Cross-Linguistic Influence of Similar Phonological Contrasts for Heritage Bilingual Children in the United Kingdom

Supervisors: Dr. Sam Kirkham and Dr. Claire Nance
Submitted for examination for the award of MPhil Linguistics

Maximilian Topps
Lancaster University

31st March 2022

Author: Maximilian Topps
Date: 31st March 2022
Award: MPhil Linguistics
Supervisors: Dr. Sam Kirkham and Dr. Claire Nance

Contents

List of Figures

List of Tables

Abstract

Children who grow up learning a heritage language at home, and a majority language from school and peers, represent a case of bilingualism in which there is less distinction between a "first" and "second" language. This thesis investigates the similar contrasts within two separate heritage language communities, in order to observe patterns that occur between heritage language groups.

A contrast pairing for each pair of languages was chosen: one for English-Polish heritage bilingual children and one for English-Chinese children. It was hypothesized that there would be cross-linguistic influence present, and that the acoustic cues of the English contrast would influence the production and perception of the heritage language contrast.

Children between ages 7;0 and 9;6 performed a picture-naming task and a perceptual categorisation task, involving a phonetic continuum across a lexical minimal pair. The data include heritage bilinguals as well as control samples of monolinguals of each language.

Results suggest that the heritage bilinguals reliably produced and perceived each contrasts. A degree of cross-linguistic influence was also present, and was seen both spectrally and temporally. In production, a comparison of vowel duration between heritage speakers and monolingual English speakers showed a significant effect ($p < 0.001$), with heritage speakers relying less on the temporal cue. This was not repeated in perception. The effect is evaluated with respect to phonetic assimilation, presenting evidence for a shared phonetic space in which mutual influence occurs.

Acknowledgements

Chapter 1

Introduction

The thesis will address the subject of phonological and phonetic acquisition in heritage bilinguals. In doing this, the investigation will involve two main foci that will be considered. Firstly, the bilingual acquisition patterns of heritage bilinguals when compared to other forms of bilingualism, and secondly, the potential interactions and interfaces between production and perception during bilingual acquisition. This introduction will briefly outline these subjects, and will be followed with a literature review that discusses the theoretical framework in more detail.

An infant growing up in a bilingual environment must learn to process and produce more than one language. As a result, the child must develop strategies for each language with which they are in contact. Evidence suggests that bilingual infants develop unique phonetic strategies that are unlike monolingual peers, potentially with the goal of discriminating such subtle categorical differences (Højen and Flege, 2006). Historically, research on bilingual phonological acquisition has investigated bilinguals who have begun learning an L2 after childhood, or bilinguals who have been exposed to two languages for their whole lives. For example, contemporary models of L2 learning such as the SLM (Flege, 1995) and PAM-L2 (Best and Tyler, 2007b) have been constructed primarily on the behavior of traditionally sequential bilinguals, but are often used to understand bilingual phonology. Whilst recent literature has taken more interest in heritage bilinguals, there still exists a substantial deficit in heritage bilingual research, as noted by Simonet (2016).

Heritage bilinguals usage and exposure for each language may ebb and flow over their lifetimes. Due to this, they represent a group of interest for research that investigates the relationship between multiple languages that exist within an individual's phonoglogical repertoire. The interaction between the phonological systems of the heritage language and majority language during childhood development result in cross-linguistic influence (CLI). This may process may also occur differently between groups of heritage bilinguals, as one pair of languages would naturally differ in ways that another pair may not. For example, one pair of languages may differ in the presence of tonality, while another pair may differ

in their quantities of tones. Considering the number of potential language pairs, and the number of ways they can differ, the vectors for cross-lingusitic influence are difficult to make generalised statements about. Therefore, this project has prioritised an approach that considers multiple heritage demographics.

Early L2 learners occupy a blurred space between what is considered simultaneous and sequential bilinguals (Au et al., 2002; Polinsky and Kagan, 2007; Cabo and Rothman, 2012). An example of individuals who fall within this ambiguous group are children growing up in immigrant or indigenous families who speak a heritage language at home but acquire a majority language through education and pre-school. Across bilingualism literature and policy, the term heritage language has been applied to highly diverse range of linguistic situations (Polinsky and Kagan, 2007). Polinsky and Kagan (2007) state that defining a heritage language is a matter of its social status as a minority language, including minority languages spoken by immigrant and indigenous families (Kelleher, 2010; Montrul, 2018).

Using some terminology can be misleading, and has historically been conflated with incomplete acquisition (Kupisch and Rothman, 2018). This issue is exacerbated by usage of dominance as a tenet of heritage bilingualism modelling, which connotes incomplete acquisition, as noted by Kupisch and Rothman (2018). As well as this, the theoretical differences between heritage bilingualism and simultaneous bilingualism are fuzzy at best.

For heritage bilinguals, the status of either language as an L1 or L2 (in accordance with common bilingual terminology) is ambiguous. All three of the models discussed - the SLM and PAM-L2, assume the L1 to be an initial state of L2 acquisition, from which assimilation, adaptation, or copying is performed. However, it is presumed in those frameworks that the initial state only has knowledge of the L1, and that L2 is introduced afterwards, with less exposure and practice. This is not necessarily true in simultaneous bilingualism or heritage bilingualism. For example, a child starting school may have minimal exposure to the majority language if they had learned another language at home. This adds complexity to the quantification of heritage language and majority language exposure and use. In terms of L1 and L2, very common heritage bilingualism scenarios bring into question what the "initial state" actually is for a heritage learner during schooling. This is particularly relevant in childhood, during which linguistic input from family members is still high. Exposure such as this can allow the phonological inventory of the child to become attuned to the contrasts of both languages similar to that of a simultaneous bilingual.

As well as the motivation to research heritage bilinguals for linguitsic theory, there is also an important social reason to research this demographic. As pointed out by Safford and Drury (2013), the infrastructure in the United Kingdom, including education and healthcare. Curdt-Christiansen (2020) identifies the sociopolitical difficulties faced by families when they are users of a minority language. Simonet (2016) refers to Rao and Ronquest (2015), who identify a specific lack of research that seeks to advance heritage bilingual pedagogy for Spanish-English bilinguals, but these sentiments may be echoed for many other heritage populations. These attitudes are corroborated by Arnot et al. (2014),

who report "language exclusion" for EAL children. This can result in the usage of the term heritage bilingualism to have a political meaning as well, regarding language status, and the perpetuation of some languages being perceived as lesser (Leeman and Serafini, 2016). Thus, the definition moving forward should be narrowed further for the purpose of a useful research demographic, and to avoid political implication.

There is a dearth of support and information for the families of heritage language learners. For example, Yu (2012), in a study that interviewed the parents of heritage bilinguals, reported: "Interviewees reported that they adopted language practices perceived to be advantageous to intervention access and wellness. They valued Chinese language but did not pursue its use if it was believed to hinder the children's overall development of English acquisition".

It is common for bilinguals to have a degree of asymmetrical proficiency and usage of the languages that they speak (Gertken et al., 2014). This is often presumed to be the case for heritage speakers or speakers of a minority language though may not be the case. A language used by the community surrounding the speaker and is encountered more often in the media they consume, and gets more practice. In cases of cross-linguistic influence a directional tendency for the results may be predicted by such asymmetrical exposure. Therefore, a potential way of investigating this case for heritage speakers is to observe their production and perception of cross-linguistically similar contrasts, and assess to direction, and degree to which, assimilation has taken place.

When investigating bilingual acquisition cases such as these, we must consider the processes by which the phonological inventory is formed. Building and attuning a phonological inventory is a task that involves both producing and perceiving language. However, research is often limited to investigating one or the other. This is not to imply that researchers have failed, or that researching each of these processes individually is not useful. It is important to research both of these processes independently. For example, second language speech perception experiments have demonstrated the ability for language users to subconsciously create and adjust phonetic boundaries for linguistic purpose (Escudero and Boersma, 2004; Escudero, 2005) whilst production studies have suggested that sufficient practice and exposure can allow even later second language learners to produce monolingual-like targets (Vihman, 2017). It is a significant challenge to conduct research that involves both production and perception. This is mostly due to the differences in mechanisms between production and perception, which create difficulties in experimental design and in comparative methods.

The aim of this project is to contribute to the understanding of bilingual phonology in a heritage setting, and to provide experimental accounts of the interaction between heritage bilingual systems. In doing so, models of bilingual phonology can be developed within the broader bilingual spectrum. A second contribution of this project is the descriptive account of young heritage bilingual learners, which may benefit clinical and pedagogical strategies.

