of English liquid tokens produced by a total of 41 L1 Japanese speakers and 14 L1 North American English speakers, suggest that (1) L1 Japanese speakers exhibit different and variable tongue dorsum movement across vowel contexts compared to that of L1 North American English speakers and that (2) L1 Japanese speakers may employ articulatory strategies that are not readily used in the production of L1 North American English speakers in realising phonetic

details involved in English liquids. These findings result from this thesis' focus on *articulation* and *dynamics*, demonstrating that this is a promising path to move L2 speech production research forward. More broadly, this PhD thesis demonstrates that research gaps still exist even in such a well-studied, long-standing research topic in phonetics and second language acquisition research. This could be taken as a reminder that no research is complete on its own and there is a lot more to explore, especially in things that we often take for granted such as speech articulation. Acquiring a second language is an effortful venture that sometimes feels like navigating oneself without clear indications of whether one is doing right or wrong; I hope that this PhD research provides my fellow second language learners (at least those who struggle with the /r/-/l/ problem) a clearer map to overcome the pronunciation difficulty by visualising the (once) invisible.

Consolidated list of references

- Akamatsu, T. (1997). *Japanese phonetics: Theory and practice*. Lincom Europa.
- Al-Tamimi, J., & Palo, P. (2024). Retraction of the whole tongue induced by pharyngealisation in Levantine Arabic: A between-subject account using static and dynamic PCA and GAMMs. In I. Wilson, A. Mizogushi, J. Perkins, J. Villegas, & N. Yamane (Eds.), *Ultrafest XI: Extended Abstracts* (pp. 79–83). <https://doi.org/10.5281/zenodo.12578650>
- Alwan, A., Narayanan, S., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *The Journal of the Acoustical Society of America*, 101(2), 1078–1089. <https://doi.org/10.1121/1.417972>
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42(4), 529–555. <https://doi.org/10.1111/j.1467-1770.1992.tb01043.x>
- Antolík, T. K., Pillot-Loiseau, C., & Kamiyama, T. (2019). The effectiveness of real-time ultrasound visual feedback on tongue movements in L2 pronunciation training: Japanese learners' progress on the French vowel contrast /y/-/u/. *Journal of Second Language Pronunciation*, 5(1), 72–97. <https://doi.org/10.1075/jslp.16022.ant>
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder.

- Behavior Research Methods*, 52(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2019). An acoustic analysis of American English liquids by adults and children: Native English speakers and native Japanese speakers of English. *The Journal of the Acoustical Society of America*, 146(4), 2671–2681. <https://doi.org/10.1121/1.5130574>
- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, 32(2), 233–250. [https://doi.org/10.1016/S0095-4470\(03\)00036-6](https://doi.org/10.1016/S0095-4470(03)00036-6)
- Aoyama, K., Guion, S. G., Flege, J. E., Yamada, T., & Akahane-Yamada, R. (2008). The first years in an L2-speaking environment: A comparison of Japanese children and adults learning American English. *International Review of Applied Linguistics in Language Teaching*, 46(1), 61–90. <https://doi.org/10.1515/IRAL.2008.003>
- Aoyama, K., Hong, L., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2023). Relationships between acoustic characteristics and intelligibility scores: A re-analysis of Japanese speakers' productions of American English liquids. *Language and Speech*, 1030–1045. <https://doi.org/10.1177/00238309221140910>
- Arai, T. (2013). On why Japanese /r/ sounds are difficult for children to acquire. *Interspeech 2013*, 2445–2449. <https://doi.org/10.21437/Interspeech.2013-568>
- Archibald, J. (2021). Ease and difficulty in L2 phonology: A mini-review. *Frontiers in Communication*, 6, 1–7. Retrieved September 6, 2022, from <https://doi.org/10.3389/fcomm.2021.626529>
- Articulate Instruments. (2008). *Ultrasound stabilisation headset: Users manual revision 1.5*. Edinburgh.
- Articulate Instruments. (2019). Articulate Assistant Advanced version 218 [computer software]. <https://www.articulateinstruments.com/aaa/>

- Articulate Instruments. (2022). Articulate Assistant Advanced version 220 [computer software]. <https://www.articulateinstruments.com/aaa/>
- Asano, Y., & Gubian, M. (2018). “Excuse meeee!”: (Mis)coordination of lexical and paralinguistic prosody in L2 hyperarticulation. *Speech Communication, 99*, 183–200. <https://doi.org/10.1016/j.specom.2017.12.011>
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511801686>
- Barlow, J. A., Branson, P. E., & Nip, I. S. B. (2013). Phonetic equivalence in the acquisition of /l/ by Spanish–English bilingual children. *Bilingualism: Language and Cognition, 16*(1), 68–85. <https://doi.org/10.1017/S1366728912000235>
- Barreda, S. (2021). Fast Track: Fast (nearly) automatic formant-tracking using Praat. *Linguistics Vanguard, 7*(1). <https://doi.org/10.1515/lingvan-2020-0051>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Beristain, A. M. (2022). *The acquisition of acoustic and aerodynamic patterns of coarticulation in second and heritage languages* (Doctoral dissertation). University of Illinois Urbana-Champaign. Illinois, United States.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 171–204). York Press.
- Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*(3), 305–330. [https://doi.org/10.1016/S0095-4470\(19\)30637-0](https://doi.org/10.1016/S0095-4470(19)30637-0)
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J.

- Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins Publishing Company. <https://doi.org/10.1075/llt.17.07bes>
- Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer. Retrieved February 21, 2022, from <https://www.fon.hum.uva.nl/praat/>
- Borchers, H. W. (2023). Pracma: Practical numerical math functions. <https://doi.org/10.32614/CRAN.package.pracma>
- Bradlow, A. R. (2008). Training non-native language sound patterns. In J. G. H. Edwards & M. Zampini L (Eds.), *Phonology and second language acquisition* (pp. 287–308). John Benjamins Publishing Company.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61(5), 977–985. <https://doi.org/10.3758/BF03206911>
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of Acoustical Society of America*, 101(4), 2299–2310. <https://doi.org/10.1121/1.418276>
- Brekelmans, G., Lavan, N., Saito, H., Clayards, M., & Wonnacott, E. (2022). Does high variability training improve the learning of non-native phoneme contrasts over low variability training? A replication. *Journal of Memory and Language*, 126, 104352. <https://doi.org/10.1016/j.jml.2022.104352>
- Browman, C. P., & Goldstein, L. (1995). Dynamics and Articulatory Phonology. In R. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 175–193). The MIT Press.
- Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219–252. <https://doi.org/10.1017/S0952675700000658>
- Bryfonski, L. (2023). Is seeing believing?: The role of ultrasound tongue imaging and oral corrective feedback in L2 pronunciation development. *Journal of Second*

- Language Pronunciation*, 9(1), 103–129. <https://doi.org/10.1075/jslp.22051.bry>
- Bürkner, P.-C. (2017). Brms : An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1). <https://doi.org/10.18637/jss.v080.i01>
- Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage*, 19(1), 113–124. [https://doi.org/10.1016/S1053-8119\(03\)00020-X](https://doi.org/10.1016/S1053-8119(03)00020-X)
- Campbell, F., Gick, B., Wilson, I., & Vatikiotis-Bateson, E. (2010). Spatial and temporal properties of gestures in North American English /r/. *Language and Speech*, 53(1), 49–69. <https://doi.org/10.1177/0023830909351209>
- Carter, P. (2002). *Structured variation in British English liquids: The role of resonance* (Doctoral dissertation). University of York, York, United Kingdom.
- Carter, P., & Local, J. (2007). F2 variation in Newcastle and Leeds English liquid systems. *Journal of the International Phonetic Association*, 37(2), 183–199. <https://doi.org/10.1017/S0025100307002939>
- Catford, J. C. (1988). *A practical introduction to phonetics*. Clarendon Press.
- Chang, C. B. (2019). The phonetics of second language learning and bilingualism. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge handbook of phonetics* (1st ed., pp. 427–447). Routledge. <https://doi.org/10.4324/9780429056253-16>
- Chen, S., Whalen, D. H., & Mok, P. P. K. (2024). Production of the English /ɹ/ by Mandarin–English bilingual speakers. *Language and Speech*, 1–38. <https://doi.org/10.1177/00238309241230895>
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row Publishers.

- Chung, H., & Kim, Y. (2021). Acoustic characteristics of Korean-English bilingual speakers' /l/ and the relationship to their foreign accent ratings. *Journal of Communication Disorders*, *94*, 106157. <https://doi.org/10.1016/j.jcomdis.2021.106157>
- Colantoni, L., Kochetov, A., & Steele, J. (2021). Articulatory settings and L2 English coronal consonants. *Phonetica*, *78*(4), 273–316. <https://doi.org/10.1515/phon-2021-2007>
- Colantoni, L., Kochetov, A., & Steele, J. (2023a). Articulatory insights into the L2 acquisition of english-/l/ allophony. *Language and Speech*, 1–33. <https://doi.org/10.1177/00238309231200629>
- Colantoni, L., Kochetov, A., & Steele, J. (2023b). L1 influence on the L2 acquisition of English word-final nasal place contrasts: An electropalatographic study of L1 Japanese and Spanish learners. *Journal of the Association for Laboratory Phonology*, *14*(1), 1–45. <https://doi.org/10.16995/labphon.6434>
- Colantoni, L., & Steele, J. (2008). Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics*, *29*(3), 489–534. <https://doi.org/10.1017/S0142716408080223>
- Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech: Theory and practice*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139087636>
- Collins, B., Mees, I., & Carley, P. (2013). *Practical phonetics and phonology: A resource book for students* (3rd ed.). Routledge. <https://doi.org/10.4324/9780203080023>
- Coretta, S. (2020). *Vowel duration and consonant voicing: A production study* (Doctoral dissertation). University of Manchester. Manchester, United Kingdom. <https://research.manchester.ac.uk/en/studentTheses/vowel-duration-and-consonant-voicing-a-production-study>
- Coretta, S. (2021). Rticulate: Ultrasound Tongue Imaging in R. <https://doi.org/10.5281/zenodo.7048602>

- Cronenberg, J., Gubian, M., Harrington, J., & Ruch, H. (2020). A dynamic model of the change from pre- to post-aspiration in Andalusian Spanish. *Journal of Phonetics*, 83, 101016. <https://doi.org/10.1016/j.wocn.2020.101016>
- Daigaku Eigo Kyoiku Gakkai Kihongo Kaitei Tokubetsu Inkaï. (2016). *Daigaku eigo kyōiku gakkai kihongo risuto: Shin JACET 8000 [The Japan Association for College English Teachers basic word list: New JACET 8000]*. Kirihara Shoten.
- Davidson, L. (2005). Addressing phonological questions with ultrasound. *Clinical Linguistics & Phonetics*, 19(6-7), 619–633. <https://doi.org/10.1080/02699200500114077>
- Davidson, L. (2006a). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America*, 120(1), 407–415. <https://doi.org/10.1121/1.2205133>
- Davidson, L. (2006b). Phonotactics and articulatory coordination interact in phonology: Evidence from nonnative production. *Cognitive Science*, 30(5), 837–862. https://doi.org/10.1207/s15516709cog0000_73
- Davidson, L. (2011). Phonetic and phonological factors in the second language production of phonemes and phonotactics. *Language and Linguistics Compass*, 5(3), 126–139. <https://doi.org/10.1111/j.1749-818X.2010.00266.x>
- Delattre, P., & Freeman, D. C. (1968). A dialect study of American R's by x-ray motion picture. *Linguistics*, 6(44), 29–68. <https://doi.org/10.1515/ling.1968.6.44.29>
- Derrick, D. (2011). *Kinematic patterning of flaps, taps and rhotics in English* (Doctoral dissertation). University of British Columbia. Vancouver, British Columbia, Canada. <https://dx.doi.org/10.14288/1.0072056>
- Derrick, D., Carignan, C., Chen, W.-r., Shujau, M., & Best, C. T. (2018). Three-dimensional printable ultrasound transducer stabilization system. *The Journal of the Acoustical Society of America*, 144(5), EL392–EL398. <https://doi.org/10.1121/1.5066350>

- Derrick, D., & Gick, B. (2011). Individual variation in English flaps and taps: A case of categorical phonetics. *The Canadian Journal of Linguistics / La revue canadienne de linguistique*, 56(3), 307–319. <https://doi.org/10.1353/cjl.2011.0024>
- Escudero, P. (2000). The perception of English vowel contrasts: Acoustic cue reliance in the development of new contrasts. In A. James & J. Leather (Eds.), *Proceedings of the 4th international symposium on the acquisition of second-language speech, New Sounds* (pp. 122–131). <https://www.fon.hum.uva.nl/paul/p2/papers/NewSoundsproc.pdf>
- Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* (Doctoral dissertation). Utrecht University. Utrecht, the Netherlands.
- Esling, J. H., & Wong, R. F. (1983). Voice quality settings and the teaching of pronunciation. *TESOL Quarterly*, 17(1), 89–95. <https://doi.org/10.2307/3586426>
- Espinal, A., Thompson, A., & Kim, Y. (2020). Acoustic characteristics of American English liquids /ɹ/, /l/, /ɹl/ produced by Korean L2 adults. *The Journal of the Acoustical Society of America*, 148(2), EL179–EL184. <https://doi.org/10.1121/10.0001758>
- Espy-Wilson, C. Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English. *The Journal of the Acoustical Society of America*, 92(2), 736–757. <https://doi.org/10.1121/1.403998>
- Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., & Alwan, A. (2000). Acoustic modeling of American English /r/. *The Journal of the Acoustical Society of America*, 108(1), 343–356. <https://doi.org/10.1121/1.429469>
- Flege, J. E. (1986). Effects of equivalence classification on the production of foreign language speech sounds. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 9–40). De Gruyter. <https://doi.org/10.1515/9783110878486-003>

- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15(1), 47–65. [https://doi.org/10.1016/S0095-4470\(19\)30537-6](https://doi.org/10.1016/S0095-4470(19)30537-6)
- Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America*, 89(1), 395–411. <https://doi.org/10.1121/1.400473>
- Flege, J. E. (1992). The intelligibility of English vowels spoken by British and Dutch talkers. In R. D. Kent (Ed.), *Studies in speech pathology and clinical linguistics* (pp. 157–232). John Benjamins Publishing Company. <https://doi.org/10.1075/sspcl.1.06fle>
- Flege, J. E. (1995). Second language speech learning theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.
- Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (Eds.), *Laboratory phonology 9* (pp. 353–381). De Gruyter Mouton.
- Flege, J. E., Aoyama, K., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r) applied. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (1st ed., pp. 84–118). Cambridge University Press. <https://doi.org/10.1017/9781108886901.003>
- Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (1st ed., pp. 3–83). Cambridge University Press. <https://doi.org/10.1017/9781108886901.002>
- Flege, J. E., Fletcher, S. G., McCutcheon, M. J., & Smith, S. C. (1986). The physiological specification of American English vowels. *Language and Speech*, 29(4), 361–388. <https://doi.org/10.1177/002383098602900404>

- Flege, J. E., Mackay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, 23(4), 567–598. <https://doi.org/10.1017/S0142716402004046>
- Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce english /ɹ/ and /l/ Accurately. *Language and Speech*, 38(1), 25–55. <https://doi.org/10.1177/002383099503800102>
- Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *The Journal of the Acoustical Society of America*, 99(2), 1161–1173. <https://doi.org/10.1121/1.414884>
- Fowler, C. A. (2015). The segment in Articulatory Phonology. In E. Raimy & C. E. Cairns (Eds.), *The segment in phonetics and phonology* (pp. 23–43). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118555491.ch2>
- Fraley, C., Raftery, A. E., Scrucca, L., Murphy, T. B., & Fop, M. (2023). Mclust: Gaussian mixture modelling for model-based clustering, classification, and density estimation. <https://doi.org/10.32614/CRAN.package.mclust>
- Franke, M., & Roettger, T. B. (2019). Bayesian regression modeling (for factorial designs): A tutorial. *PsyArXiv*. <https://doi.org/10.31234/osf.io/cdxv3>
- Fukaya, T., & Byrd, D. (2005). An articulatory examination of word-final flapping at phrase edges and interiors. *Journal of the International Phonetic Association*, 35, 45–58. <https://doi.org/10.1017/S0025100305001891>
- Gick, B. (1999). A gesture-based account of intrusive consonants in English. *Phonology*, 16(1), 29–54. <https://doi.org/10.1017/S0952675799003693>
- Gick, B., Allen, B., Roewer-Després, F., & Stavness, I. (2017). Speaking tongues are actively braced. *Journal of Speech, Language, and Hearing Research*, 60(3), 494–506. https://doi.org/10.1044/2016_JSLHR-S-15-0141
- Gick, B., Bacsfalvi, P., Bernhardt, B. M., Oh, S., Stolar, S., & Wilson, I. (2007). A motor differentiation model for liquid substitutions in children's speech.

- Proceedings of the Meeting Acoustics, 1*, 060003. <https://doi.org/10.1121/1.2951481>
- Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. G. Hansen Edwards & M. Zampini (Eds.), *Studies in bilingualism* (pp. 309–322). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.36.15gic>
- Gick, B., & Campbell, F. (2003). Intergestural Timing in English /r/. In M.-J. Solé, D. Recasens, & J. G. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 1911–1914). https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/p15_1911.html
- Gick, B., Campbell, F., Oh, S., & Tamburri-Watt, L. (2006). Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*, *34*(1), 49–72. <https://doi.org/10.1016/j.wocn.2005.03.005>
- Gordon, P. C., Keyes, L., & Yung, Y.-F. (2001). Ability in perceiving nonnative contrasts: Performance on natural and synthetic speech stimuli. *Perception & Psychophysics*, *63*(4), 746–758. <https://doi.org/10.3758/BF03194435>
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia*, *9*(3), 317–323. [https://doi.org/10.1016/0028-3932\(71\)90027-3](https://doi.org/10.1016/0028-3932(71)90027-3)
- Grosjean, F. (2008). *Studying bilinguals*. Oxford University Press.
- Gubian, M., Torreira, F., & Boves, L. (2015). Using functional data analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics*, *49*, 16–40. <https://doi.org/10.1016/j.wocn.2014.10.001>
- Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., & Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *The Journal of the Acoustical Society of America*, *105*(5), 2854–2865. <https://doi.org/10.1121/1.426900>

- Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, *107*(5), 2711–2724. <https://doi.org/10.1121/1.428657>
- Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, *21*(2), 205–228. <https://doi.org/10.1017/S0142716400002034>
- Hardcastle, W., & Barry, W. (1989). Articulatory and perceptual factors in /l/ vocalisations in English. *Journal of the International Phonetic Association*, *15*(2), 3–17. <https://doi.org/10.1017/S0025100300002930>
- Harper, S., Goldstein, L., & Narayanan, S. (2020). Variability in individual constriction contributions to third formant values in American English /ɹ/. *The Journal of the Acoustical Society of America*, *147*(6), 3905–3916. <https://doi.org/10.1121/10.0001413>
- Harper, S., Goldstein, L., & Narayanan, S. S. (2016). L2 acquisition and production of the English rhotic pharyngeal gesture. *Interspeech 2016*, 208–212. <https://doi.org/10.21437/Interspeech.2016-658>
- Hashi, M., Honda, K., & Westbury, J. R. (2003). Time-varying acoustic and articulatory characteristics of American English [ɹ]: A cross-speaker study. *Journal of Phonetics*, *31*(1), 3–22. [https://doi.org/10.1016/S0095-4470\(02\)00062-1](https://doi.org/10.1016/S0095-4470(02)00062-1)
- Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America*, *125*(1), 469–479. <https://doi.org/10.1121/1.3021295>
- Hattori, K., & Iverson, P. (2011). Examination of the relationship between L2 perception and production: An investigation of English /r/-/l/ perception

- and production by adult Japanese speakers. In M. Nakano (Ed.), *Interspeech workshop on second language studies: Acquisition, learning, education and technology* (pp. 2–4). https://www.isca-archive.org/l2ws_2010/hattori10_l2ws.pdf
- Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, & J. L. M. Trim (Eds.), *In honour of Daniel Jones* (pp. 73–84). Longman.
- Hoole, P., & Pouplier, M. (2017). Öhman returns: New horizons in the collection and analysis of imaging data in speech production research. *Computer Speech & Language*, *45*, 253–277. <https://doi.org/10.1016/j.csl.2017.03.002>
- Howson, P. J., & Redford, M. A. (2021). The acquisition of articulatory timing for liquids: Evidence from child and adult speech. *Journal of Speech, Language, and Hearing Research*, *64*(3), 734–753. https://doi.org/10.1044/2020_JSLHR-20-00391
- Hwang, Y., Lulich, S. M., & de Jong, K. J. (2019). Articulatory and acoustic characteristics of the Korean and English word-final laterals produced by Korean female learners of American English. *The Journal of the Acoustical Society of America*, *146*(5), EL444–EL450. <https://doi.org/10.1121/1.5134656>
- Ingvallson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate /r-ɹ/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition*, *15*(2), 255–274. <https://doi.org/10.1017/S1366728911000447>
- Iskarous, K., & Pouplier, M. (2022). Advancements of phonetics in the 21st century: A critical appraisal of time and space in Articulatory Phonology. *Journal of Phonetics*, *95*, 101195. <https://doi.org/10.1016/j.wocn.2022.101195>
- Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r-/ɹ/ to

- Japanese adults. *The Journal of the Acoustical Society of America*, 118(5), 3267–3278. <https://doi.org/10.1121/1.2062307>
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1), B47–B57. [https://doi.org/10.1016/S0010-0277\(02\)00198-1](https://doi.org/10.1016/S0010-0277(02)00198-1)
- Jochim, M., Winkelmann, R., Jaensch, K., Cassidy, S., & Harrington, J. (2023). emuR - Main package of the EMU Speech Database Management System version 2.4.2. <https://CRAN.R-project.org/package=emuR>
- Johnson, K. (2008). *Quantitative methods in linguistics*. Wiley-Blackwell.
- Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, 94(2), 701–714. <https://doi.org/10.1121/1.406887>
- Kallioinen, N., Paananen, T., Bürkner, P.-C., & Vehtari, A. (2023). *Detecting and diagnosing prior and likelihood sensitivity with power-scaling*. <https://doi.org/10.48550/arXiv.2107.14054>
- Katz, W. F., Mehta, S., & Wood, M. (2018). Effects of syllable position and vowel context on Japanese /r/: Kinematic and perceptual data. *Acoustical Science and Technology*, 39(2), 130–137. <https://doi.org/10.1250/ast.39.130>
- Kawahara, S., & Matsui, M. F. (2017). Some aspects of Japanese consonant articulation: A preliminary EPG study. *ICU Working Papers in Linguistics (ICUWPL)*, 2, 9–20. http://user.keio.ac.jp/~kawahara/pdf/ICUWP_KawaMatsui.pdf
- Keating, P. A. (1985). Universal phonetics and the organization of grammars. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 115–132). Academic Press.
- King, H., & Ferragne, E. (2020). Loose lips and tongue tips: The central role of the /r/-typical labial gesture in Anglo-English. *Journal of Phonetics*, 80, 100978. <https://doi.org/10.1016/j.wocn.2020.100978>

- Kirkham, S. (2017). Ethnicity and phonetic variation in Sheffield English liquids. *Journal of the International Phonetic Association*, 47(1), 17–35. <https://doi.org/10.1017/S0025100316000268>
- Kirkham, S. (2024). TadaR: R interface to Task Dynamic Application (v 1.0.0). <https://doi.org/10.5281/zenodo.13329512>
- Kirkham, S., & McCarthy, K. M. (2021). Acquiring allophonic structure and phonetic detail in a bilingual community: The production of laterals by Sylheti-English bilingual children. *International Journal of Bilingualism*, 25(3), 531–547. <https://doi.org/10.1177/1367006920947180>
- Kirkham, S., & Nance, C. (2017). An acoustic-articulatory study of bilingual vowel production: Advanced tongue root vowels in Twi and tense/lax vowels in Ghanaian English. *Journal of Phonetics*, 62, 65–81. <https://doi.org/10.1016/j.wocn.2017.03.004>
- Kirkham, S., & Nance, C. (2022). Diachronic phonological asymmetries and the variable stability of synchronic contrast. *Journal of Phonetics*, 94, 101176. <https://doi.org/10.1016/j.wocn.2022.101176>
- Kirkham, S., Nance, C., Littlewood, B., Lightfoot, K., & Groarke, E. (2019). Dialect variation in formant dynamics: The acoustics of lateral and vowel sequences in Manchester and Liverpool English. *The Journal of the Acoustical Society of America*, 145(2), 784–794. <https://doi.org/10.1121/1.5089886>
- Kirkham, S., Strycharczuk, P., Gorman, E., Nagamine, T., & Wrench, A. (2023). Co-registration of simultaneous high-speed ultrasound and electromagnetic articulography for speech production research. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 942–946). Guarant International. <https://guarant.cz/icphs2023/145.pdf>
- Kirkham, S., Turton, D., & Leemann, A. (2020). A typology of laterals in twelve English dialects. *The Journal of the Acoustical Society of America*, 148(1), EL72–EL76. <https://doi.org/10.1121/10.0001587>

- Klein, H. B., McAllister Byun, T., Davidson, L., & Grigos, M. I. (2013). A multidimensional investigation of children's /r/ productions: Perceptual, ultrasound, and acoustic measures. *American Journal of Speech-Language Pathology*, 22(3), 540–553. [https://doi.org/10.1044/1058-0360\(2013/12-0137\)](https://doi.org/10.1044/1058-0360(2013/12-0137))
- Kochetov, A. (2018). Linguopalatal contact contrasts in the production of Japanese consonants: Electropalatographic data from five speakers. *Acoustical Science and Technology*, 39(2), 84–91. <https://doi.org/10.1250/ast.39.84>
- Kochetov, A. (2020). Research methods in articulatory phonetics I: Introduction and studying oral gestures. *Language and Linguistics Compass*, 14(4), 1–29. <https://doi.org/10.1111/lnc3.12368>
- Kochetov, A. (2022, March 7–8). *Production of English phonemic contrasts and allophony by Japanese learners: Electropalatographic evidence* [Keynote address]. Phonology Festa 17, Tokyo, Japan.
- Krakow, R. A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics*, 27(1), 23–54. <https://doi.org/10.1006/jpho.1999.0089>
- Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Elsevier. <https://doi.org/10.1016/B978-0-12-405888-0.09999-2>
- Kubozono, H. (2015). Loanword phonology. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 313–362). DE GRUYTER. <https://doi.org/10.1515/9781614511984.313>
- Ladefoged, P., & Ferrari Disner, S. (2012). *Vowels and consonants* (3rd ed.). Wiley-Blackwell.
- Ladefoged, P., & Johnson, K. (2010). *A course in phonetics, international edition* (6th edition). Wadworth.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell Publishers.