With this in mind, the research questions of this project are as follows:

- What degree of cross-linguistic influence is present in young heritage bilinguals?

- Do production and perception align during heritage language acquisition?

- Can the existing models of bilingual phonological acquisition predict these outcomes?

Chapter 2

Literature Review

This chapter will offer an overview of the present literature and theories relevant to the thesis project. Among the concepts that will be addressed are: the classification of bilinguals, cross-linguistic influence (CLI), similar contrasts, acoustic cues, and models of second language acquisition.

2.1 Bilingual Classification

In the simplest interpretation, a bilingual individual is one who can speak and understand two languages (Moradi, 2014). The presence of the "bi-" prefix suggests that the word refers specifically to individuals who are proficient in only two languages, but it is widely understood as referring to individuals proficient in more than one language (Wei, 2013). The potentially misleading terminology arises from the difficulty of defining words such as "bilingual" in the first place. Speakers described as bilingual occupy a very broad range – from those with exposure to a second language up to full competency (García and Wei, 2014). Exploring contemporary literature on the subject makes it clear that quantifications of n-lingualism are not discrete (Flege et al., 2003; Stow and Dodd, 2003).

The broad definition of bilingualism necessitates a set of more specific nomenclature that attempts to address the differences between bilingual individuals. This variation may occur on many different scales, including exposure and competence for each language, social contexts in which each language is used, and similarities and differences between the systems of each language. Contemporary literature separates bilinguals into broad categories: simultaneous (early) and sequential (late) (Bialystok and Hakuta, 1999; Moradi, 2014). However, these categories contain extremely diverse populations, and the unique experience of language that each person acquires is likely to form their own "individual bilingualism" (Moradi, 2014; Baker and Jones, 1998).

Historically, the age of immersion (AOI) has been used as a theoretical predictor for foreign accent and a heuristic for exposure (Johnson and Newport,

1989; Dekeyser, 2000; Munro and Mann, 2002). From this, the age at which an individual begins to learn a language has been considered the primary factor – sometimes the only factor – in classifying bilingualism, and this approach is apparent in the nomenclature (Genesee, 2015).

2.1.1 Simultaneous Bilingualism

Generally, this refers to an infant who has experienced simultaneous exposure and usage of two languages during the "optimal period", which refers to approximately the first three years of life (Sebastian Galles et al., 2005; Mccarthy et al., 2013). It is during this period that neuroplasticity is at is maximal level due to the ongoing development of the brain's physiology as well as a child's cognitive skills (Flege, 1995; Iverson et al., 2003). The presence of two interfacing phonological systems during this period causes bilingual development to exhibit differences to that of monolinguals.

Early bilingualism is sometimes referred to this as bilingual first language acquisition (BFLA) (De Houwer, 2009). This term emphasises the high propensity of the simultaneous bilingual to achieve L1-like proficiency in both or all their languages (Tsukada et al., 2005). Simultaneous bilingualism is typically observed in households in which parents are from different language backgrounds (De Houwer, 2009), or in bilingual communities (Hakuta, 1986).

The development of contrasts and categories is varied between monolinguals and bilinguals (Bosch and Sebastián-Gallés, 2003; Sebastián-Gallés and Bosch, 2009; Garcia-Sierra et al., 2011). Studies using both existence contrasts such as Garcia-Sierra et al. (2011), which observed /t/-/d/ contrasts in 6-9 month-old Spanish-English simultaneous bilingual infants, as well as those using non-words, show that some contrasts may be slower to establish for bilingual infants. These differences are also present in production, and may also be more persistent in production (Vihman, 1996).

2.1.2 Sequential Bilingualism

Sequential bilingualism may be a result of situations such as, but not limited to, the formal acquisition of an L2 language through schooling, or immigration to a region in which there is a majority language other than an individual's L1 (Baker and Jones, 1998). Montrul (2008) suggests that a sequential bilingual can be broadly defined as acquiring an L2 later than the "optimal period" of the first few years of life, and beyond the systematic implementation of the L1. The span of such a definition may cause discursive issues. However, the end of the optimal period, and the degree to which "L1 filter" (the initial framework through which a second language is interpreted) becomes effective, is not an easily defined boundary. As a result, many language learning trajectories may not be predicted by the existing L2 acquisition models. Montrul (2008) points out that the beginning of the L2 acquisition may occur during childhood, late adulthood, or anywhere in between. Sequential bilingualism is often characterised by the presence of a "foreign accent" (Flege, 1995; Piske et al., 2001) and "incomplete

acquisition" Montrul (2008). This is theorised to be a result of an L1 filter through which an L2 is understood (McCarthy, 2015; Iverson et al., 2003).

As a result of sequential bilingualism, these speakers are likely to exhibit a "foreign accent", which is predicted by the SLM detailed by Flege (1995) and Piske et al. (2001). The SLM is discussed in more detail in Section 2.4.2. Due to the robust representation of phonetic contrasts for an individual's L1, the L2 learner is likely to employ the nearest L1 approximation of an L2 gesture. It has been shown that L2 speakers are able to mitigate foreign accent and establish new categories with explicit training as opposed to implicit learning (Vihman, 2017). This is predicted by the observations that phonologies are never complete and that language users continue to refine and adapt their phonological representations well into their adult lives (Flege et al., 2003; Flege, 2007). Further evidence for the continued malleability of sound categories can be found in the influence of L2 acquisition on the phonology of an L1 (Chang, 2009; Simonet, 2014).

2.1.3 Heritage Bilingualism

Classifying heritage bilingualism has been a matter of debate in contemporary literature, without a consensus on definition, which is pointed out by Benmamoun et al. (2013) and Peyton et al. (2008).Benmamoun et al. (2013) move towards a definition of heritage bilingualism similar to that which is discussed by Valdés (2005) and Chang et al. (2011) as a speaker who mostly uses a majority language but whose initial language exposure (which according to some definitions may be considered an L1) was to a heritage language. This definition should be narrowed further for the purpose of a useful research demographic. This study considers heritage bilingualism in the form of young children who are still in active use of the heritage language at home, but also attend a primary school taught in the majority language.

These speakers have been grouped under different terminological categories such as English additional language (EAL) in pedagogical policy of the Anglosphere (Strand, 2016; Anderson, 2008), and early second language acquisition (ESLA) (De Houwer, 2011). These terms have a consistent underlying meaning: a child whose primary language exposure at home, and before schooling, is a minority language. Typically, a heritage bilingual will become attuned to the heritage language prenatally and during infancy, but through education, media, and socialisation, will learn the majority language and use it more often (Benmamoun et al., 2013).

Whilst a simultaneous bilingual who undergoes bilingual first language acquisition is likely to use each of their immersed languages to some regular degree, a heritage speaker may exclusively use one language for certain relatives, or in one distinct situation such as a place of worship (Ruiz-Felter et al., 2016). However, heritage speakers remain distinct from sequential bilinguals in many cases pertaining to their non-trivial exposure to both languages during early childhood (Benmamoun et al., 2013). Non-trivial exposure comprises linguistic interaction with parents, family friends, and media such as television which would occur

with some regularity and child engagement. With parents, this is likely to be in the heritage language, whilst media is likely to be in the majority language.

Exposure such as this can allow the phonological inventory of the child to become attuned to the contrasts of both languages similar to that of a simultaneous bilingual. Chang et al. (2011) found that heritage speakers of Mandarin were highly proficient at maintaining both cross-linguistic and within-language contrasts of Mandarin and English plosives. Similarly, Chang (2016) found that heritage speakers of Korean performed with L1-like proficiency in the perception of Korean unreleased stops and with better than L1-like proficiency when perceiving English unreleased stops. Tamburelli et al. (2015) corroborates the ability of heritage learners to be L1-like in both languages by demonstrating acceleration in the acquisition of fricative clusters by Polish heritage bilinguals.

Heritage bilinguals may exhibit behaviours and patterns similar to both simultaneous bilingualism and sequential L2 acquisition (Montrul, 2011; Fowler et al., 2008; Polinsky, 2018; Fabiano-Smith and Barlow, 2010; Rothman, 2009). For example, foreign accent or perceptual assimilation have been observed in both heritage speakers and sequential L2 learners (Stangen et al., 2015; Flege et al., 1999; Yeni-Komshian et al., 2000). It is intuitive to suggest that this vast variation is a result of the heterogeneity of heritage bilingualism, even within narrow groupings of age-matched children in factors such as language input and practice.

The complexity of the relationship between the home language and the majority language brings into question what the "initial state" is for a heritage learner during schooling. This is particularly relevant in childhood, during which linguistic input from parents is still high. Considering this, observation of the similar scenario, and consideration of cross-linguistic influence may present evidence towards discerning this relationship. Exposure such as this can allow the phonological inventory of the child to become attuned to the contrasts of both languages like that of a simultaneous bilingual.

2.2 Cross-Linguistic Influence

Cross-linguistic influence (CLI), also known as Language Transfer, is a process by which a language user may incorporate features from one language when using another (Smith and Kellerman, 1986). This may include knowledge such as the psychological organisation and structure of phonological units, the perception of multiple *(Lx)* segments as a single *(Ly)* segment, or phonetic and prosodic patterns that form the basis of accent features (Jarvis and Pavlenko, 2008).

It has been theorised that the characteristics of cross-linguistic influence arise from a shared phonological system within the psychology of the speaker, and the asymmetry in phonetic inventories between languages (Tsukada et al., 2005). Examples of evidence are the systematic variations in accented speech from bilinguals with a shared L1, and perceptual assimilation of L2 categories to L1 categories, for example Japanese speakers observed by (Flege, 1995). It follows that a heritage bilingual may be susceptible to the same process. It

should be noted that, whilst cross-linguistic influence is an easy explanation for such variation, another consideration for heritage language variation is that sociolinguistic trends and patterns are not always identical to those of a native or monolingual demographic (Valdés, 2005). In these models, phonological knowledge exists in the same dynamic system, and is activated flexibly depending on context (Grosjean, 1997; Herdina et al., 2004).

Within the shared phonological model, Grosjean (1997) defines dynamic states of language activtion named language modes. Language modes are defined by Grosjean (2013) as "[...] the state of activation of the bilingual's languages and language-processing mechanisms at a given point in time". One's place on the continuum at a given time is defined by the degree to which one or both languages is activated, with a monolingual end representing activation of mainly one language, and a bilingual end representing more balanced, simultaneous activation of both languages.

The degree to which a language is activated is controlled by the linguistic and social environment surrounding a bilingual (Grosjean, 2013). The changes are not generally conscious decisions, but are subconscious reactions to effects such as "their language proficiency, their relationship, their language-mixing habits and attitudes toward language mixing, their mode of interaction, and so forth" (Grosjean, 2013). In a shared phonological system, the use of language modes would enable a bilingual to tune their perceptual expectations and in order to maximise their understanding of spoken language in the vicinity, using the previously mentioned information.

Other models suggest interdependence between phonological systems, but not unity (Paradis and Genesee, 1996; Hossein Keshavarz and Ingram, 2002). Crucially, in neither of these paradigms are bilinguals defined as "two monolinguals in one" - the interaction between their languages results in a unique linguistic state. Given the interactions between cohabiting linguistic systems during the acquisition process, research can bring attention to the processing and organisation of phonological knowledge as acquisition occurs.

Whilst the debate of the degree to which transfer or unity between phonological systems remains open, the proposal that there is at least some degree of mutual influence is widely accepted (Simonet, 2016). This is further supported by Antoniou et al. (2011) and Sancier and Fowler (1997), who exemplified dynamic shifts in cross-linguistic influence which appeared to be the result of recent exposure, causing phonetic targets to more closely resemble the language of recent exposure. This influence may manifest in considerably different ways including boundary shifts, categorical narrowing, and complete categorical assimilation (Clarke and Garrett, 2004). Cross-linguistic links are complex and have varied results (Amengual, 2012).

Historically, research on cross-linguistic influence has primarily observed the case of L1 influence on the L2 within sequential bilinguals, and it is from these observations that many L2 language learning models are derived. For example, So and Best (2010) suggested that similarities and differences between tonal systems of an L1 and L2 played "critical roles in the perception of non-native tones". Other research has diversified this focus and considered bidirectional influence.

This research includes Miller et al. (2006), who observed that oral language skills in Spanish and English bilinguals were factors in predicting reading both between and within each language. These examples are not limited to phonetics and phonology: another example is Zhou et al. (2021), who observed bidirectional influence in Cantonese-English bilinguals. Zhou et al. (2021) states: "[...] the direction of influence goes from the weaker to the stronger language and from the stronger to the weaker language" in the case of grammatical object ommission.

Bidirectional cross-linguistic influence is an area of interest because it offers insight into the processes behind language acquisition and phonological processing Schertz et al. (2019). For speakers whose use and exposure to their languages is generally variable over time, the effects may be particularly salient. Interactions between an individual's languages are especially relevant in contemporary models of bilingualism that propose interdependence, or total unity, between an individual's phonological systems. Bidirectional cross-linguistic influence presents itself as a research focus because it offers insight into the processes behind language acquisition and phonological processing (Schertz et al., 2019).

2.2.1 CLI in Heritage Bilinguals

Previous experiments suggest that heritage speakers are able to achieve L1-like proficiency in both the majority and heritage languages. Chang et al. (2011) found that heritage speakers of Mandarin were highly proficient at maintaining both cross-linguistic and within-language contrasts of Mandarin and English plosives. Similarly, Chang (2016) found that heritage speakers of Korean performed with L1-like proficiency in the perception of Korean unreleased stops and with better than L1-like proficiency when perceiving English unreleased stops. Tamburelli et al. (2015) corroborates the ability of heritage learners to be L1-like in both languages by demonstrating acceleration in the acquisition of fricative clusters by Polish heritage bilinguals.

On the other hand, some research suggests that contrasts may develop differently compared to monolingual peers, or show accentedness that reflects patterns of the majority language (Au et al., 2002; Best et al., 1994; Newman and Wu, 2011; Amengual, 2018; Nagy, 2018). For example, Stangen et al. (2015) found that, in a study of Turkish heritage speakers in Germany, a foreign accent was present for most speakers in one language regardless of age of onset, which suggests different learning trajectories for heritage speakers owing to the social status of the heritage language. When observing cognates produced by Spanish heritage speakers, Amengual (2012) found that phonetic interference was present in vowel onset time (VOT), which supports the presence of cross-linguistic influence. Lee and Iverson (2012) demonstrated that young bilinguals of Korean and English showed different stop productions to age-matched monolingual peers at five years of age but not at ten, which is interpreted as a mixed system that develops into two separate ones.

In light of this research, it can be asserted that heritage bilinguals may exhibit cross-linguistic influence and patterns similar to both simultaneous bilin-

gualism and sequential L2 acquisition (Montrul, 2011; Fowler et al., 2008; Polinsky, 2018; Fabiano Smith and Goldstein, 2010; Rothman, 2009). This complicates the classification of different types of bilingualism into neat boundaries, and suggests that, while some behaviour is shared, not all behaviour is. In terms of modelling bilingualism, this is encouragement to test many different individual bilingual scenarios.

2.2.2 The Similar Scenario

Categories or contrasts with phonetic similarity across languages are likely to show evidence of cross-linguistic influence. This may be a result of difficulty distinguishing between one language's phonemes and another's, or a lack of utility in doing so if the individual can understand and be understood regardless. This is exemplified in what Escudero and Boersma (2004) describe as the "similar scenario". The similar scenario occurs as a two-to-two mapping between L1 and L2, or heritage and majority in the case of heritage bilinguals, sounds (Escudero, 2005; Chang, 2009). Figure 2.1 shows examples presented by Escudero (2005) that depict a visual representation of the two-to-two mapping.

Such a relationship occurs when categories from one language share articulatory and acoustic qualities with those of another, which suggests phonetic similarity. Phonemes in the similar scenario which are mapped to each other may or may not have the same IPA symbol, but are nevertheless generally distinct in some way (Escudero, 2005). An inexperienced or naive listener may have difficulty identifying the acoustic differences between similar contrasts. On the other hand, to some experienced listeners the phonemes in the similar scenario may appear to have easily discernible phonetic differences (Chang, 2019). The similarity of two speech sounds is complex to define, but in the case of this study is defined through phonetic and phonological correlates. For example, similar formant frequencies, and a contrast incorporating these moving in the same direction.