- Laver, J. (1994). *Principles of phonetics*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139166621>
- Lawson, E., & Stuart-Smith, J. (2019). The effects of syllable and sentential position on the timing of lingual gestures in /l/ and /r/. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 547–551). Australasian Speech Science; Technology Association Inc.
- Lawson, E., Stuart-Smith, J., & Rodger, L. (2019). A comparison of acoustic and articulatory parameters for the GOOSE vowel across British Isles Englishes. *The Journal of the Acoustical Society of America*, 146(6), 4363–4381. <https://doi.org/10.1121/1.5139215>
- Lawson, E., Stuart-Smith, J., Scobbie, J., Yaeger-Dror, M., & Maclagan, M. (2011). Liquids. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A student's guide* (pp. 72–86). Routledge.
- Lee-Kim, S.-I., Davidson, L., & Hwang, S. (2013). Morphological effects on the darkness of English intervocalic /l/. *Journal of the Association for Laboratory Phonology*, 4(2). <https://doi.org/10.1515/lp-2013-0015>
- Léger, A., King, H., & Ferragne, E. (2023). Is rhoticity on the tip of your tongue? Tongue shapes for English /r/ in French learners with ultrasound. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 2741–2745). Guarant International. <https://guarant.cz/icphs2023/1011.pdf>
- Lenneberg, E., H. (1967). *Biological foundations of language*. Wiley.
- Lenth, R. V., Buerkner, P., Herve, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2022). *Emmeans: Estimated marginal means, aka least-squares means*. Retrieved March 19, 2022, from <https://rvlenth.github.io/emmeans/>
- Li, M., Kambhamettu, C., & Stone, M. (2005). Automatic contour tracking in ultrasound images. *Clinical Linguistics & Phonetics*, 19(6-7), 545–554. <https://doi.org/10.1080/02699200500113616>

- Li, N. H., & Juffs, A. (2014). The influence of moraic structure on L2 English syllable-final consonants. In A. Albright & M. A. Fullwood (Eds.), *Proceedings of the annual meetings on phonology* (pp. 1–12). <https://doi.org/10.3765/amp.v2i0.3767>
- Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods*, *38*(2), 202–210. <https://doi.org/10.3758/bf03192770>
- Lindau, M. (1985). The story of /r/. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 157–68). Academic Press.
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242–1255. <https://doi.org/10.1121/1.408177>
- Llompарт, M., Eger, N. A., & Reinisch, E. (2021). Free allophonic variation in native and second language spoken word recognition: The case of the German rhotic. *Frontiers in Psychology*, *12*, 1–12. <https://doi.org/10.3389/fpsyg.2021.711230>
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886. <https://doi.org/10.1121/1.1894649>
- Macdonald, R., & Stuart-Smith, J. (2024). Coarticulation guides sound change: An acoustic-phonetic study of real-time change in word-initial /l/ over four decades of Glaswegian. In F. Kleber & T. Rathcke (Eds.), *Speech dynamics: Synchronic variation and diachronic change*. De Gruyter Mouton.
- Mackenzie, S., Olson, E., Clayards, M., & Wagner, M. (2018). North American /l/ both darkens and lightens depending on morphological constituency and segmental context. *Laboratory Phonology*, *9*(1). <https://doi.org/10.5334/labphon.104>

- Maekawa, K. (2019). Nihongo ragyo shiin no choon: Riaru taimu MRI ni yoru kansatsu [Articulation of Japanese /r/: A real-time MRI study]. *Proceedings of the 33rd National Conference of Phonetic Society of Japan*, 98–103.
- Maekawa, K. (2023). Articulatory characteristics of the Japanese /r/: A real-time MRI study. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 992–996). Guarant International. <https://guarant.cz/icphs2023/443.pdf>
- Makino, T. (2009). Vowel substitution patterns in Japanese speakers' English. In B. Čubrović & T. Paunovic (Eds.), *Ta(l)king English phonetics across frontiers* (pp. 19–32). Cambridge Scholars.
- Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of Effect Existence and Significance in the Bayesian Framework. *Frontiers in Psychology*, 10, 1–14. <https://doi.org/10.3389/fpsyg.2019.02767>
- Manuel, S. (1999). Cross-language studies: Relating language-particular coarticulation patterns to other language-particular facts. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation* (1st ed., pp. 179–198). Cambridge University Press. <https://doi.org/10.1017/CBO9780511486395.009>
- Masaki, S., Akahane-Yamada, R., Tiede, M., Shimada, Y., & Fujimoto, I. (1996). An MRI-based analysis of the English /r/ and /l/ articulations. In H. T. Bunnell & W. Idsardi (Eds.), *Proceedings of Fourth International Conference on Spoken Language Processing*. (pp. 1581–1584). <https://doi.org/10.1109/ICSLP.1996.607922>
- Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9), 1281–1289. <https://doi.org/10.1038/s41593-018-0209-y>
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using kald. *Interspeech 2017*, 498–502. <https://doi.org/10.21437/Interspeech.2017-1386>

- McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and STAN* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9780429029608>
- Mennen, I., Scobbie, J., de Leeuw, E., Schaeffler, S., & Schaeffler, F. (2010). Measuring language-specific phonetic settings. *Second Language Research*, 26(1), 13–41. <https://doi.org/10.1177/0267658309337617>
- Mielke, J., Baker, A., & Archangeli, D. (2016). Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɾ/. *Language*, 92(1), 101–140. <https://doi.org/10.1353/lan.2016.0019>
- Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 18(5), 331–340. <https://doi.org/10.3758/BF03211209>
- Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics*, 9(3), 283–303. [https://doi.org/10.1016/S0095-4470\(19\)30972-6](https://doi.org/10.1016/S0095-4470(19)30972-6)
- Mokhtari, P., Kitamura, T., Takemoto, H., & Honda, K. (2007). Principal components of vocal-tract area functions and inversion of vowels by linear regression of cepstrum coefficients. *Journal of Phonetics*, 35(1), 20–39. <https://doi.org/10.1016/j.wocn.2006.01.001>
- Moore, J., Shaw, J., Kawahara, S., & Arai, T. (2018). Articulation strategies for English liquids used by Japanese speakers. *Acoustical Science and Technology*, 39(2), 75–83. <https://doi.org/10.1250/ast.39.75>
- Morimoto, M. (2020). *Geminated liquids in Japanese: A production study* (Doctoral dissertation). University of California Santa Cruz, California, United States.
- Morimoto, M. (2021). Articulatory preference in Japanese liquids and F3 in English: A preliminary report. *ICU Working Papers in Linguistics: Selected Papers from the 5th Asian Junior Linguists Conference (AJL5)*, 15, 1–6. <https://doi.org/10.34577/00004829>

- Munro, M. J., & Derwing, T. M. (2015). A prospectus for pronunciation research in the 21st century: A point of view. *Journal of Second Language Pronunciation*, 1(1), 11–42. <https://doi.org/10.1075/jslp.1.1.01mun>
- Nagamine, T. (2022). Acquisition of allophonic variation in second language speech: An acoustic and articulatory study of English laterals by Japanese speakers. *Interspeech 2022*, 644–648. <https://doi.org/10.21437/Interspeech.2022-11020>
- Nagamine, T. (2023). Dynamic tongue movements in L1 Japanese and L2 English liquids. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 2442–2446). Guarant International. <https://guarant.cz/icphs2023/198.pdf>
- Nagamine, T. (2024a, May 13–17). *Acquisition of articulatory dynamics in second language speech: Japanese speakers' production of English and Japanese liquids* [Poster presentation]. The 13th International Seminar of Speech Production, Autrans, France.
- Nagamine, T. (2024b). Formant dynamics in second language speech: Japanese speakers' production of English liquids. *The Journal of the Acoustical Society of America*, 155(1), 479–495. <https://doi.org/10.1121/10.0024351>
- Nagamine, T. (in prep.). L1 Japanese speakers use a single articulatory strategy to produce onset-coda allophony in L2 English liquids. *To be submitted to Language and Speech*.
- Nagamine, T. (In revision). Learning to resist: Japanese speakers' production of liquid-vowel coarticulation in L2 English. *Submitted to the Journal of Phonetics*.
- Nagamine, T. (revised & resubmitted). Quantifying between-speaker variation in ultrasound tongue imaging data. *Journal of Phonetic Society of Japan*.
- Nagle, C. L., & Baese-Berk, M. M. (2022). Advancing the state of the art in L2 speech perception-production research: Revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition*, 44(2), 1–26. <https://doi.org/10.1017/S0272263121000371>

- Nakamura, M. (2001). *Articulatory organisation in Japanese: An EPG study* (Doctoral dissertation). University College London. London, United Kingdom.
- Nance, C. (2014). Phonetic variation in Scottish Gaelic laterals. *Journal of Phonetics*, 47, 1–17. <https://doi.org/10.1016/j.wocn.2014.07.005>
- Nance, C., Dewhurst, M., Fairclough, L., Forster, P., Kirkham, S., Nagamine, T., Turton, D., & Wang, D. (2023). Acoustic and articulatory characteristics of rhoticity in the North-West of England. In S. Radek & V. Jan (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 3573–3577). Guarant International. <https://guarant.cz/icphs2023/217.pdf>
- Nance, C., & Kirkham, S. (2022). Phonetic typology and articulatory constraints: The realisation of secondary articulations in Scottish Gaelic rhotics. *Language*, 419–460. <https://dx.doi.org/10.1353/lan.0.0268>
- Nance, C., & Kirkham, S. (2023). Producing a smaller sound system: Acoustics and articulation of the subset scenario in Gaelic–English bilinguals. *Bilingualism: Language and Cognition*, 1–13. <https://doi.org/10.1017/S1366728923000688>
- Narayanan, S. S., Alwan, A. A., & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals. *The Journal of the Acoustical Society of America*, 101(2), 1064–1077. <https://doi.org/10.1121/1.418030>
- Nogita, A. (2016). *L2 letter-sound correspondence: Mapping between English vowel graphemes and phonemes by Japanese EAL learners* (Doctoral dissertation). University of Victoria. Victoria, British Columbia, Canada.
- Oh, E. (2008). Coarticulation in non-native speakers of English and French: An acoustic study. *Journal of Phonetics*, 36(2), 361–384. <https://doi.org/10.1016/j.wocn.2007.12.001>
- Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39(1), 151–168. <https://doi.org/10.1121/1.1909864>

- Olsen, M. K. (2012). The L2 acquisition of Spanish rhotics by L1 English speakers: The effect of L1 articulatory routines and phonetic context for allophonic variation. *Hispania*, 95(1), 65–82. <https://doi.org/10.1353/hpn.2012.a469667>
- Otake, T. (2015). Mora and mora-timing. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 493–524). De Gruyter. <https://doi.org/10.1515/9781614511984.493>
- Palo, P., Schaeffler, S., & Scobbie, J. (2014). Pre-speech tongue movements recorded with ultrasound. In S. Fuchs, M. Grice, A. Hermes, L. Lancia, & D. Mücke (Eds.), *Proceedings of the 10th international seminar on speech production* (pp. 300–303). <http://www.issp2014.uni-koeln.de/>
- Plug, L., & Ogden, R. (2003). A parametric approach to the phonetics of postvocalic /r/ in dutch. *Phonetica*, 60(3), 159–186. <https://doi.org/10.1159/000073501>
- Potts, A., & Baker, P. (2012). Does semantic tagging identify cultural change in British and American English? *International Journal of Corpus Linguistics*, 17(3), 295–324. <https://doi.org/10.1075/ijcl.17.3.01pot>
- Preston, J. L., McAllister Byun, T., Boyce, S. E., Hamilton, S., Tiede, M., Phillips, E., Rivera-Campos, A., & Whalen, D. H. (2017). Ultrasound images of the tongue: A tutorial for assessment and remediation of speech sound errors. *Journal of Visualized Experiments*, (119), 55123. <https://doi.org/10.3791/55123>
- Proctor, M. (2011). Towards a gestural characterization of liquids: Evidence from Spanish and Russian. *Journal of the Association for Laboratory Phonology*, 2(2), 451–485. <https://doi.org/10.1515/labphon.2011.017>
- Proctor, M. (2021). Consonants. In J. Setter & R.-A. Knight (Eds.), *The Cambridge handbook of phonetics* (pp. 65–105). Cambridge University Press. <https://doi.org/10.1017/9781108644198.004>
- Proctor, M., Walker, R., Smith, C., Szalay, T., Goldstein, L., & Narayanan, S. (2019). Articulatory characterization of English liquid-final rimes. *Journal of Phonetics*, 77, 100921. <https://doi.org/10.1016/j.wocn.2019.100921>

- Pucher, M., Klingler, N., Luttenberger, J., & Spreafico, L. (2020). Accuracy, recording interference, and articulatory quality of headsets for ultrasound recordings. *Speech Communication*, 123, 83–97. <https://doi.org/10.1016/j.specom.2020.07.001>
- R Core Team. (2021). R: A Language and Environment for Statistical Computing version 4.1.2 [computer software]. <https://www.R-project.org/>
- R Core Team. (2022). R: A Language and Environment for Statistical Computing version 4.2.2 [computer software]. <https://www.R-project.org/>
- R Core Team. (2023). R: A Language and Environment for Statistical Computing version 4.3.2 [computer software]. <https://www.R-project.org/>
- Ramanarayanan, V., Lammert, A., Goldstein, L., & Narayanan, S. (2014). Are articulatory settings mechanically advantageous for speech motor control? *PLoS ONE*, 9(8), e104168. <https://doi.org/10.1371/journal.pone.0104168>
- Ramsay, J., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and MATLAB*. Springer Science & Business Media.
- Rebernik, T., Jacobi, J., Jonkers, R., Noiray, A., & Wieling, M. (2021). A review of data collection practices using electromagnetic articulography. *Journal of the Association for Laboratory Phonology*, 12(1). <https://doi.org/10.5334/labphon.237>
- Recasens, D. (1991). On the production characteristics of apicoalveolar taps and trills. *Journal of Phonetics*, 19(3-4), 267–280. [https://doi.org/10.1016/S0095-4470\(19\)30344-4](https://doi.org/10.1016/S0095-4470(19)30344-4)
- Recasens, D. (1996). An articulatory-perceptual account of vocalization and elision of dark /l/ in the Romance languages. *Language and Speech*, 39(1), 63–89. <https://doi.org/10.1177/002383099603900104>
- Recasens, D. (2011). Differences in base of articulation for consonants among Catalan dialects. *Phonetica*, 67(4), 201–218. <https://doi.org/10.1159/000322312>