Figure 2.1: Similar scenario examples as presented by Escudero (2005)

Target L2	L1	Target L2	L1	Target L2	L1
Sc. English	Spanish	C. French	C. English	Spanish	Dutch
/i/——/i/		/ɛ/——/ɛ/		/u/——/u/	
/ɪ/——/e/		/æ/——/æ/		/o/——/ɔ/	

Results of similar sound interaction may include both dissimilation through category narrowing and total categorical assimilation (Flege, 1995). The sharing of cognitive resources for different languages and their phonologies introduces more complex cue-weighting challenges for the heritage bilingual (Escudero, 2005).

The sometimes contradictory empirical evidence regarding the interdependence of articulation and perception also encourages exploration (Schmitz et al.,

2018). Studies of perception and production within young heritage bilinguals have been carried out in the past on a variety of languages, and have provided valuable insight into bilingualism (Lee and Iverson, 2012; Oh et al., 2011; Turner et al., 2014; Schertz et al., 2019).

Likewise, whilst cross-linguistic influence may be bidirectional, the tendency for categorical contrasts is to narrow, assimilate, or shift rather than to disappear in the presence of a similar L2 contrast. While exposure and practice to both languages is consistent and high, one may expect a contrast to be maintained.

2.2.3 Acoustic Cue-Weighting

Among the ways in which languages differ is their implementation of acoustic cue weighting, making it a suitable test case in cross-linguistic studies. An acousitc cue in this case is a feature of the acoustic signal such as a formant frequency. The weight of such a cue is its relevance in determining the phonological sense of the acoustic signal. Speech sounds are multivariate acoustic signals and listeners distinguish between phonemic and phonetic categories by paying attention to the acoustic properties of the speech signal, which allows for perceptual categorisation (Escudero, 2005). Not all of the acoustic properties of a signal may carry linguistic value. Indeed, the salient properties of the speech signal may be vastly different across dialects and languages. In any case, it is often a small subset of these acoustic properties that constitute the necessary information for categorisation to take place (Tyler et al., 2014). These are the acoustic cues by which a category is distinguished. Furthermore, the cues within this subset may vary in their salience for the purpose of categorisation. Contrasts within a language may rely on different cue weighting strategies for reliable categorisation. Acoustic cues exist across multiple dimensions, and include spectral characteristics such as formant frequencies and pitch, as well as the temporal characteristics such as duration or voice onset time.

However, it has been shown that attention to specific cues is not constant throughout language acquisition. Ohde and German (2011) demonstrated that children and old as 11 were not adult-like in their perceptual categorisation, and relied on different cue-weighting strategies. Ohde and German (2011) attributed the results to the children assigning higher relative cue weighting to formant transitions. Hazan and Barrett (2000) corroborate the difference between adult and child perceptual cue weighting (Mayo and Turk, 2004; Nittrouer, 2005). There also exists variation in cue weighting between monolinguals and bilinguals (Escudero and Boersma, 2004), who must learn to incorporate the contrasts of both or all of their languages. In doing so, this can lead to assimilation or dissimilation between contrasts that are similar between each language, leading to increased or decreased reliance on certain cues when compared to monolingual counterparts.

2.3 The Relationship Between Age and Acquisition

The "critical period" hypothesis was detailed by Lenneberg (1967). It posits that there is a biologically-defined window that is ideal for acquiring language. After this "critical period", language learning becomes significantly more difficult Lenneberg (1967). The critical period hypothesis has undergone a large amount of research and debate since its introduction. Over time, more serious criticism has been brought forward against the tenets of the hypothesis.

Whilst many perceptual and productive contrasts are established within the first few years of language use, development of these categories continues long after childhood (Nittrouer, 2005). It is suggested that an individual's phonology remains malleable into adolescence and continues to develop throughout one's lifespan (Berken et al., 2017; Hazan and Barrett, 2000). However, as the individual ages, the flexibility of their phonetic capabilities is diminished (Flege et al., 2003). Explanations for this are both cognitive and neurophysiological, with decreased neuroplasticity as a factor as well as what may be considered statistical reinforcement of categories that creates robust prototypical categories that are difficult to manipulate (Berken et al., 2017).

It is widely established that infants learn to perceive and identify phonemic contrasts before they are able to produce them (Eimas et al., 1971). Intuitively speaking, infants are capable of receiving sensory input and attempting to understand or categorise it far earlier than they can acquire the motor skills necessary to reproduce them. Empirical evidence such as the famous Fis phenomenon support this line of thought (Brown and Berko, 1960; Smith et al., 1973). It is shown that infants are universal listeners (Berken et al., 2017). This is corroborated by findings that show the perceptual representations within a child's lexicon lack detail when compared to those of adults (Altvater-Mackensen and Fikkert, 2010). An inference that can be made is that phonetic skill, as well as phonological knowledge, is derived from experience and practise.

Past the first four years, it was once theorised that phonetic development more or less ceases, and an individual's linguistic inventory solidifies quickly (Lenneberg, 1967). Whilst there is evidence that, as brain development slows, phonetics and phonology become less flexible over time, it does not cease to change (Bialystok and Hakuta, 1994; Raumolin-Brunberg, 2009). As late as adolescence, children who have been placed in a foreign language environment through immigration have demonstrated the ability to create new phonetic contrasts and achieve native-like phonetic proficiency in an L2 in which they are immersed (Birdsong, 2019; Bongaerts, 1999).

The necessary articulatory skill for speech production is acquired through constant practice. This begins with non-linguistic vocalisations commonly referred to as cooing. This is followed by babbling, and with the acquisition of lexical knowledge, fully realised words. This practice is what allows children to form bonds between patterns of acoustical information with the articulatory control that they build in practice (Messum and Howard, 2015). Messum and

Howard (2015) name this process Mirrored Equivalence. Practicing the production and perception of speech informs the formation of phonological neighbourhoods (Gahl and Strand, 2016). Whilst an infant may understand simple patterns of sound as speech, and lexical units that they encounter frequently, they have yet to fully network their lexical, articulatory, and phonological knowledge.

As described by Byun (2016), children may exhibit an "articulatory mapping" (henceforth A-map) in their phonological patterns. The premise of the A-map, as posited by Byun (2016), is that there exists conflict in child speech motor planning between the desire to imitate adult productions of speech sounds, and the child's own practised gestures that they can perform reliably. The various goals of speech production, and the various methods through which they are achieved, are certainly complicating factors when attempting to model phonological development.

2.4 Acquisition Models

Models of L2 phonological acquisition have been built with various interpretations of bilingualism and cross-linguistic influence in mind, and have often been derived from research on sequential bilinguals. The three models of L2 phonological acquisition addressed within the framework of this thesis are the SLM (Flege, 1995), the PAM-L2 (Best and Tyler, 2007b) and the L2LP (Escudero, 2005). Heritage bilinguals may exhibit behaviours and patterns similar to both simultaneous bilingualism and sequential L2 acquisition (Montrul, 2011; Fowler et al., 2008; Polinsky, 2018; Fabiano-Smith and Barlow, 2010; Rothman, 2009). Therefore, these models may not accurately predict heritage learning outcomes accurately, especially considering the heterogeneity of heritage bilingual experience. However, the underlying linguistic processes described in L2 acquisition models are likely to provide a framework for investigating the modelling of heritage bilingualism.

2.4.1 Native Langauge Magnet Theory

The common tenet of the second language acquisition models below is that there exists an initial "L1 state", which functions as the set of toosl from which L2 competence begins to arise. The L1 state is a result of perceptual conditioning decsribed by Kuhl (1993) as the Native Language Magent (NLM). The NLM causes infant linguistic perception to become attuned to discerning the phonetic input of the first 6 months of infancy Kuhl (1993), which is in many cases limited to the L1 of the infant. The NLM was expanded by Kuhl et al. (2008) to the NLM-e, which extends the model beyond the first 6 months, and includes consideration for infant-directed speech (ID). The increased intelligibility and emphasis during ID is theorised by Kuhl et al. (2008) to encourage the "magnet" to identify L1 categories. For many heritage bilinguals, the influence of ID within the NLM may support outcomes more similar to sequential bilinguals, given the heritage langauge input during infancy.

2.4.2 Speech Learning Model

The Speech Learning Model (SLM) was derived empirically from studies of foreign accentedness and second language perception (Flege, 1995). It is a model that hypothesises the existence of phonetic categories stored in long-term memory.

Figure 2.2: Demonstration of a unimodal (one category) and bimodal (two categories) input distribution
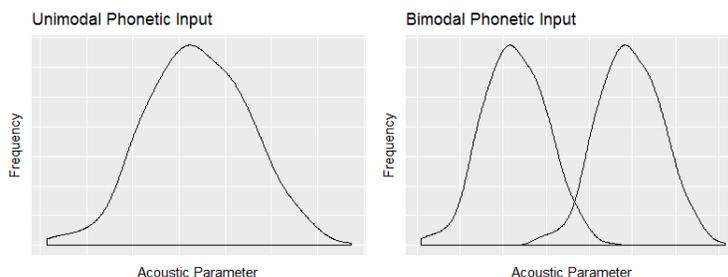
Phonetic categories according to Flege (1995) are prototypical representations of a linguistic unit, stored in long term memory as phonetic detail. Categories in the SLM are statistically derived from incoming speech, and categorisation occurs when there are multiple discernible maxima in the distribution of one or more phonetic features of incoming speech signals. In Figure 2.2, the scenario on the left would result in the formation of one phonetic category over time, whilst the scenario on the right would divide into two over time. The information that comprises a phonetic category is the acoustic detail that can be used to describe its prototypical form, and that which differentiates it from other categories. Thus, the SLM posits that distributional learning occurs perceptually at the phonetic level. When perceiving speech, the incoming signal is considered with regards to its phonetic closeness to the distributional maxima that have formed within a listener's long term memory. The signal is then categorised and ascribed linguistic value on this basis.

Revised Speech Learning Model (SLM-r)

During the course of this research project, the SLM has undergone revisions which are detailed by (Flege and Bohn, 2021). The SLM-r is built upon the same fundamental linguistic mechanism as the SLM. Specifically, the core phonetic unit in the SLM-r is still the category (as opposed to the contrast), and each phonetic category is developed and reinforced over time given based on statistical properties of an individual's language exposure. The SLM-r modifies parts of the SLM to reflect changes in the understanding of bilingual acquisition since the SLM's inception. One of these changes is emphasis on the individual bilinguals, instead of differences between groups of speakers (Flege and Bohn, 2021).

The original SLM suggested that the accuracy of L2 production was constrained by the accuracy of L2 perception, and that perception placed an "upper limit" on production (Flege and Bohn, 2021). This revision is notable within the context of this study, given that the Flege and Bohn (2021) explicitly states that "segmental production and perception coevolve without precedence".

In accordance with the previously mentioned criticism of the critical period hypothesis, the SLM-r has replaced the tenet of "age" with "category precision". The hypothesis of category precision states that "the more precisely defined L1 categories are at the time of first exposure to an L2, the more readily the phonetic difference between an L1 sound and the closest L2 sound will be discerned and a new phonetic category formed for the L2 sound" (Flege and Bohn, 2021). In other words, the more reinforcement and statistical tuning a category has undergone, the less likely a similar L2 sound will be assimilated by it.

Another relevant change in the SLM-r is that the original SLM raised questions of whether sufficient experience with an L2 would lead to monolingual-like results eventually, but Flege and Bohn (2021) states that "it seems evident, at least to us, that L2 1earners can never perfectly match monolingual native speakers of the target L2".

2.4.3   Discussion of the SLM

In the context of bilingualism, the SLM requires that the initial state of L2 learning is, as described by Flege (1995), a "grid" through which L2 sounds are interpreted. This follows logically, as the phonetic categories that have formed in a listener's long term memory do not require "empty space" between them. Instead, incoming speech is fitted to its most appropriate category. The SLM hypothesises three cases for this filter-like L1-L2 interaction. The consequences of these cases are derived from the statistical method of categorisation.

Flege (1995) and Flege and Bohn (2021) suggest that an L2 sound that is sufficiently different to existing L1 phonetic categories such that it forms a new distributional maximum is likely to form a new phonetic category. The propensity for new category formation is described as a function of the perceptible phonetic difference between the closest L1 categories as well as the age of the listener. According to Flege et al. (1997), new L2 phonetic categories can be formed "readily" until approximately six years of age in typical children. As age increases, the space for new sounds to be categorised diminishes, which is a potential result of the gradual statistical strengthening of L1 phonetic categories.

An implication of this behaviour is that perceptual factors are a large predictor in the degree of accentedness for an L2, which according to the SLM is a direct result of L2 sounds being indistinguishable to L1 categories for the listener. In this case, the targets of production are the phonetic categories formed by the speaker's perceptual experience. Thus, without the knowledge of separate categories, a speaker uses those most appropriate from their L1.

As an extension of accentedness, which may be considered to be crosslinguistic influence from L1 to L2, the SLM suggests that the reverse may occur. Once again, this result can be intuited from the possible outcomes of similar

L1 and L2 sounds. In the event that a new category is formed from an L2 sound, there may be a requirement for similar L1 categories to become more specific in order to accommodate the distributional peaks of the new category. Alternatively, in the event that a similar but perceivably different L2 sound is assimilated into an L1 category, this may result in the widening of the L1 category as an effect of the new statistical input.

2.4.4 Perceptual Assimilation Model

The Perceptual Assimilation Model (PAM) was initially described by Best (1995) in order to describe naïve perception of non-native speech sounds. Later, the PAM was adapted to accommodate the perceptual mechanisms of sequential bilinguals by Best and Tyler (2007b). It is a direct realist model, as opposed to the representative realist approach of the SLM. This implies that there is no long-term storage of phonetic exemplars in memory due to the assumption that the acoustic signal already contains phonemc information. Empirically, the observations of the PAM corroborated those of the SLM regards to the assimilation of non-native speech sounds by those of the L1 (Best and Tyler, 2007b). However, the epistemological and phonological differences between these models may result in differing predictions.

The central tenet of the PAM-L2 is perceptual assimilation, a process by which a new sound is understood in terms of existing L1 exemplars. When perceiving a new sound, the listener will compare the sound to existing exemplars within their established phonological space, then, its closeness to an ideal exemplar of the L1 category will decide how it is interpreted Best and Tyler (2007b).

Figure 2.3: Schematic of contrast assimilation in the PAM. It is a derivative of "sketch of the human vocal tract" by Wikimedia user Tavin, modified and used under licence CC-BY-3.0.
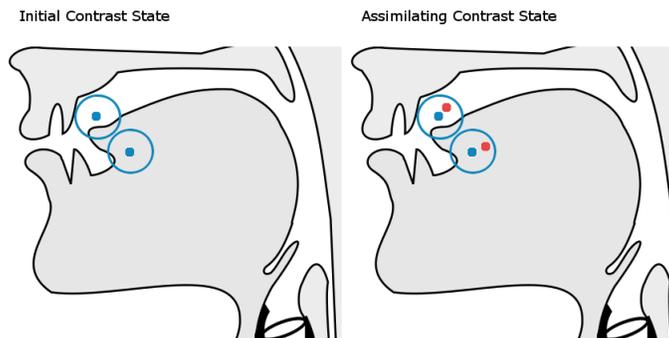
The PAM-L2 is not limited to phonetic-only distinction between speech sounds that the SLM suggests. Instead, the PAM-L2 postulates that the pho-

netic detail of a speech signal is mapped to phonological constraints in order to distinguish linguistic contrasts. Whilst these remain distributionally derived like the phonetic categories described by Flege (1995), the PAM-L2's considers contrasts as opposed to categories. The phonetic detail perceived in the PAM-L2's description takes the form of gestural constellations as opposed to the acoustic detail which is described by the SLM. Figure 2.3 shows a hypothetical situation in which an L1 contrast, shown in blue, is formed in the front of the vowel space. Following this, an L2 contrast shown in red is acquired and mapped to the L1 contrast given its perceptual closeness to the L1 contrast.

2.4.5 Discussion of the PAM-L2

Therefore, according to the PAM-L2, the process of forming the perceptual ability to differentiate between speech sounds arises from the mapping of articulatory information directly to abstract phonological rules. This postulate implies that a listener must also be able to produce a contrast, or at least have articulatory knowledge of its production, in order to perceive it. This conceptualisation of speech perception can be likened to the deliberate honing of a set of perceptual parameters which are applied to incoming acoustic signals, rather than referenced to long term memory (Tyler et al., 2014). The parameters themselves are built actively by the individual in order to extract articulatory features from the speech signal, which allows linguistic information to be interpreted.

The PAM describes four results that arise from the perception of non-native speech sounds, the occurrence of which is determined by the goodness of fit to an existing L1 contrast. These include a one-to-one mapping, two-to-one (for one or more L2 contrasts that fall within the boundaries of one L1 contrast), one-to-two (for an L2 contrast that may fall either side of an L1 boundary), or an inability to categorise if the L2 contrast is far enough removed from those of the L1. Criticism of the PAM-L2 is drawn from its embodied, direct-realist ontological approach.