- Recasens, D. (2012). A cross-language acoustic study of initial and final allophones of /l/. *Speech Communication*, 54(3), 368–383. <https://doi.org/10.1016/j.specom.2011.10.001>
- Recasens, D., & Espinosa, A. (2005). Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /l/: Evidence from two Catalan dialects. *Journal of the International Phonetic Association*, 35(1), 1–25. <https://doi.org/10.1017/S0025100305001878>
- Recasens, D., & Espinosa, A. (2007). Phonetic typology and positional allophones for alveolar rhotics in Catalan. *Phonetica*, 64(1), 1–28. <https://doi.org/10.1159/000100059>
- Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *The Journal of the Acoustical Society of America*, 125(4), 2288–2298. <https://doi.org/10.1121/1.3089222>
- Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *The Journal of the Acoustical Society of America*, 102(1), 544–561. <https://doi.org/10.1121/1.419727>
- Recasens, D., & Rodríguez, C. (2016). A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound. *Journal of Phonetics*, 59, 58–75. <https://doi.org/10.1016/j.wocn.2016.09.002>
- Recasens, D., & Rodríguez, C. (2017). Lingual articulation and coarticulation for Catalan consonants and vowels: An ultrasound study. *Phonetica*, 74(3), 125–156. <https://doi.org/10.1159/000452475>
- Reidy, P. F. (2016). Spectral dynamics of sibilant fricatives are contrastive and language specific. *The Journal of the Acoustical Society of America*, 140(4), 2518–2529. <https://doi.org/10.1121/1.4964510>
- Rimac, R., & Smith, B. L. (1984). Acoustic characteristics of flap productions by American English-speaking children and adults: Implications concerning the

- development of speech motor control. *Journal of Phonetics*, 12(4), 387–396. [https://doi.org/10.1016/S0095-4470\(19\)30898-8](https://doi.org/10.1016/S0095-4470(19)30898-8)
- Riney, T. J., Takada, M., & Ota, M. (2000). Segmentals and global foreign accent: The Japanese flap in EFL. *TESOL Quarterly*, 34(4), 711–737. <https://doi.org/10.2307/3587782>
- Saito, K. (2011). Identifying problematic segmental features to acquire comprehensible pronunciation in EFL settings: The case of Japanese learners of English. *RELC Journal*, 42(3), 363–378. <https://doi.org/10.1177/0033688211420275>
- Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, 55(3), 866–900. <https://doi.org/10.1002/tesq.3027>
- Saito, K., & Munro, M. J. (2014). The early phase of /ɹ/ production development in adult Japanese learners of English. *Language and Speech*, 57(4), 451–469. <https://doi.org/10.1177/0023830913513206>
- Saito, K., & van Poeteren, K. (2018). The perception–production link revisited: The case of Japanese learners’ English /ɹ/ performance. *International Journal of Applied Linguistics*, 28(1), 3–17. <https://doi.org/10.1111/ijal.12175>
- Schwartz, G., & Kaźmierski, K. (2020). Vowel dynamics in the acquisition of L2 English – an acoustic study of L1 Polish learners. *Language Acquisition*, 27(3), 227–254. <https://doi.org/10.1080/10489223.2019.1707204>
- Scobbie, J., Gibbon, F., Hardcastle, W. J., & Fletcher, P. (1996). Covert contrast as a stage in the acquisition of phonetics and phonology: Working paper. *QMC Working Papers in Speech and Language Sciences*, 1, 43–62. <https://eresearch.qmu.ac.uk/handle/20.500.12289/10>
- Scobbie, J., Lawson, E., Cowen, S., Cleland, J., & Wrench, A. (2011). A common co-ordinate system for mid-sagittal articulatory measurement. *QMU CASL Working Papers*, 20, 1–4. <https://eresearch.qmu.ac.uk/handle/20.500.12289/3597>

- Scobbie, J., Stuart-Smith, J., & Lawson, E. (2012). Back to front: A socially-stratified ultrasound tongue imaging study of Scottish English /u/. *Italian Journal of Linguistics/Rivista di Linguistica*, 24(1), 103–148. <http://linguistica.sns.it/RdL/2012.html>
- Sereno, J., Lammers, L., & Jongman, A. (2016). The relative contribution of segments and intonation to the perception of foreign-accented speech. *Applied Psycholinguistics*, 37(2), 303–322. <https://doi.org/10.1017/S0142716414000575>
- Setter, J., & Jenkins, J. (2005). Pronunciation. *Language Teaching*, 38(1), 1–17. <https://doi.org/10.1017/S026144480500251X>
- Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3(3), 243–261. <https://doi.org/10.1017/S0142716400001417>
- Shimizu, K., & Dantsuji, M. (1983). A study on the perception of /r/ and /l/ in natural and synthetic speech sounds. *Studia phonologica*, 17, 1–14. [https://doi.org/10.1016/S0095-4470\(19\)30972-6](https://doi.org/10.1016/S0095-4470(19)30972-6)
- Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/-/l/. *Journal of Phonetics*, 66, 242–251. <https://doi.org/10.1016/j.wocn.2017.11.002>
- Slud, E., Stone, M., Smith, P. J., & Goldstein Jr., M. (2002). Principal components representation of the two-dimensional coronal tongue surface. *Phonetica*, 59(2-3), 108–133. <https://doi.org/10.1159/000066066>
- Solon, M. (2017). Do learners lighten up? Phonetic and allophonic acquisition of Spanish /l/ by English-speaking learners. *Studies in Second Language Acquisition*, 39(4), 801–832. <https://doi.org/10.1017/S0272263116000279>
- Someda, T. (1966). Ei, futsugo to no hikaku ni okeru nihongo no chōon no ippanteki haikai ni tsuite. [General articulatory settings of Japanese in comparison to English and French]. *Onsei no Kenkyu*, 12, 327–346.

- Song, J. Y., & Eckman, F. (2021). Using ultrasound tongue imaging to study covert contrasts in second-language learners' acquisition of English vowels. *Language Acquisition*, 28(4), 344–369. <https://doi.org/10.1080/10489223.2021.1910266>
- Sóskuthy, M. (2017). Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. *arXiv*. <https://doi.org/10.48550/arXiv.1703.05339>
- Sóskuthy, M., Foulkes, P., Hughes, V., & Haddican, B. (2018). Changing words and sounds: The roles of different cognitive units in sound change. *Topics in Cognitive Science*, 10(4), 787–802. <https://doi.org/10.1111/tops.12346>
- Spreafico, L., Pucher, M., & Matosova, A. (2018). UltraFit: A speaker-friendly head-set for ultrasound recordings in speech science. *Interspeech 2018*, 1517–1520. <https://doi.org/10.21437/Interspeech.2018-995>
- Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21(3), 291–311. [https://doi.org/10.1016/S0095-4470\(19\)31340-3](https://doi.org/10.1016/S0095-4470(19)31340-3)
- Stevens, K. N. (2000). *Acoustic phonetics*. The MIT Press.
- Stolar, S., & Gick, B. (2013). An index for quantifying tongue curvature. *Canadian Acoustics*, 41(1), 11–15. <https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2598>
- Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, 19(6-7), 455–501. <https://doi.org/10.1080/02699200500113558>
- Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, 36(2), 131–145. <https://doi.org/10.3758/BF03202673>
- Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. Zampini (Eds.), *Studies in Bilingualism* (pp. 153–191). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.36.09str>

- Strycharczuk, P., Derrick, D., & Shaw, J. (2020). Locating de-lateralization in the pathway of sound changes affecting coda /l/. *Journal of the Association for Laboratory Phonology*, 11(1), 21. <https://doi.org/10.5334/labphon.236>
- Strycharczuk, P., Lloyd, S., & Scobbie, J. (2023). Apparent time change in the articulation of onset rhotics in Southern British English. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 3602–3606). Guarant International. <https://guarant.cz/icphs2023/315.pdf>
- Strycharczuk, P., & Scobbie, J. (2017). Fronting of Southern British English high-back vowels in articulation and acoustics. *The Journal of the Acoustical Society of America*, 142(1), 322–331. <https://doi.org/10.1121/1.4991010>
- Sudo, M., Kiritani, S., & Sawashima, M. (1983). The articulation of Japanese intervocalic /d/ and /r/: An electro-palatographic study. *Annual Bulletin of RILP*, 17, 55–59.
- Sudo, M., Kiritani, S., & Yoshioka, H. (1982). An electro-palatographic study of Japanese intervocalic /r/. *Annual Bulletin of Research Institute of Logopedics and Phoniatrics (RILP)*, 16, 21–25.
- Święciński, R. (2013). An EMA study of articulatory settings in Polish speakers of English. In E. Waniek-Klimczak & L. R. Shockey (Eds.), *Teaching and researching English accents in native and non-native speakers* (pp. 73–82). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24019-5_6
- Takagi, N. (1993). *Perception of American English /r/ and /l/ by adult Japanese learners of English: A unified view* (Doctoral dissertation). University of California, Irvine. California, United States.
- Takebayashi, S., & Saito, H. (1998). *Eigo onseigaku nyumon [Introduction to English phonetics]*. Taishukan Shoten.
- Thomson, R., & Isaacs, T. (2009). Within-category variation in L2 English vowel learning. *Canadian Acoustics*, 138–139. <http://jcaa.caa-aca.ca/index.php/jcaa/article/view/2172/>

- Tiede, M. (2021). GetContours: tongue contour fitting software [Computer program]. <https://github.com/mktiede/GetContours>
- Tiede, M., Boyce, S. E., Holland, C. K., & Choe, K. A. (2004). A new taxonomy of American English /r/ using MRI and ultrasound. *The Journal of the Acoustical Society of America*, 115(5), 2633–2634. <https://doi.org/10.1121/1.4784878>
- Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952–981. <https://doi.org/10.1080/01690960903498424>
- Tsui, H. M.-L. (2012). *Ultrasound speech training for Japanese adults learning English as a second language* (Master's thesis). University of British Columbia. Vancouver, British Columbia, Canada. <https://doi.org/10.14288/1.0073242>
- Turton, D. (2014). *Variation in English /l/: Synchronic reflections of the life cycle of phonological processes* (Doctoral dissertation). University of Manchester. Manchester, United Kingdom. Retrieved March 20, 2022, from https://research.manchester.ac.uk/files/54558782/FULL_TEXT.PDF
- Turton, D. (2017). Categorical or gradient? An ultrasound investigation of /l/-darkening and vocalization in varieties of English. *Journal of the Association for Laboratory Phonology*, 8(1), 1–13. <https://doi.org/10.5334/labphon.35>
- Turton, D. (2023). Sociophonetics and laterals. In *The Routledge handbook of sociophonetics* (1st ed., pp. 214–236). Routledge. <https://doi.org/10.4324/9781003034636-11>
- Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A sound approach to language matters: In honor of Ocke-Schwen Bohn* (pp. 607–630). Aarhus University. <https://doi.org/10.7146/aul.322.218>

- van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology, 6*, Article1000. <https://doi.org/10.3389/fpsyg.2015.01000>
- van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2020). Itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. <https://doi.org/10.32614/CRAN.package.itsadug>
- Vance, T. J. (1987). *An introduction to Japanese phonology*. State University of New York Press.
- Vance, T. J. (2008). *The sounds of Japanese*. Cambridge University Press.
- Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics, 71*, 147–161. <https://doi.org/10.1016/j.wocn.2018.07.008>
- Wang, Y., Bundgaard-Nielsen, R. L., Baker, B. J., & Maxwell, O. (2023). Difficulties in decoupling articulatory gestures in L2 phonemic sequences: The case of Mandarin listeners' perceptual deletion of English post-vocalic laterals. *Phonetica, 80*(1-2), 79–115. <https://doi.org/10.1515/phon-2022-0027>
- Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America, 130*(3), 1606–1617. <https://doi.org/10.1121/1.3621306>
- Watson, C. I., & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America, 106*(1), 458–468. <https://doi.org/10.1121/1.427069>
- Wells, J. C. (1982). *Accents of English*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511611759>
- Wells, J. C. (2008). *Longman pronunciation dictionary* (3rd ed.). Pearson Education Ltd.
- West, P. (1999a). The extent of coarticulation of English liquids: An acoustic and articulatory study. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic*

- Sciences* (pp. 1901–1904). https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/p14_1901.html
- West, P. (1999b). Perception of distributed coarticulatory properties of English /l/ and /r/. *Journal of Phonetics*, 27(4), 405–426. <https://doi.org/10.1006/jpho.1999.0102>
- Westbury, J. R. (1994). On coordinate systems and the representation of articulatory movements. *The Journal of the Acoustical Society of America*, 95(4), 2271–2273. <https://doi.org/10.1121/1.408638>
- Whalen, D. H., Iskarous, K., Tiede, M., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S. (2005). The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech, Language, and Hearing Research*, 48(3), 543–553. [https://doi.org/10.1044/1092-4388\(2005/037\)](https://doi.org/10.1044/1092-4388(2005/037))
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemond, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, 70, 86–116. <https://doi.org/10.1016/j.wocn.2018.03.002>
- Wieling, M., & Tiede, M. (2017). Quantitative identification of dialect-specific articulatory settings. *The Journal of the Acoustical Society of America*, 142(1), 389–394. <https://doi.org/10.1121/1.4990951>
- Wiese, R. (2011). The representation of rhotics. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 1–19). Wiley-Blackwell. <https://doi.org/10.1002/9781444335262.wbctp0030>

- Wikse Barrow, C., Włodarczak, M., Thörn, L., & Heldner, M. (2022). Static and dynamic spectral characteristics of Swedish voiceless fricatives. *The Journal of the Acoustical Society of America*, 152(5), 2588–2600. <https://doi.org/10.1121/10.0014947>
- Williams, D., & Escudero, P. (2014). A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *The Journal of the Acoustical Society of America*, 136(5), 2751–2761. <https://doi.org/10.1121/1.4896471>
- Wilson, I. (2014). Using ultrasound for teaching and researching articulation. *Acoustical Science and Technology*, 35(6), 285–289. <https://doi.org/10.1250/ast.35.285>
- Wilson, I., & Gick, B. (2014). Bilinguals use language-specific articulatory settings. *Journal of Speech, Language, and Hearing Research*, 57(2), 361–373. https://doi.org/10.1044/2013_JSLHR-S-12-0345
- Wilson, I., & Kanada, S. (2014). Pre-speech postures of second-language versus first-language speakers. *Journal of the Phonetic Society of Japan*, 18(2), 106–109. https://doi.org/10.24467/onseikenkyu.18.2_106
- Winter, B. (2019). *Sensory linguistics*. John Benjamins Publishing Company.
- Winter, B. (2020). *Statistics for linguists: An introduction using R*. Routledge.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370279>
- Wrench, A., & Balch-Tomes, J. (2022). Beyond the edge: Markerless pose estimation of speech articulators from ultrasound and camera images using DeepLabCut. *Sensors*, 22(3), 1133. <https://doi.org/10.3390/s22031133>
- Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics*, 52(4), 376–392. <https://doi.org/10.3758/BF03206698>
- Yamane, N., Howson, P., & Po-Chun (Grace), W. (2015). An ultrasound examination of taps in Japanese. In The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic*

- Sciences* (pp. 1–5). The International Phonetic Association. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0815.pdf>
- Yazawa, K., Whang, J., Kondo, M., & Escudero, P. (2020). Language-dependent cue weighting: An investigation of perception modes in L2 learning. *Second Language Research*, 36(4), 557–581. <https://doi.org/10.1177/0267658319832645>
- Ying, J., Shaw, J. A., Carignan, C., Proctor, M., Derrick, D., & Best, C. T. (2021). Evidence for active control of tongue lateralization in Australian English /l/. *Journal of Phonetics*, 86, 101039. <https://doi.org/10.1016/j.wocn.2021.101039>
- Ying, J., Shaw, J. A., Kroos, C., & Best, C. T. (2012). Relations between acoustic and articulatory measurements of /l/. *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, 109–112. <https://assta.org/proceedings/sst/SST-12/SST2012/PDF/AUTHOR/ST120075.PDF>
- Zharkova, N. (2013). Using ultrasound to quantify tongue shape and movement characteristics. *The Cleft Palate-Craniofacial Journal*, 50(1), 76–81. <https://doi.org/10.1597/11-196>
- Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/. *The Journal of the Acoustical Society of America*, 123(6), 4466–4481. <https://doi.org/10.1121/1.2902168>
- Zhou, Y., Bhattacharjee, S., Carroll, C., Chen, Y., Dai, X., Fan, J., Gajardo, A., Hadjipantelis, P. Z., Han, K., Ji, H., Zhu, C., Lin, S.-C., Dubey, P., Müller, H.-G., & Wang, J.-L. (2022). Fdapace: Functional data analysis and Empirical Dynamics.
- Zimmermann, G. N., Price, P., & Ayusawa, T. (1984). The production of English /r/ and /l/ by two Japanese speakers differing in experience with English.

Consolidated list of references

Journal of Phonetics, 12(3), 187–193. [https://doi.org/10.1016/S0095-4470\(19\)30873-3](https://doi.org/10.1016/S0095-4470(19)30873-3)

Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t, d/ in American English. *The Journal of the Acoustical Society of America*, 66(4), 1039–1050. <https://doi.org/10.1121/1.383323>

Appendix A

Participant information (1)

This section presents demographic information of individual participants collected with the demographic questionnaire (Appendix [F](#)).

- **Speaker ID:** Anonymised speaker ID
- **Gender:** Speaker's identified gender. F = female, M = male
- **Age:** Speaker's age at the time of recording
- **Country:** Speaker's country of origin. US = the United States
- **Region:** Speaker's region of origin
- **L1:** Speaker's first language(s)
- **Parent(s)' L1:** Speaker's parent(s)' first language
- **Languages until 13:** Languages that each speaker uses until 13 years of age

Table A.1: Speaker’s demographic information (1/4)

Speaker ID	Gender	Age	Country	Region	L1	Parent(s)’ L1	Languages until 13
4ps8zx	F	27	US	Wisconsin	English	English	English
5jzj2h	M	26	US	Pensylvania	English	English	English
5upwe3	M	24	US	Iowa	English	English	English
6p63jy	F	29	US	Oregon	English	English	English
bfwizh	M	27	Canada	British Columbia	English	English	English
byxcff	F	21	US	New Jersey	English	English	English
cu2jce	F	43	Poland	Krakow	Polish, English	Polish, English	Polish, English, French
ds6umh	F	31	US	Texas	English	English	English
h5s4x3	F	39	US	California	English	English	English
jcy8xi	F	32	Canada	Ontario	English	Cantonese	English, French, Cantonese
m46dhf	F	22	US	New York	English	Chinese	English
tay55n	F	30	Canada	Ontario	English	Ga/Fante	English
we8z58	F	25	US	Texas	English	English	English
xub9bc	F	29	Canada	Toronto	English	English	English

Table A.2: Speaker’s demographic information (2/4)

Speaker ID	Gender	Age	Country	Region	L1	Parent(s)’ L1	Languages until 13
2d57ke	M	20	Japan	Aichi	Japanese	Japanese	Japanese
2drb3c	F	20	Japan	Hyogo	Japanese	Japanese	Japanese
2zy9tf	M	19	Japan	Chiba	Japanese	Japanese	Japanese
3bcpyh	M	20	Japan	Kagawa	Japanese	Japanese	Japanese
3hsubn	F	19	Japan	Hyogo	Japanese	Japanese	Japanese
3pzrts	M	19	Japan	Mie	Japanese	Japanese	Japanese
3wy8us	F	19	Japan	Aichi	Japanese	Japanese	Japanese
54i2ks	M	20	Japan	Aichi	Japanese	Japanese	Japanese
7cd4t4	F	20	Japan	Hyogo	Japanese	Japanese	Japanese
9c4efu	F	22	Japan	Hyogo	Japanese	Japanese	Japanese
9zxyng	F	19	Japan	Hyogo	Japanese	Japanese	Japanese
a2kyah	F	21	Japan	Osaka	Japanese	Japanese	Japanese
a3h8n6	F	22	Japan	Fukuoka	Japanese	Japanese	Japanese
b6z2c	F	19	Japan	Hyogo	Japanese	Japanese	Japanese
birw55	F	20	Japan	Shimane	Japanese	Japanese	Japanese

Table A.3: Speaker’s demographic information (3/4)

Speaker ID	Gender	Age	Country	Region	L1	Parent(s)’ L1	Languages until 13
bj8mjm	M	18	Japan	Aichi	Japanese	Japanese	Japanese
c5y8z6	F	19	Japan	Shizuoka	Japanese	Japanese	Japanese
c7cr26	F	20	Japan	Hyogo	Japanese	Japanese	Japanese
cdsju7	F	21	Japan	Osaka	Japanese	Japanese	Japanese
dbtzn2	M	19	Japan	Aichi	Japanese	Japanese	Japanese
dcxuft	M	20	Japan	Aichi	Japanese	Japanese	Japanese
f9japd	M	21	Japan	Kagoshima	Japanese	Japanese	Japanese
fgd95u	F	20	Japan	Aichi	Japanese	Japanese	Japanese
fkewjr	F	18	Japan	Aichi	Japanese	Japanese	Japanese
heat7g	F	19	Japan	Fukui	Japanese	Japanese	Japanese
hgrist	M	20	Japan	Hyogo	Japanese	Japanese	Japanese
i3wa7f	M	21	Japan	Hyogo	Japanese	Japanese	Japanese
i7xs9b	M	21	Japan	Aichi	Japanese	Japanese	Japanese
j586ts	F	19	Japan	Aichi	Japanese	Japanese	Japanese
kjn9n4	F	20	Japan	Shizuoka	Japanese	Japanese	Japanese
m5r28t	F	18	Japan	Gifu	Japanese	Japanese	Japanese

Table A.4: Speaker’s demographic information (4/4)

Speaker ID	Gender	Age	Country	Region	L1	Parent(s)’ L1	Languages until 13
mgh8ee	M	20	Japan	Hyogo	Japanese	Japanese	Japanese
s6a8gh	M	20	Japan	Shiga	Japanese	Japanese	Japanese
srky8g	F	21	Japan	Aichi	Japanese	Japanese	Japanese
th7uwk	F	20	Japan	Kyoto	Japanese	Japanese	Japanese
uig6n9	F	18	Japan	Aichi	Japanese	Japanese	Japanese
ut4e5m	F	19	Japan	Aichi	Japanese	Japanese	Japanese
wrgwc3	F	21	Japan	Okayama	Japanese	Japanese	Japanese
z3n578	M	20	Japan	Hyogo	Japanese	Japanese	Japanese
zajk25	M	21	Japan	Hyogo	Japanese	Japanese	Japanese
zz3r2g	F	21	Japan	Okinawa	Japanese	Japanese	Japanese

Appendix B

Participant information (2)

This section presents information of individual participants regarding their language experience collected with the demographic questionnaire (Appendix [F](#)).

- **Speaker ID:** Anonymised speaker ID
- **Fluency:** Participant's self-evaluation of their own English ability. 1 = I do not speak English at all., 7 = No problem in using English in daily life.
- **Familiarity:** Participant's self-evaluation of their familiarity with English., 1 = I am not accustomed to it at all. 7 = I'm fully accustomed to it.
- **Use:** Participant's self-evaluation of the amount of English use per week. 1 = I do not use English., 7 = I only use English every day.
- **Conversation:** Participant's self-evaluation of the amount of English conversation per week (i.e., involving speaking). 1 = I do not speak English at all., 7 = I only speak English with people.
- **English study** (for L1 Japanese-speaking participants' only): Length of formal instruction of English study (in years)
- **Other languages:** Languages other than English or Japanese they speak. The number in parenthesis indicates self-evaluated fluency. 1 = I don't speak it fluently at all., 7 = No problem in using it in daily life.

Table B.1: Speaker's language experience (1/4)

Speaker	Fluency	Familiarity	Use	Conversation	English study	Other languages
4ps8zx	7	7	7	7		French (2)
5jzj2h	7	7	7	7		
5upwe3	7	7	7	6		Portuguese (7), Spanish (7), French (5), Turkish (5), Arabic (5), Italian (3), German (3), Russian (3) French (6), Polish (2)
6p63jy	7	7	7	5		
bfwizh	7	7	7	7		
byxcff	7	7	7	7		
cu2jce	7	7	4	4		Polish (7), French (6)
ds6umh	7	7	7	7		
h5s4x3	7	7	7	7		French (2)
jcy8xi	7	7	6	6		French (6), Cantonese (6)
m46dhf	7	7	6	6		Chinese (7)
tay55n	7	7	7	7		French (5)
we8z58	7	7	7	7		
xub9bc	7	7	7	7		

Table B.2: Speaker’s language experience (2/4)

Speaker	Fluency	Familiarity	Use	Conversation	English study	Other languages
2d57ke	4	4	3	1	9.5	Chinese (2)
2drb3c	4	4	4	4	11	Chinese (2), Korean (2)
2zy9tf	4	4	3	2	12	
3bcpyh	4	4	4	4	8	
3hsubn	4	4	3	3	13	
3pzrts	2	3	4	4	5	
3wy8us	4	4	4	5	10	
54i2ks	3	3	5	5	10	
7cd4t4	5	5	5	5	8	Korean (6)
9c4efu	5	6	5	0	11	Chinese (4)
9zxyng	4	5	5	3	8	Chinese (2), Korean (2)
a2kyah	5	5	4	2	10	Chinese (2)
a3h8n6	5	4	5	2	11	Chinese (2)
b6z2c	4	5	5	4	14	Korean (NA)
birw55	4	4	4	3	8	Chinese (2)

Table B.3: Speaker's language experience (3/4)

Speaker	Fluency	Familiarity	Use	Conversation	English study	Other languages
bj8mjm	3	3	2	2	6	
c5y8z6	3	3	5	4	7	
c7cr26	3	3	5	3	8	
cdsju7	4	4	5	2	9	
dbtzn2	4	4	3	1	6	
dcxuft	6	7	4	4	15	
f9japd	4	4	4	3	10	
fgd95u	5	5	4	0	7	
fkcwjr	3	3	3	0	7	
heat7g	2	3	2	2	11	
hgrist	4	3	3	3	8	Chinese (2)
i3wa7f	5	6	4	1	8	Korean (3)
i7xs9b	5	5	5	4	10	Chinese (2)
j586ts	3	3	3	2	9	
kjn9n4	1	1	3	2	10	

Table B.4: Speaker's language experience (4/4)

Speaker	Fluency	Familiarity	Use	Conversation	English study	Other languages
m5r28t	3	4	3	5	9	
mgh8ee	3	4	3	2	10	
s6a8gh	5	4	5	5	8	French (3)
srky8g	4	4	5	4	15	Chinese (2)
th7uwk	4	4	3	3	13	Chinese (2)
uig6n9	3	5	6	6	7	Swedish (1)
ut4e5m	4	4	1	1	10	
wrgwc3	4	4	3	3	12	
z3n578	3	3	4	3	10	Chinese (2)
zajk25	6	6	3	0	9	
zz3r2g	4	4	3	1	9	

Appendix C

Participant information (3)

This section presents information of individual participants regarding their overseas experience, occupation and previous training in linguistics/phonetics collected with the demographic questionnaire (Appendix **F**).

- **Speaker ID:** Anonymised speaker ID
- **Overseas:** Participant's overseas experience.
 - L1 English speakers: free description on any overseas experience from home countries
 - L1 Japanese speakers: length of experience of staying in an English-speaking country (unit: weeks). 1 month = 4 weeks (0.25 * 4)
- **Occupation:** Participant's occupation at the time of recording.
- **Linguistics:** Participant's former experience in Linguistics/Phonetics training
 - 1 None: No experience
 - 2 Class: I have taken a class (module) on linguistics/phonetics.
 - 3 Major: I have majored in linguistics/phonetics.
 - 4 Seminar: I have written my dissertation in linguistics/phonetics.