2.4.6 Second Language Linguistic Perception Model

The Second Language Linguistic Perception Model borrows from computational models of distributional learning and combines some properties from both the SLM and PAM-L2 (Escudero, 2005; Escudero and Boersma, 2007). The L2LP draws from Optimality Theory, which posits that linguistic processing relies on the optimal satisfaction of a learned set of constraints (Smolensky and Prince, 1993).The L2LP states that optimal perception boundaries are derived from distributional category-forming much like the SLM. However, a salient difference between the L2LP and the SLM is that the L2LP postulates that phonemic categorisation occurs in long-term memory, rather than phonetic categorisation.

Figure 2.4: Flow chart depicting the trajectory of copying an initial L1 state into two perception grammars which are continually developed: the L1 state and the L2 state
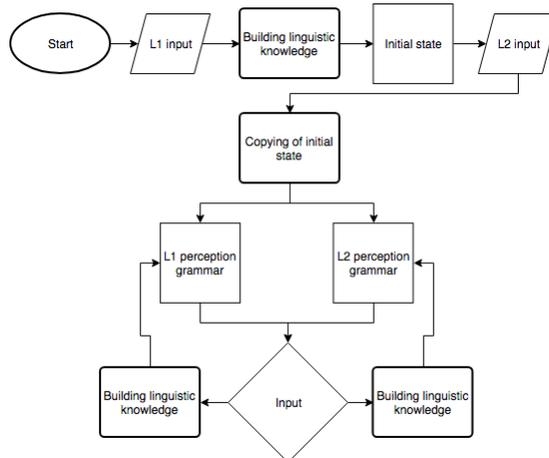
2.4.7 Perception Grammars in the L2LP

The L2LP's phonemic categorisation is carried out by language-specific perception grammars. Figure 2.4 shows a simplified schematic flowchart of the process by which a speaker is first attuned to a L1, and then later encounters the L2. A perception grammar, in the description of the L2LP, can be considered similarly to the sets of articulatory constraints determined by the PAM-L2. However, whilst the constraints of the PAM-L2 are gestural, the constraints of the L2LP are the realtive weight of acoustic cues for the cateogorisation of a speech sound. These perception grammars are encoded by exposure to speech signals, which allow acoustic boundaries to be placed between contrastive phonemes (Escudero, 2005).In other words, the perception grammar is a series of phonetic rules that are applied to an incoming speech signal in order to determine which phonemic category it belongs to.

2.4.8 Discussion of the L2LP

As a result, linguistic influence from L2 to L1 is less likely to occur due to the separation of the L1 and L2 phonologies. This is the distinguishing factor between the L2LP, which posits that individuals separate their "perception grammars", and the previously discussed models, which posit assimilation of the L2 system by the L1 (Colantoni, 2015). The L2LP predicts that L2 phonology begins as a direct map of L1 phonology – in other words, the current perception grammar of an individual's L1 becomes the initial perception grammar for learning L2. As learning continues, additional constraints are developed upon the L2 perception grammar such that it becomes distinct from that of the L1,

with new categories being formed as necessary. A prediction that can be derived from this tenet is that the presence of a majority language, or temporary language mode (Grosjean, 2012b, 1998), will influence perceptual preference for cross-linguistically similar speech sounds.

Crucially, the L2LP presents more space for individuals to achieve a monolingual-like perceptual boundaries with sufficient learning. The existence of separate perception grammars allows individuals to maximise the phonetic space for individual phonemes. This is due to the ability to use cues from the ambient language setting to engage perception grammars. Ju and Luce (2004) corroborate the concept of language-specific perception strategies during lexical access in order to reduce interference from the non-activated language(s). This postulation is expanded upon by Bosch and Ramon-Casas (2011), who note that in order to adapt the transferred boundaries through lexical and contextual cues, the speaker must first memorise the lexical items which are minimally contrastive for such a boundary.

2.5   Conclusion and Research Questions

Research on bilingualism has made considerable progress as research methods become more sophisticated and as the field of linguistics moves away from anglocentricity. However, due to the nature of much linguistic research especially in phonetics and phonology, experiments are often conducted on small scales and can struggle to make broader claims. As a result, many different models and interpretations of results can arise when there is a very large amount of potentially confounding or contextual factors. This creates an incentive for researchers to conduct experiments that can create parallels between contexts and factors such that more robust conclusions can be inferred.

The following research questions have been derived:

2.5.1   What degree of cross-linguistic influence is present in young heritage bilinguals?

The presence of cross-linguistic influence in bilingual speakers is well-documented. However, the interaction between linguistic systems is especially relevant in the case of heritage bilingualism. Heritage bilinguals are characterised by language use and exposure that is not only asymmetrical between the majority language and heritage language, but that is considerably variable over time. This relationship is exemplified the exposure in utero and infancy to the heritage language, during which implicit knowledge is implemented. The majority language only becomes the more used language once schooling begins and the child is immersed. In some cases, schooling is a child's first exposure to the majority language. Arising from this timeline is a complex psycholinguistic relationship between the languages spoken by a heritage speaker.

These factors can cause difficulty in discerning the relationship between the

two languages, and what the results of this relationship are in terms of cross-linguistic influence.

2.5.2 Do production and perception align during heritage language acquisition?

As covered in the literature review, experimental data in phonetics has often focused on either production or perception. In many cases, this focus is appropriate for the research task at hand. However, it should not be assumed that these two domains are congruent. Many heritage bilinguals may grow up sounding functionally indistinguishable from a monolingual in their respective languages. However, others may not. In order to explore both the underlying representation of cross-linguistically similar sounds this study will address both the production and the perception of this scenario.

2.5.3 Can the existing models of bilingual phonological acquisition predict these outcomes?

Towards informing current and future models of bilingual acquisition, discussing results within the framework of existing models of bilingualism is necessary for their development. The models addressed in the literature review require testing across the entire bilingual landscape.

For heritage speakers, the status of either language as an L1 or L2 is ambiguous. All three of the models discussed assume the L1 to be an initial state of L2 acquisition, from which assimilation, adaptation, or copying is performed. However, the complexity of heritage and majority language exposure and use, as opposed to L1 and L2, brings into question what the "initial state" is for a heritage learner during schooling. This is particularly relevant in childhood, during which linguistic input from parents is still high. In light of this, observation of the similar scenario, and consideration of cross-linguistic influence may present evidence towards discerning this relationship.

Chapter 3

Methodology

3.1 Introduction

The MPhil project has been planned to investigate contemporary issues in heritage bilingual language acquisition. This will be achieved by collecting data primarily from young heritage language speakers. This methodology section will first introduce the research questions that the thesis will address. Following this, the method of investigation will be described. The description will discuss both the pilot study and the main study in order to discuss both the methodological decisions made during the thesis project as well as the process they went through. A final section will describe the statistical approach the to study.

To summarise the questions derived from the existing literature, this thesis seeks to examine the following broad question: what happens when heritage speakers learn cross-linguistically similar sounds? Observing the same interaction in a different bilingual subset is likely to be beneficial for both informing linguistic theory and for the heritage community itself, or speech and language therapists and educators with whom they work.

3.2 Pilot Study

A pilot study was conducted in order to test a potential approach to the research questions. The pilot study was conducted at a co-operating Polish supplementary Saturday school run by the Polish community in a city in the North of England. Children attending the school were taught Polish by native Polish adults. The study was carried out one-on-one in a quiet room on the school premises by the author with permission acquired by written consent from the Head of the school and consent forms returned by parents. The participants for the pilot were 11 children aged 6;5 to 9;6. All 11 participants were heritage speakers of Polish residing in the North of England.

The pilot study comprised two parts which lasted approximately 20 minutes. This time frame was chosen in order to maintain participant interest and concentration. The first part of the experiment was a picture-naming task that targeted the relevant vowel contrasts in both languages. The second part of the experiment, which directly followed the production condition, was a forced-choice lexical decision task from acoustic input. The researcher recorded isolated tokens of a minimal triplet in the Lancaster Phonetics Lab, and these recordings were used for endpoints in the perception condition. Three six-step phonetic continua were produced from each pair using Praat software scripts from those provided on the website of Winn (2017). This method was refined and modified in order to design the main methodology.

3.3   Main Study

3.3.1   Procedure

The languages were separated in order to accommodate language modes in the experimental design, which would allow focused activation of the language-specific contrasts (Grosjean, 2012a). Further to this, each language block was delivered in the language of its subject for the same language modes effect. Therefore the experiment would follow the pattern of: English production, English perception, heritage language production, and then heritage language perception.

3.3.2   Production

The production task comprised 14 English stimuli followed by 14 heritage language stimuli, which included the relevant contrasting phonemes in lexical nouns as well as two filler stimuli for each language. A stimulus would be a single word including one of the relevant phonemes, elicited using a picture-naming task presented using a flipbook containing simple cartoon images. Children were asked "What is this?", or the heritage language equivalent, when shown each picture from a flip book. The image order was consistent every time, and children were asked to repeat the word at least once in order to elicit multiple samples and to ensure at least one clear sample for each stimulus word. If a child was unable to identify the image, or used an unexpected word, they were first given hints or descriptions of the noun. Failing this, participants were allowed to read the word from an otherwise hidden list. Recordings were acquired using a lapel condenser microphone connected to a USBPre2 digital audio interface, the input from which was recording using Audacity open source software (Audacity Team, 2019).

Once production data had been recorded, each token was labelled and analysed using Praat (Boersma, 2001), with formants values extracted from the temporal midpoint of each vowel using Praat's automated Burg algorithm. The spectral slice at the midpoint was chosen because they are monophthongs oc-

curring in a variety of phonetic contexts. This approach was taken because of the steady state of monophthongs, and to minimise potential skewing by phonetic contexts which could occur from spectral slices near segmental boundaries. Once the data was processed, it was analysed without normalisation. This was because of the exploratory nature of the pilot study. If the data were to be used in a future investigation, normalisation would be used.

3.3.3 Perception

For the perception task, methodological inspiration was taken from the method used byEscudero (2005) for investigating cue-weighting. The perception tasks comprised two 2-dimensional phonetic continua, one for each within-language contrast, which vary the vowel's spectral characteristics of $F1 - F2$ on one dimension, and the vowel's duration on the other dimension. For this perception condition, within-language minimal pairs were used.

The researcher recorded isolated tokens of the stimulus words in an acoustically controlled environment, and these recordings were used for the basis of resynthesis in the perception condition. A 36-point 2-dimensional continuum was produced from each pair using Praat software scripts from Boersma (2001); Winn (2017). The dimensions of the continua were spectral and temporal, comprising 6 steps for each dimension. The same formant continuum script from the pilot study was applied to the initial token pairs, creating a continuum for each minimal pair with the formant values represented on the y axis of Figure 8. Following this, a duration continuum script, also from Winn (2017), was used on each continuum step. This populated the x axis on Figure 8, creating 2-dimensional phonetic continua, with each unique F2-F1 and duration combination represented. The script extracted a user-selected vowel segment, and resynthesized the vowel for each duration step using Praat's manipulation functions before splicing it between user-specified initial and final segments. Each language condition included 36 tokens, with six increments on each dimension. The stimuli were compared with recordings from native speakers, both auditorily and quantitatively using formant Hz values, to verify their suitability for the task.

The perception stimuli were delivered using Praat's inbuilt experiment scripting feature. The stimuli were presented in isolated tokens using a pair of Sennheiser HD 201 headphones connected to the researcher's laptop in a randomised order. In order to elicit responses, children were asked to point to the correct word on paper, and were allowed to hear repeats of a stimulus if they were not sure. The child was asked, in the relevant language (English, Polish, or Chinese, depending on the task), "Which did you hear?". The researcher would not say either choice to avoid priming a particular response, and reassured the participant that there was no wrong answer if there was hesitation. Once an answer was given, the researcher would input the response on the laptop with the screen facing away. Then, the next stimulus would play automatically.

3.3.4   Participants

Participants were sought between the ages of 7 and 9 years inclusively. This approximate age range was specified due to the status of phonological and lexical development typical for children, which would accommodate the demands of the task whilst also being somewhat "stable" (James et al., 2001). All of the participants were regular pupils of the Polish supplementary school. A language use and exposure questionnaire was given to parents alongside consent forms in order to acquire more detailed information regarding linguistic input at home for each participant. The questionnaire revealed that all of the participants' parents were Polish L1, English L2 speakers and provided information on relevant details such as siblings and the balance of English and Polish used in the household. Polish was the primary home language for all participants, with English being the language used in the day-to-day education and social activities.

3.4   Quantitative Analysis

For each token, measurements of $F2$, $F1$, $F2 - F1$ (post-calculation), and *duration* were extracted. These were scaled using the `scale` function in R R Core Team (2018). This normalization allowed the data to be compared across speakers, whose anatomy may differ, and improved modeling by giving the variables a consistent scale. Formant frequencies were extracted using Praat at the midpoint of each vowel. The formants were detected in Praat by setting the `Maximum formant (Hz)` to 5000, and the `Number of formants` to 3. The same measurements were recorded for the perception stimuli, so that the results can be compared between conditions.

To compare spectral and temporal cues, formant frequencies were reduced to one dimension by using F2-F1 to represent the spectral cue. This measure was chosen because both contrasts feature these acoustic dimensions. The trajectory between each contrast involves moving the tongue back and lower. Both of these movements lower F2-F1, making it a reasonable method of reducing the dimensions of the spectral cue (Kuhl et al., 1997; Watt and Fabricius, 2002).

In order to assess the presence of the each cue for both the production and perception conditions, a likelihood ratio test was conducted on nested models for each cue. For each language a model was fitted with vowel as the dependent variable, duration and F2-F1 measurements as fixed effects, with the word as a random effect. A nested model was created with the relevant cue absent, and the models were compared using the `anova` function in R, in order to perform a χ^2 comparison of the models. For each combination of population, condition, and language, both a full model and a null model were calculated for model comparisons.

The hypothesis suggests that the inclusion of each cue should have a significant effect on the performance of the model, because its value would be a predictor of the vowel, increasing the model's accuracy when compared to using the spectral or temporal data alone. A significant result would suggest that the

variable in question is the cause of the differences between models.

To compare production and perception, a new model was fitted using normalised F2-F1 measurements as the dependent variable. This model included an interaction between *duration* and *condition* (i.e. production or perception) in order to investigate whether the condition changed the relationship between the temporal cue and dependent variable. The hypothesis suggests that the condition will not effect the contrasts. For this to be supported, there will be no significance for the interaction term.

3.5 Conclusion

The method of the thesis study was constructed around creating a strong and unique data set that is able to provide progress towards modelling heritage bilingual acquisition. Doing this requires data that can enable comparisons across factors such as different languages, and the differences or similarities between perception and production. With these priorities in mind, two parallel experiments were conducted for different demographics of heritage bilinguals.

Chapter 4

Production and Perception of Cross-linguistically Similar Contrasts by Polish-English Heritage Bilingual Children

4.1 Introduction

This study considers heritage bilingualism in the form of young children who are still in active use of the heritage language at home, but also attend a primary school taught in the majority language. The study investigates the similar contrasts of English /i/-/ɪ/ and Polish /i/-/ɨ/ acquired by English-Polish heritage bilingual children. The study seeks to evaluate the state of cross-linguistic influence within these speakers in both production and perception scenarios. Few studies have observed production and perception within the same sample in the context of heritage bilingualism, and this approach may contribute to the understanding of the relationship between them. The rest of this section briefly outlines the specific context of English-Polish heritage bilinguals in this study.

4.2 Chinese Population in the UK

The Polish-born population is the largest overseas-born population in the UK, with approximately 922,000 members as of 2017 (The Office for National Statistics, 2018). This allows the study to reach out to a broad number of institutions and individuals in order to find a substantial sample. This rapid growth has enabled a widespread and growing population of Polish-English bilingual and

heritage speakers. The growth in the Polish-born population in the UK requires that adequate support for bilingual children is present in both schools and healthcare, and investigating bilingual acquisition among these children moves towards this goal.

4.3   Polish and English Phonological Interactions

Koźbiał (2015) explains the consequences of the interactions between Polish and English phonology. Examples of this effect include identifying and producing contrasts between the pairs /i/-/ɪ/ and /ɔː/-/ɒ/. In realizing these contrasts, it was noted by Szpyra-Kozłowska (2003) that Polish speakers "employ fewer contrasts", resulting in some English contrasts not being present. Tamburelli et al. (2015) noted that Polish "allows for a greater range" of consonant clusters than English, particularly for affricates, exemplified in words such as *chrząszcz* (beetle). As a result of this, Polish-English bilingual children exhibited acceleration in the acquisition of English consonant clusters.