Table C.1: Speaker’s occupation and experience (1/4)

Speaker ID	Overseas	Occupation	Linguistics
4ps8zx	Netherlands, 5 months; UK, 3 years	Editor	4 Seminar
5jzj2h	UK, 3 years	Christian Charity staff	1 None
5upwe3	Lived in US until age 22, abroad for 2 years since (UK, 2 months)	Student, communication leadership consultant	3 Major
6p63jy	Poland, 1.5 months; France, 3.5 years; UK, 4 years	Sustainability consultant	2 Class
bfwizh	UK, 1 year 2 months	Former PG student	4 Seminar
byxceff	UK, 2 months	Student	2 Class
cu2jce	Australia, 1 year; UK, 10 years	Writer, linguist	4 Seminar
ds6umh	UK, 6 years	PhD student	1 None
h5s4x3	UK, 3 months	student	1 None
jcy8xi	France, 4 years; UK, 4 years UAE, 1.75 years	Sales admin	1 None
m46dlhf	Taiwan, 17 years	PG student	3 Major
tay55n	Ghana, 6 weeks; Switzerland, 3 months	Intelligence analyst	1 None
we8z58	Spain, 6 months	Marketing, sales	1 None
xub9bc	Thailand, 4 months; UK, 1 year	Senior analyst	1 None

Table C.2: Speaker's occupation and experience (2/4)

Speaker ID	Overseas	Occupation	Linguistics
2d57ke	0	UG student	1 None
2drb3c	1	UG student	2 Class
2zy9tf	0	UG student	2 Class
3bcpyh	0.5	UG student	1 None
3hsubn	0	UG student	3 Major
3pzrts	0	UG student	1 None
3wy8us	0	UG student	2 Class
54i2ks	0	UG student	2 Class
7cd4t4	0.5	UG student	1 None
9c4efu	0	UG student	3 Major
9zxyng	0	UG student	2 Class
a2kyah	5	UG student	3 Major
a3h8n6	0.25	UG student	4 Seminar
b62z2c	0.75	UG student	1 None
birw55	0.25	UG student	2 Class

Table C.3: Speaker's occupation and experience (3/4)

Speaker ID	Overseas	Occupation	Linguistics
bj8mjm	0	UG student	2 Class
c5y8z6	0	UG student	2 Class
c7cr26	0	UG student	1 None
cdsju7	4	UG student	2 Class
dbtzn2	0	UG student	2 Class
dcxuft	0.5	UG student	2 Class
f9japd	1.5	UG student	2 Class
fgd95u	0.5	UG student	2 Class
fkewjr	0	UG student	1 None
heat7g	0	UG student	2 Class
hgrist	1.5	UG student	2 Class
i3wa7f	0.25	UG student	2 Class
i7xs9b	1.5	UG student	2 Class
j586ts	0	UG student	2 Class
kjn9n4	0	UG student	1 None

Table C.4: Speaker’s occupation and experience (4/4)

Speaker ID	Overseas	Occupation	Linguistics
m5r28t	0	UG student	2 Class
mgh8ee	1.5	UG student	4 Seminar
s6a8gh	0	UG student	1 None
srky8g	0.5	UG student	4 Seminar
th7uwk	1	UG student	1 None
uig6n9	0	UG student	1 None
ut4e5m	0.5	UG student	2 Class
wrgwc3	4.25	UG student	2 Class
z3n578	0	UG student	2 Class
zajk25	4	UG student	2 Class
zz3r2g	4	UG student	2 Class

Appendix D

Information sheet

Notes:

- Amendments were made in the English version of the Information sheet. This is because the ethics application initially was intended for an experiment involving both ultrasound and electromagnetic articulography (EMA). At a later stage of the PhD project, I decided not to use EMA so I omitted the content related to EMA and adjusted the amount of compensation to the participants. £12 was rounded up to £15 for L1 English-speaking participants due to vouchers used for payment only available in multiples of a £5.
- In the Japanese versions, used at two different institutions respectively, I am listed as one of the researchers (rather than the principal investigator) in the project. This is due to regulations related to the eligibility of ethics application at the respective institutions.

Participant information sheet

Project: Investigating speech articulation of English and Japanese sounds

For further information about how Lancaster University processes personal data for research purposes and your data rights please visit our webpage: www.lancaster.ac.uk/research/data-protection

I am a PhD student at Lancaster University, and I would like to invite you to take part in a research study about: Analysing articulation and acoustics in native and nonnative speech.

Please take time to read the following information carefully before you decide whether or not you wish to take part.

What is the study about?

This study aims to investigate how your tongue moves when pronouncing speech sounds.

Why have I been invited?

I have approached you because I am interested in understanding how native and nonnative speakers of English would differ in the movement of articulators (e.g., tongue, lips, etc). I would be very grateful if you would agree to take part in this study.

What will I be asked to do if I take part?

If you decided to take part, this would involve a visit to our Phonetics Lab at Lancaster University to engage in: 1) reading short sentences in English (and in Japanese, if you are a native Japanese participant) with specialised equipment being attached either on your tongue or under your chin, or both, and 2) listening to a set of sounds and make certain judgements on them. It may take up to **2–3 1.5 hours** to complete all above procedures.

What are the possible benefits from taking part?

This study endeavours to take on new approaches to analysing speech sounds by looking closely and directly into articulation of first and second language speech, so that we will be able to uncover what was once 'invisible' to us. This has been an elusive part of research, and your cooperation will be valuable. Upon completion of all tasks, we will pay you an equivalent amount of ~~20-GBP~~ **2,000 JPY (approximately £12)** to thank you for your time and participation.

Do I have to take part?

No. It's completely up to you to decide whether or not you take part. Your participation is voluntary.

What if I change my mind?

If you change your mind, you are free to withdraw at any time during your participation in this study. If you want to withdraw, please let me know, and I will destroy any data you contributed to the study. However, it is difficult and often impossible to take out data from one specific participant when this has already been anonymised or pooled together with other people's data. Therefore, you can only withdraw up to 2 weeks after taking part in the study – after this date, the data will have been anonymised and I will not be able to identify which is your data.

What are the possible disadvantages and risks of taking part?

Taking part will mean investing 1.5 hours of your time including briefing, preparation, experimental sessions and finishing-up. It is unlikely that there will be any major disadvantages to taking part.

If you are asked to participate in the data collection session using EMA (electromagnetic articulography), however, you may find having sensors attached to your tongue unusual. Also, there is possibility for minor discomfort in removing the tongue sensors due to the dental glue used to affix the sensors. However, it should be stressed that the researcher will use latex gloves when attaching sensors to your tongue. Therefore, **you must not take part in the EMA session if you have a latex allergy.**

If you are asked to participate in the speech experiment using only ultrasound, there is no established risks associated with its use.

Will my data be identifiable?

I will keep all personal information about you (e.g. your name and other information about you that can identify you) confidential, that is I will not share it with others, except for the people mentioned in the following section for research purposes. I will remove any personal information from the written record of your contribution. All reasonable steps will be taken to protect the anonymity of the participants involved in this project.

How will we use the information you have shared with us and what will happen to the results of the research study?

I will use the obtained data for research purposes only. This will include my PhD thesis and other publications, for example journal articles, academic presentations and research meetings. I may also have to play very short audio extracts of speech at academic conferences to demonstrate the speech phenomena investigated in the study.

The data obtained from you will be shared with my PhD supervisors, Dr Claire Nance and Dr Sam Kirkham in the Department of Linguistics and English Language, given the nature of the research being a PhD project. Also, other research collaborators may also need to look into the data if necessary, provided they have received proper guidance on the data security.

How my data will be stored

Your data will be stored in encrypted files (that is no-one other than me, the researcher will be able to access them) and on password-protected computers. I will store hard copies of any data securely in locked cabinets in my office. I will keep data that can identify you separately from non-personal information. In accordance with University guidelines, I will keep the data securely for a minimum of ten years.

What if I have a question or concern?

If you have any queries or if you are unhappy with anything that happens concerning your participation in the study, please contact myself (Takayuki Nagamine, t.nagamine@lancaster.ac.uk) in Room C29, Department of Linguistics and English Language, Lancaster University, County South, Lancaster, LA1 4YL, United Kingdom.

You can also contact my supervisors: Dr Claire Nance (email: c.nance@lancaster.ac.uk) or Dr Sam Kirkham (email: s.kirkham@lancaster.ac.uk), both of whom are based at: Department of Linguistics and English Language, Lancaster University, County South, Lancaster, LA1 4YL, United Kingdom.

If you have any concerns or complaints that you wish to discuss with a person who is not directly involved in the research, you can also contact: Professor Jonathan Culpeper, Head of Department (email: j.culpeper@lancaster.ac.uk)

This study has been reviewed and approved by the Faculty of Arts and Social Sciences and Lancaster Management School's Research Ethics Committee.

Thank you for considering your participation in this project.

超音波舌断層撮像を用いた英語発音に関する研究

研究に関する説明書

はじめに

この説明文書は、「超音波舌断層撮像を用いた英語発音に関する研究」にご参加・ご協力をお願いするにあたって、研究・調査の内容についてご説明し、ご理解いただくために用意したものです。研究責任者または研究実施者から説明をお受けになり、本説明文書をお読みになってご理解いただいた上で、この研究・調査に参加されるかどうかを決めてください。内容についてわからないこと、お尋ねになりたいことなどがありましたら、研究責任者又は研究実施者までご遠慮なくご質問ください。

なお、本研究計画は、神戸学院大学、ランカスター大学、名城大学所属の教員および大学院生の共同研究で、「人を対象とする生命科学・医学系研究に関する倫理指針」に則り、神戸学院大学における人を対象とする非医学系研究倫理審査委員会の審査を受けて、神戸学院大学長・名城大学長の許可を得て実施するものです。研究・調査に参加・ご協力いただける場合には、同意書にご署名をお願いします。

1. 研究の目的

この研究は、英語音声を発音するときに、舌や唇などの「調音器官」がどのように使われているかを調べることを目的としています。本研究の結果などは成績等の評価とは一切関係しません。

2. 実施方法

この研究は、神戸学院大学において、2022年9月から2025年3月まで実施される予定です。ただし、参加・協力いただく方に研究に参加していただくのは、2022年9月から10月31日の期間内の1日です。神戸学院大学において、日本語を母語とし、英語を学習している約20名の方にご協力をお願いしています。

本研究にご参加いただける場合、以下のような作業をお願いします：

- | |
|--|
| <p>1. 音声発話課題: 顎の下に超音波機器を装着し、英語や日本語の単語や短文を読み上げる
(所要時間: 約1時間～1時間半程度 実施方法: 大学にて対面実施)</p> <p>2. 音声聞き取り課題: 英単語を聞き、聞こえてきた単語を選ぶ
(所要時間: 約30分程度 実施方法: 音声発話課題に引き続き、大学にて対面実施)</p> |
|--|

本研究により得られたデータは、研究目的にのみ使用されます。電磁的データは、パスワード保護がなされたハードディスクや、研究者以外がアクセスすることのできないクラウドに保管されます。紙面で収集されたデータは、研究室の施錠可能なロッカー等において厳重に保管されます。

皆さんから得られたデータを用い、博士論文やその他の出版物(例えば、雑誌記事、学会発表、研究会など)において成果を公表する予定です。その際、学会発表等において、音声の実演するために、ごく短い音声の抜粋を流すことがあります。

本研究は、神戸学院大学(兵庫県)・名城大学(愛知県)・ランカスター大学(英国)の三機関による共同研究です。本研究により得られたデータは、これらの大学からプロジェクトに参画している研究者間で共有されます。

3. 研究対象者として選定された理由

本研究では、以下の条件に該当する方に対し、ご協力をお願いしています。あなたは、以下の条件を全て満たしていることから、この度研究参加の依頼をさせていただいております。

- (1) 日本語を母語とすること
- (2) 18歳以上であること
- (3) 日本の小学校～高等学校における英語教育を経験していること
- (4) 学部1・2年生の場合、長期海外滞在経験がないこと(約1ヶ月未満を基準とする)
- (5) TOEIC や英検等の英語語学試験の受験経験があり、スコアの提供ができること
- (6) 音声発話・聞き取りが正常であること

4. 研究対象者に生じる負担並びに予測されるリスク及び利益

本研究は、「超音波舌断層撮像」という技術を使用します。超音波技術は、産婦人科等で子宮内の様子を観察するために使用されるいわゆる「エコー技術」です。このことからわかるように、超音波は皆さんの人体に悪影響を与えることはなく、安全性が確立されています。

超音波機器を顎下に固定するため、研究参加者の皆さんには特別なヘッドセットを装着していただきます。これも、人体に悪影響を与えるものではないですが、約1時間程度装着していただくため、疲労を伴うことがあります。また、口唇の動きを撮影した映像からは、現段階では個人の特定はほぼ不可能ですが、将来的に画像認識技術の発展により、個人の特定が可能となる場合があります。そのため、研究成果の公表時には、画像の解像度を落とす等によりその可能性を最小化することとし、口唇データにかかる個人情報に厳重な取り扱いをいたしますので、どうかご安心ください。

本研究により得られたデータは、希望に応じて、皆さんにお返しします。英語 L・R 音の聞き取り能力や、皆さんが発話した英語 L・R 音の正確性に関する英語母語話者の判断の結果を参照することで、皆さんが英語発音学習において現在どのような立ち位置にいるのかを把握することが可能になります。さらに、希望者に対しては、超音波画像をもとに、英語 L・R 音の発音における舌の使い方についてフィードバックを実施することも可能です。皆さんの英語発音学習の一助として、どうぞお役立てください。

5. 同意の撤回

この研究・調査に参加するかどうかは、あなたの自由な意思でお決めください。参加に同意していただける場合には、同意書に署名をお願いします。この研究・調査に参加されている期間中いつでも同意を取り消すことができます。

しかし、あなたが参加を取りやめたい場合は、**実験終了後2週間以内(2022年11月中旬まで)**に、**研究代表者にお知らせください**。その時点で、あなたに関する全てのデータを破棄いたします。これ以後は、ある特定個人のデータのみを取り出すことは困難となります。

研究・調査への協力をお断りになったり、協力を取り消される場合であっても、研究・調査の関係者との人間関係が気まづくなったり、何らかの不利益を被ることは全くありませんので、どうぞご安心ください。

6. 研究に関する情報公開の方法

この研究結果は、学会や学術雑誌、あるいは学術論文等で公表します。ただし、参加いただいた方の個人情報(名前や住所、電話番号など)あるいは個人を特定し得る情報の公表は一切いたしません。

本研究により得られた結果については、基本的に成果物を研究対象者に電子的に送付する形で実施しますが、研究参加者の皆さんからの開示の求めがあった場合は、他の方々の個人情報の保護や、研究の知的財産権等に支障がない範囲で報告させていただきます。

この研究の計画や方法について、もっと詳しくお知りになりたい場合には、研究責任者までご連絡ください。この調査・研究に参加・協力していただいている他の方々の個人情報の保護や、研究の知的財産権等に支障がない範囲で、研究計画書を閲覧していただくか、研究責任者等からご説明等をさせていただきます。

7. 個人情報等の取り扱い及び保管の方法

データ収集終了後2週間(2022年11月中旬頃)を経たのち、全てのデータは匿名化されます。それ以後は、特定の個人のデータのみを取り出すことは不可能となります。神戸学院大学、名城大学から得られた個人情報は、ID化し個人を識別できない情報に加工されます。個人情報は、データ収集終了2週間後に完全に削除されますので、それ以後は個人の特定は不可能となります。

研究により得られた電磁的資料は、ID化処理を施した上で(1)個人情報保管責任者の所有するパスワード保護された外付けハードドライブ及び(2)ランカスター大学のパスワード保護された OneDrive フォルダーに保存されます。また、紙面による質問紙調査の結果は、個人情報保管責任者の研究室において、施錠のできるキャビネットに保管します。

本研究において収集されたデータは、今後様々な分析・考察等が見込まれるため、基本的には廃棄を前提とはせず、上記に記された厳重な保管方法により、期限を定めず保管することといたします。

8. 研究における利益相反等の情報

この研究にかかる費用は、「公益財団 村田学術振興財団 2022年度研究助成(受給者:長峯貴幸)」から支出されます。特定の企業等との間に研究結果や研究対象者の保護に影響を及ぼす可能性のある経済的利益関係等の利益相反の状況はありません。

9. 謝礼

本研究に参加してくださった方には、謝礼として1時間あたり ¥1,000 をお支払いいたします。また、実験から得られたデータをもとに「発音カルテ(仮称)」を作成し、皆さんにお返しする予定です。皆さんの英語 L・R 音の発音における舌の画像や、他の種々の分析の結果を、発音学習にお役立てください。

10. 研究体制及び研究に対する相談等の問い合わせ先

本研究についてご質問がありましたら、以下の研究者にいつでもお問い合わせください。

研究代表者

中西のりこ(神戸学院大学 グローバル・コミュニケーション学部 英語コース 教授)

nakanisi@gc.kobegakuin.ac.jp

研究責任者

長峯貴幸(ランカスター大学 言語学・英語学研究科 博士課程2年)

t.nagamine@lancaster.ac.uk

西尾由里(名城大学 外国語学部国際英語学科 教授)

ynishio@meijo-u.ac.jp

超音波舌断層撮像を用いた英語発音に関する研究

研究に関する説明書（名城大学）

はじめに

この説明文書は、「超音波舌断層撮像を用いた英語発音に関する研究」にご参加・ご協力をお願いするにあたって、研究・調査の内容についてご説明し、ご理解いただくために用意したものです。研究責任者または研究実施者から説明をお受けになり、本説明文書をお読みになってご理解いただいた上で、この研究・調査に参加されるかどうかを決めてください。内容についてわからないこと、お尋ねになりたいことなどがありましたら、研究責任者又は研究実施者までご遠慮なくご質問ください。

なお、本研究計画は、神戸学院大学、ランカスター大学、名城大学所属の教員および大学院生の共同研究で、「人を対象とする生命科学・医学系研究に関する倫理指針」に則り、神戸学院大学における人を対象とする非医学系研究倫理審査委員会の審査を受けて、神戸学院大学長・名城大学長の許可を得て実施するものです。研究・調査に参加・ご協力いただける場合には、同意書にご署名をお願いします。

1. 研究の目的

この研究は、英語発音を録音するときに、舌や唇などの「調音器官」がどのように使われているかを調べることを目的としています。本研究の結果などは成績等の評価とは一切関係しません。

2. 実施方法

この研究は、神戸学院大学において、2022年9月から2025年3月まで実施される予定です。ただし、参加・協力いただく方に研究に参加していただくのは、2022年9月から10月31日の期間内の1日です。名城大学において、日本語を母語とし、英語を学習している約20名の方にご協力をお願いしています。本研究にご参加いただける場合、以下のような作業をお願いします：

1. 音声発話課題：顎の下に超音波機器を装着し、英語や日本語の単語や短文を読み上げる

（所要時間：約1時間～1時間半程度 実施方法：大学にて対面実施）

2. 音声聞き取り課題：英単語を聞き、聞こえてきた単語を選ぶ

（所要時間：約30分程度 実施方法：音声発話課題に引き続き、大学にて対面実施）

本研究により得られたデータは、研究目的にのみ使用されます。電磁的データは、パスワード保護がなされたハードディスクや、研究者以外がアクセスすることのできないクラウドに保管されます。紙面で収集されたデータは、研究室の施錠可能なロッカー等において厳重に保管されます。

皆さんから得られたデータを用い、博士論文やその他の出版物（例えば、雑誌記事、学会発表、研究会など）において成果を公表する予定です。その際、学会発表等において、音声の実演するために、ごく短い音声の抜粋を流すことがあります。

本研究は、神戸学院大学（兵庫県）・名城大学（愛知県）・ランカスター大学（英国）の三機関による共同研究です。本研究により得られたデータは、これらの大学からプロジェクトに参画している研究者間で共有されます。

3. 研究対象者として選定された理由

本研究では、以下の条件に該当する方に対し、ご協力をお願いしています。あなたは、以下の条件を全て満たしていることから、この度研究参加の依頼をさせていただいております。

- (1) 日本語を母語とすること
- (2) 18歳以上であること
- (3) 日本の小学校～高等学校における英語教育を経験していること
- (4) 学部1・2年生の場合、長期海外滞在経験がないこと（約1ヶ月未満を基準とする）
- (5) TOEIC や英検等の英語語学試験の受験経験があり、スコアの提供ができること
- (6) 音声発話・聞き取りが正常であること

4. 研究対象者に生じる負担並びに予測されるリスク及び利益

本研究は、「超音波舌断層撮像」という技術を使用します。超音波技術は、産婦人科等で子宮内の様子を観察するために使用されるいわゆる「エコー技術」です。このことからわかるように、超音波は皆さんの人体に悪影響を与えることはなく、安全性が確立されています。

超音波機器を顎下に固定するため、研究参加者の皆さんには特別なヘッドセットを装着していただきます。これも、人体に悪影響を与えるものではないですが、約1時間程度装着していただくため、疲労を伴うことがあります。また、口唇の動きを撮影した映像からは、現段階では個人の特定はほぼ不可能ですが、将来的に画像認識技術の発展により、個人の特定が可能となる場合があります。そのため、研究成果の公表時には、画像の解像度を落とす等によりその可能性を最小化することとし、口唇データにかかる個人情報に厳重な取り扱いをいたします。

本研究により得られたデータは、希望に応じて、皆さんにお返しします。英語 L・R 音の聞き取り能力や、皆さんが発話した英語 L・R 音の正確性に関する英語母語話者の判断の結果を参照することで、皆さんが英語発音学習において現在どのような立ち位置にいるのかを把握することが可能になります。さらに、希望者に対しては、超音波画像をもとに、英語 L・R 音の発音における舌の使い方についてフィードバックを実施することも可能です。皆さんの英語発音学習の一助として、どうぞお役立てください。

5. 同意の撤回

この研究・調査に参加するかどうかは、あなたの自由な意思でお決めください。参加に同意していただける場合には、同意書に署名をお願いします。この研究・調査に参加されている期間中いつでも同意を取り消すことができます。

しかし、あなたが参加を取りやめたい場合は、**実験終了後2週間以内(2022年11月中旬まで)**に、**研究代表者にお知らせください**。その時点で、あなたに関する全てのデータを破棄いたします。これ以後は、ある特定個人のデータのみを取り出すことは困難となります。

研究・調査への協力をお断りになったり、協力を取り消される場合であっても、研究・調査の関係者との人間関係が気まづくなったり、何らかの不利益を被ることは全くありませんので、どうぞご安心ください。

6. 研究に関する情報公開の方法

この研究結果は、学会や学術雑誌、あるいは学術論文等で公表します。ただし、参加いただいた方の個人情報(名前や住所、電話番号など)あるいは個人を特定し得る情報の公表は一切いたしません。

本研究により得られた結果については、基本的に成果物を研究対象者に電子的に送付する形で実施しますが、研究参加者の皆さんからの開示の求めがあった場合は、他の方々の個人情報の保護や、研究の知的財産権等に支障がない範囲で報告させていただきます。

この研究の計画や方法について、もっと詳しくお知りになりたい場合には、研究責任者までご連絡ください。この調査・研究に参加・協力していただいている他の方々の個人情報の保護や、研究の知的財産権等に支障がない範囲で、研究計画書を閲覧していただくか、研究責任者等からご説明等をさせていただきます。

7. 個人情報等の取り扱い及び保管の方法

データ収集終了後2週間(2022年11月中旬頃)を経たのち、全てのデータは匿名化されます。それ以後は、特定の個人のデータのみを取り出すことは不可能となります。神戸学院大学、名城大学から得られた個人情報は、ID化し個人を識別できない情報に加工されます。個人情報は、データ収集終了2週間後に完全に削除されますので、それ以後は個人の特定は不可能となります。

研究により得られた電磁的資料は、ID化処理を施した上で(1)個人情報保管責任者の所有するパスワード保護された外付けハードドライブ及び(2)ランカスター大学のパスワード保護された OneDrive フォルダーに保存されます。また、紙面による質問紙調査の結果は、個人情報保管責任者の研究室において、施錠のできるキャビネットに保管します。

本研究において収集されたデータは、今後様々な分析・考察等が見込まれるため、基本的には廃棄を前提とはせず、上記に記された厳重な保管方法により、期限を定めず保管することといたします。

8. 研究における利益相反等の情報

この研究にかかる費用は、「公益財団 村田学術振興財団 2022年度研究助成(受給者:長峯貴幸)」から支出されます。特定の企業等との間に研究結果や研究対象者の保護に影響を及ぼす可能性のある経済的利益関係等の利益相反の状況はありません。

9. 謝礼

本研究に参加してくださった方には、謝礼として1時間あたり ¥1,000 をお支払いいたします。また、実験から得られたデータをもとに「発音カルテ(仮称)」を作成し、皆さんにお返しする予定です。皆さんの英語 L・R 音の発音における舌の画像や、他の種々の分析の結果を、発音学習にお役立てください。

10. 研究体制及び研究に対する相談等の問い合わせ先

本研究についてご質問がありましたら、以下の研究者にいつでもお問い合わせください。

研究代表者

中西のりこ(神戸学院大学 グローバル・コミュニケーション学部 英語コース 教授)

nakanisi@gc.kobegakuin.ac.jp

研究責任者

長峯貴幸(ランカスター大学 言語学・英語学研究科 博士課程2年)

t.nagamine@lancaster.ac.uk

西尾由里(名城大学 外国語学部国際英語学科 教授)

ynishio@meijo-u.ac.jp

Appendix E

Consent form

CONSENT FORM



Name of Researchers: Takayuki Nagamine

Email: t.nagamine@lancaster.ac.uk

Please tick each box

1. I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	<input type="checkbox"/>
2. I understand that my participation is voluntary and that I am free to withdraw at any time during my participation, and within 2 weeks after I took part in the study in this study, without giving any reason.	<input type="checkbox"/>
3. I understand that any information given by me may be used in future reports, academic articles, publications or presentations by the researcher/s, but my personal information will not be included, and all reasonable steps will be taken to protect the anonymity of the participants involved in this project.	<input type="checkbox"/>
4. I understand that my name/my organisation's name will not appear in any reports, articles or presentation without my consent.	<input type="checkbox"/>
5. I understand that any recorded data will be transcribed, and that data will be protected on encrypted devices and kept secure.	<input type="checkbox"/>
6. I understand that data will be kept according to University guidelines for a minimum of 10 years after the end of the study.	<input type="checkbox"/>
7. I agree to take part in the above study.	<input type="checkbox"/>

Name of Participant

Date

Signature

I confirm that the participant was given an opportunity to ask questions about the study, and all the questions asked by the participant have been answered correctly and to the best of my ability. I confirm that the individual has not been coerced into giving consent, and the consent has been given freely and voluntarily.

Signature of Researcher /person taking the consent _____ Date _____ Day/month/year

One copy of this form will be given to the participant either digitally or physically and the original kept in the files of the researcher at Lancaster University

神戸学院大学 学長 殿

同意書

私は、「超音波舌断層撮像を用いた英語発音に関する研究」の実施に際し、以下の項目につき十分説明を受け、その趣旨を理解いたしましたので、自らの自由意思により本調査に協力し、参加することに同意します。

1. 倫理審査委員会の審査を受け、神戸学院大学長の許可を受けていることについて
2. 研究機関等の名称及び研究責任者の氏名について
3. 研究の目的や意義について
4. 研究の方法について
5. 研究の対象者について
6. 研究対象者に生じる負担や予測されるリスク及び利益について
7. 研究への参加とその撤回について
8. 同意を撤回しても不利益とならないことについて
9. 研究に関する情報公開について
10. 研究に関する資料の入手・閲覧について
11. 個人情報の取扱いについて
12. 試料・情報の保管等について
13. 研究の資金源・利益相反について
14. 研究により得られた結果等の取扱いについて
15. 経済的負担又は謝礼について
16. 不特定の将来の研究利用

同意者（本人）：

同意日 _____ 年 _____ 月 _____ 日

氏名（署名） _____

同意者（代諾者）：

同意日 _____ 年 _____ 月 _____ 日 本人との続柄 _____

氏名（署名） _____

説明者：

同意日 _____ 年 _____ 月 _____ 日 大学名・職名 _____

氏名（署名） _____

***同意書は2部作成し、研究参加者・研究者が1部ずつ保管します。**

名城大学 学長 殿

同意書

私は、「超音波舌断層撮像を用いた英語発音に関する研究」の実施に際し、以下の項目につき十分説明を受け、その趣旨を理解いたしましたので、自らの自由意思により本調査に協力し、参加することに同意します。

1. 倫理審査委員会の審査を受け、名城大学長の許可を受けていることについて
2. 研究機関等の名称及び研究責任者の氏名について
3. 研究の目的や意義について
4. 研究の方法について
5. 研究の対象者について
6. 研究対象者に生じる負担や予測されるリスク及び利益について
7. 研究への参加とその撤回について
8. 同意を撤回しても不利益とならないことについて
9. 研究に関する情報公開について
10. 研究に関する資料の入手・閲覧について
11. 個人情報の取扱いについて
12. 試料・情報の保管等について
13. 研究の資金源・利益相反について
14. 研究により得られた結果等の取扱いについて
15. 経済的負担又は謝礼について
16. 不特定の将来の研究利用

同意者（本人）：

同意日 _____ 年 _____ 月 _____ 日

氏名（署名）_____

同意者（代諾者）：

同意日 _____ 年 _____ 月 _____ 日 _____ 本人との続柄

氏名（署名）_____

説明者：

同意日 _____ 年 _____ 月 _____ 日 _____ 大学名・職名

氏名（署名）_____

***同意書は2部作成し、研究参加者・研究者が1部ずつ保管します。**

Appendix F

Demographic questionnaire

For researcher use	ID	2022年 月 日 時 分 ~ 時 分
--------------------	----	------------------------

Participant Questionnaire

Thank you very much for your participation in the experiment today. I would like to ask you some questions related to yourself. The information obtained from this questionnaire will be used only for research purposes; no information that may lead to identifying you will be revealed when the results are presented in conference presentations, journal articles, or my PhD thesis.

Please only answer the questions with which you feel comfortable. It is ok for you not to answer some of the questions when you do not feel comfortable answering; in that case, it would be helpful if you could write 'XX' to indicate it.

1. About yourself

(1) What is your gender? _____

(2) Where were you born? Country: _____ State/Region: _____

(3) Where were you raised before the age of 13?

(4) How old are you? _____ years old

(5) What is your first language(s)? _____

(6) What is your parents' first language(s)? _____

(7) What language(s) did you mainly use before the age of 13? _____

(8) Do you have any history of language impairment?

No speaking hearing writing reading

(9) Have you had any experience of staying outside your home country longer than a month?

Where?	When?	For how long?
_____	_____	_____
_____	_____	_____

2. Your language backgrounds

(1) How would you assess your English ability?

I do not speak English at all. 1 --- 2 --- 3 --- 4 --- 5 --- 6 --- 7 No problems in using English in daily life.

(2) How much are you accustomed to using English?

I am not accustomed to it at all. 1 --- 2 --- 3 --- 4 --- 5 --- 6 --- 7 I'm fully accustomed to it.

(3) How much do you use English in general per week?

I do not use English at all. 1 ---- 2 ---- 3 ---- 4 ---- 5 ---- 6 ---- 7 I only use English every day.

(4) How much do you use English to talk with other people per week?

I do not speak English at all. 1 ---- 2 ---- 3 ---- 4 ---- 5 ---- 6 ---- 7 I only speak English with people.

Please go overleaf

3. About your educational background

(1) What is your occupation? _____

(2) Do you have any experience in linguistics and/or phonetics?

- None
- I have taken a class (module) on linguistics/phonetics.
- I have majored in linguistics/phonetics/
- I have written my dissertation in linguistics/phonetics/
- Miscellaneous (_____)

(3) Do you speak any other language than English? And how fluent are you in each of them?

language	fluency	remarks
	I don't speak it fluently at all.	
	1-2-3-4-5-6-7	
	No problems in using it in daily life.	
	I don't speak it fluently at all.	
	1-2-3-4-5-6-7	
	No problems in using it in daily life.	

This is the end of the demographic survey. Thank you very much for your cooperation!

研究者 使用欄	ID	2022年 月 日 時 分 ~ 時 分
------------	----	------------------------

超音波舌断層撮像を用いた英語発音に関する研究 研究参加者アンケート

この度は、皆さんの貴重な時間を頂戴し、超音波を用いた発音研究へご参加いただき、ありがとうございます。最後に、以下の設問にご回答をお願いします。回答は、差し支えのない範囲で構いませんが、できるだけ多く埋めていただくと助かります。設問の意味や回答の仕方がわからない、設問について疑問点がある等の場合は、同席する研究者にどうぞ遠慮なくお尋ねください。設問は、表・裏の両面にあります。

本アンケートから得られた情報は、「研究に関する説明書」にて想定されている用途以外への利用は一切行いません。また、研究成果の発表の際も、みなさん個人を特定できるような情報は一切開示されませんので、安心してご回答ください。終了後、アンケートは研究室にて厳重に保管されます。

1. あなた自身について

- (1) 性別 女性 男性 開示を希望しない 特定不可能 その他()
- (2) 出身地(国名・県名) 国名: 県名や地域名:
- (3) 年齢 歳
- (4) あなたの母語・第一言語 日本語 その他()
- (5) 親御さんの母語 日本語 その他()
- (6) 13歳になるまでに用いていた言語 日本語 その他()
- (7) 言語機能に関して、お医者さんから受けたことがある診断
なし 話すこと 聞くこと 書くこと 読むこと
- (8) 英語圏での留学・在住・滞在経験及びその期間

滞在国・地域の名称	時期(年・月)	期間(〇ヶ月)

2. あなたの言語経験について

- (1) あなたの英語発話能力(自己評価)
全く英語ができない 1 ----- 2 ----- 3 ----- 4 ----- 5 ----- 6 ----- 7 支障なく日常生活が送れる
- (2) 英語への慣れ
全く慣れていない 1 ----- 2 ----- 3 ----- 4 ----- 5 ----- 6 ----- 7 大いに慣れている
- (3) 1週間あたりに英語に触れる量
全く使用しない 1 ----- 2 ----- 3 ----- 4 ----- 5 ----- 6 ----- 7 英語しか触れない
(大学の授業のみである: はい いいえ)
- (4) 1週間あたりに英語を使って他人と会話する量
全く使用しない 1 ----- 2 ----- 3 ----- 4 ----- 5 ----- 6 ----- 7 英語のみで会話している
(大学の授業のみである: はい いいえ)

裏面に続きます

3. あなたの教育について

(1) 英語に関する語学検定の得点や資格

資格試験の名称	時期(年・月)	得点
TOEIC Listening/Reading		リスニング: _____
		リーディング: _____
		合計: _____
TOEFL		
その他		

(2) 最終学歴 高等学校卒 その他(_____)

(3) 現在所属する大学名・学部や学科 大学名:名城大学 学部:
学科: _____ 学年: _____

(4) 英語学習歴(年数・場所) 合計(_____)年
 詳細: 小学校(_____ 年生) 中学校から 高等学校から
学習塾・英会話等(_____ 歳から)

(5) 言語学や音声学の経験 全くない 授業で履修した コースの専攻 ゼミの専攻
その他(_____)

(6) 英語以外に話す言語(言語名・言語能力の自己評価(設問 2-(2)を参照)等を記入してください。)

言語	自己評価	備考・その他
	全くできない 1-2-3-4-5-6-7	
	支障なく日常生活が送れる	
	全くできない 1-2-3-4-5-6-7	
	支障なく日常生活が送れる	

4. その他

英語発音についてのフィードバックを希望されますか? 希望する 希望しない

*「希望する」と答えた方:希望するフィードバックの種類を選んでください。(複数回答可)

ア:自分の英語発音の超音波画像 イ:音声聞き取り課題の結果 ウ:自分の英語発音の正確性チェック

Appendix G

Word familiarity survey

Lexical familiarity

Participant ID (

) Date (

)

Please rate how familiar you think you are with the following lexical items on a scale of three:

⊙ = Yes, I know it. ○ = I think I know it. △ = I might know it. ✕ = I don't know it.

reef ()	red ()	believe ()
reap ()	loot ()	rash ()
rife ()	rag ()	lead ()
veer ()	laze ()	leech ()
lap ()	rob ()	ram ()
right ()	lewd ()	led ()
lice ()	road ()	lack ()
rot ()	lube ()	lamp ()
bereave ()	leaf ()	lose ()
rake ()	room ()	robe ()
laid ()	ruse ()	ramp ()
fear ()	reeve ()	read ()
leak ()	veal ()	leave ()
lake ()	rap ()	peer ()
lag ()	rube ()	wrong ()
peel ()	reach ()	lash ()
raid ()	leap ()	load ()
reek ()	root ()	light ()
lobe ()	rude ()	raise ()
loom ()	lock ()	rice ()
rack ()	long ()	
rock ()	rim ()	
limb ()	lot ()	
lamb ()	feel ()	
lob ()	life ()	

以下の単語に対して、どれくらい馴染みがありますか。以下の4段階で自己評価してください。

◎=確実に知っている	○=おそらく知っている	△=見たことはあると思う	✕=今回初めてみた
------------	-------------	--------------	-----------

reef ()	red ()	believe ()
reap ()	loot ()	rash ()
rife ()	rag ()	lead ()
veer ()	laze ()	leech ()
lap ()	rob ()	ram ()
right ()	lewd ()	led ()
lice ()	road ()	lack ()
rot ()	lube ()	lamp ()
bereave ()	leaf ()	lose ()
rake ()	room ()	robe ()
laid ()	ruse ()	ramp ()
fear ()	reeve ()	read ()
leak ()	veal ()	leave ()
lake ()	rap ()	peer ()
lag ()	rube ()	wrong ()
peel ()	reach ()	lash ()
raid ()	leap ()	load ()
reek ()	root ()	light ()
lobe ()	rude ()	raise ()
loom ()	lock ()	rice ()
rack ()	long ()	
rock ()	rim ()	
limb ()	lot ()	
lamb ()	feel ()	
lob ()	life ()	

Appendix H

Lexical frequency for the perception experiment

	JACET8000 Rank	JACET8000 POS	AmE06 Frequency	AmE06 Range		JACET8000 Rank	JACET8000 POS	AmE06 Frequency	AmE06 Range
life	117	Noun	833	289	rack	4273	Noun	11	7
right	124	Adverb	677	297	wrist	4675	Noun	15	13
look	137	Verb	422	238	lamb	4744	Noun	8	5
lot	171	Noun	242	135	rim	4851	Noun	10	9
room	190	Noun	439	172	ram	4917	Noun	3	3
long	196	Adjective	798	360	reef	5050	Noun	N/A	N/A
late	224	Adverb	233	161	lace	5892	Noun	13	9
law	257	Noun	275	117	raid	6017	Noun	11	8
road	264	Noun	158	81	limb	6023	Noun	4	4
wrong	377	Adjective	115	75	lime	6178	Noun	4	4
red	379	Adjective	197	125	rap	6180	Noun	5	4
light	384	Noun	263	161	rug	6337	Noun	5	5
read	444	Verb	226	130	robe	6574	Noun	12	8
list	492	Noun	103	73	lush	6665	Adjective	8	7
race	507	Noun	101	57	rip	6922	Verb	7	6
rock	584	Noun	57	41	rag	6971	Noun	6	6
led*	589	Verb	169	120	rouge	7605	Noun	1	1
lead	589	Verb	134	100	leech	N/A	N/A	N/A	N/A
reach	636	Verb	101	84	reek	N/A	N/A	N/A	N/A
raise	696	Verb	57	48	ling	N/A	N/A	N/A	N/A
rain	702	Noun	78	38	lash	N/A	N/A	N/A	N/A
lake	732	Noun	52	25	luge	N/A	N/A	N/A	N/A
ring	828	Noun	55	27	rune	N/A	N/A	N/A	N/A
rung**	828	Noun	1	1	rook	N/A	N/A	N/A	N/A
rice	910	Noun	34	20	lout	N/A	N/A	N/A	N/A
rate	1078	Noun	167	71	lob	N/A	N/A	N/A	N/A
laid**	1191	Verb	56	47	laze	N/A	N/A	N/A	N/A
leaf	1230	Noun	16	9	roam	N/A	N/A	N/A	N/A
lack	1565	Noun	91	74	roan	N/A	N/A	N/A	N/A
route	1691	Noun	54	36	loam	N/A	N/A	N/A	N/A
rare	1707	Adjective	51	41	luff	N/A	N/A	N/A	N/A
lane	1787	Noun	17	13	lair	N/A	N/A	N/A	N/A
rough	1830	Adjective	34	25	lit	N/A	N/A	39	35
root	1880	Noun	28	16	writ	N/A	N/A	13	4
rid	1897	Verb	21	18	Rick	N/A	N/A	8	6
lock	1994	Verb	17	16	loon	N/A	N/A	7	2
rush	2025	Noun	24	19	loom	N/A	N/A	6	6
rude	2112	Noun	9	7	rash	N/A	N/A	5	3
loan	2270	Noun	10	7	lick	N/A	N/A	4	4
raw	2405	Adjective	20	17	reap	N/A	N/A	3	3
load	2527	Noun	45	20	lewd	N/A	N/A	3	3
loose	2580	Adjective	46	38	lobe	N/A	N/A	3	3
lip	2860	Noun	21	19	loot	N/A	N/A	2	2
rob	2941	Verb	4	4	ruse	N/A	N/A	2	2
lag	3458	Noun	5	5	rhyme	N/A	N/A	2	2
rear	3523	Adjective	30	22	lug	N/A	N/A	2	2
lung	3843	Noun	42	13	leer	N/A	N/A	1	1
leap	3863	Noun	23	22	rot	N/A	N/A	1	1
lap	3975	Noun	21	19	lice	N/A	N/A	1	1
lid	4008	Noun	13	6	rife	N/A	N/A	1	1
leak	4101	Noun	8	6	rake	N/A	N/A	1	1

Note. JACET8000_rank = Ranking of word frequency in the JACET8000 list, JACET8000_POS = Parts of speech of a given token defined in the JACET8000 list, AmE06_frequency = The raw frequency of the token in AmE06 (max. = 1,017,879 tokens), AmE06_range = The number of sources that a given token is found (max. = 500 sources). *led = JACET8000 data show ranking and POS for *lead*. **rung = JACET8000 data show ranking and POS for *ring*. ***laid = JACET8000 data show ranking and POS for *lay*.