   The present experiment was designed to compare the cross-linguistic production and perception of vowels by young Polish heritage speakers in England. The objective of the experiment is to explore the spectral and temporal boundaries and differences between the vowels /ɪ/, /ɨ/ and /i/ for these speakers and what these imply in the context of competing models of bilingual phonology.

   Whilst there exist multiple similar scenarios between English and Polish across both vowels and consonants, the chosen subject of this study is that of English /i/-/ɪ/ and Polish /i/-/ɨ/. This similar pairing of contrasts occurs in the high-front vowel space and has been shown in the past to be a matter of difficulty for Polish-L1, English-L2 speakers (Szpyra-Kozłowska, 2003). English /i/-/ɨ/ is a contrast that varies across both spectral cues and temporal cues (Bohn and Flege, 1991). For the similar Polish contrast, /i/-/ɨ/, the contrast is spectral only. This relationship raises interesting questions of the interactions of cue-weighting in perception, as well as articulatory strategies. The asymmetrical inventory of contrastive cues for these categories is likely to exhibit cross-linguistic influence, as temporal features may be assimilated or ignored depending on exposure, proficiency, and communicative necessity (Amengual and Chamorro, 2015). This may be especially true among children, who remain in high exposure to both languages.

4.4   Hypotheses

RQ 1. Do heritage bilingual children maintain contrasts in both languages in production and perception? Hypothesis: Yes. Due to regular exposure and practice from a young age, heritage bilinguals will maintain within-language contrasts.

   RQ 2. Do heritage bilingual children produce and perceive Chinese differently to monolingual Chinese children? Hypothesis: Cross-linguistic influence

Table 4.1: Production Stimuli, Produced in Isolation

English				Polish			
/ɪ/		/i/		/ɨ/		/i/	
ship	/ʃɪp/	sheep	/ʃiːp/	ryba	/rɨba/	drzwi	/dʐvi/
pig	/pʰɪg/	feet	/fiːt/	trzy	/tzɨ/	świnia	/ɕviɲja/
chips	/tʃɪps/	beach	/biːtʃ/	tygrys	/tɨgris/	szpital	/ʂpital/
six	/sɪks/	key	/kʰiː/	mysz	/mɨʂ/	widelec	/vidɛlet͡s/
fifty	/fɪfti/	cheese	/tʃiːz/	mydlo	/mɨdwɔ/	mis	/mis/
stick	/sdɪk/	peas	/pʰiːz/	frytki	/frɨtki/	miska	/miska/
kitten	/kʰɪtən/	teeth	/tʰiːθ/	buty	/butɨ/	lis	/lis/
fish	/fɪʃ/	leaf	/liːf/	cytryna	/tsɨtrɨna/	lisc	/liʂt͡ɕ/

will lead to differences between the heritage bilinguals and monolinguals.

RQ 3. Do heritage bilingual children produce and perceive English differently to monolingual English children? Hypothesis: Cross-linguistic influence will lead to differences between the heritage bilinguals and monolinguals.

4.5 Methodology

The experiment was designed to investigate the acoustic cue-weighting of cross-linguistic phonological contrasts employed by the children during language use. The experiment conditions were separated by language and by production and perception. The experiment was separated into an English block including production and perception, which was then followed by a Polish block. The languages were separated in order to accommodate language modes in the experimental design, which would allow focused activation of the language-specific contrasts (Grosjean, 2012a). Further to this, each language block was delivered in the language of its subject for the same language modes effect.

4.5.1 Production

To elicit production data, a picture-naming task was presented using a flipbook containing simple cartoon images. Children were asked "What is in the picture?" and shown the images in sequence. In the event that a child could not identify the image, they were given hints that did not reveal the pronunciation, in order to avoid imitation. Images were chosen over orthography to avoid any "sounding out" of the stimuli. The images all depicted lexical nouns that the children were likely to be familiar with in each language. For each vowel, there were 8 words, and each flip-book contained a filler word. The stimuli were spoken twice each and recorded using Audacity (Audacity Team, 2019) and a lapel microphone routed through a Sound Devices USBPre2 soundcard, connected to the researcher's laptop at a sampling rate of 44100 Hz.

4.5.2 Perception

The perception tasks comprised two 2-dimensional phonetic continua, one for each within-language contrast, which vary on one dimension the vowel's spectral characteristics of $F1 - F2$, and on the other dimension the vowel's duration. For this perception condition, within-language minimal pairs were used. For English, the endpoints were <ship> and <sheep>. For Polish, the endpoints were <my> and <mi>, which are phonemically /mɪ/ and /mi/. The acoustic values, and a visual representation can be found in Figure 4.1.

Figure 4.1: Perception Stimuli Acoustic Values

The researcher recorded isolated tokens of the stimulus words in an acoustically controlled environment, and these recordings were used for the basis of resynthesis in the perception condition. A 36-point 2-dimensional continuum was produced from each pair using Praat software scripts from Boersma (2001); Winn (2017). The dimensions of the continua were spectral and temporal, comprising 6 steps for each dimension.

The stimuli were delivered using Praat's inbuilt experiment scripting feature. The stimuli were presented in isolated tokens using a pair of Sennheiser HD 201 headphones connected to the researcher's laptop in a randomised order. In order to elicit responses, children were asked to point to the correct word on paper, and were allowed to hear repeats of a stimulus if they were not sure.

4.5.3 Participants

The participants for the study were 58 children aged 7;0 to 9;6 ($\bar{x} = 8;3$). Six other children participated, but their results were excluded due to their ages being too far from the rest of the sample. Within the finals dataset, there were 24

heritage bilinguals of Polish and English ($\bar{x} = 8;5$, $\sigma = 0;9$), 18 English mono-
linguals ($\bar{x} = 8;3$, $\sigma = 0;7$), and 16 Polish monolinguals ($\bar{x} = 8;0$, $\sigma = 0;6$).
No children were reported to have any language, motor, or hearing impairment.
The details of the sample are outlined below.

The Polish heritage participants were UK residents and regular pupils of
Polish supplementary schools run by local Polish community at weekends. This
provided some assurance that the children had regular Polish input and practice
with both peers and adults. All but one of the participants' parents were Polish
L1, English L2 speakers. The exception to this was one parent who spoke
English L1, Polish L2. Two participating schools were located in the North of
England, and two in the South-East. Polish was the primary home language
for all participants, with English being the language used in the day-to-day
education and social activities. The research took place in a quiet room on the
school premises. The English monolinguals were recorded at a primary school
in the South of England, and the Polish monolingual data was collected at a
primary school in western Poland. All data was collected with informed parental
consent using a consent form and information sheet. The consent procedure and
methodology was approved by the Lancaster University FASS-LUMS Research
Ethics Committee.

4.6   Analysis

The total dataset comprised 2686 production tokens between all speakers, af-
ter 66 were discarded due to being unsuitable, for reasons such as interfering
background noises. The perception data comprised 3168 responses.

For each token, measurements of $F2$, $F1$, $F2 - F1$ (post-calculation), and
duration were extracted. These were scaled using the `scale` function in R
R Core Team (2018). This normalization allowed the data to be compared
across speakers, whose anatomy may differ, and improved modeling by giving
the variables a consistent scale. Formant frequencies were extracted using Praat
at the midpoint of each vowel. The formants were detected in Praat by setting
the `Maximum formant (Hz)` to 5000, and the `Number of formants` to 3. The
same measurements were recorded for the perception stimuli, so that the results
can be compared between conditions.

To compare spectral and temporal cues, formant frequencies were reduced to
one dimension by using F2-F1 to represent the spectral cue. This measure was
chosen because both contrasts feature these acoustic dimensions. The trajectory
between each contrast involves moving the tongue back and lower. Both of
these movements lower F2-F1, making it a reasonable method of reducing the
dimensions of the spectral cue (Kuhl et al., 1997; Watt and Fabricius, 2002).

In order to assess the presence of the each cue for both the production and
perception conditions, a likelihood ratio test was conducted on nested models
for each cue. For each language a model was fitted with vowel as the dependent
variable, duration and F2-F1 measurements as fixed effects, with the word and
speaker as random effects. A nested model was created with the relevant cue

absent, and the models were compared using the `anova` function in R, in order to perform a χ^2 comparison of the models. For each combination of population, condition, and language, both a full model and a null model were calculated for model comparisons.

The hypothesis suggests that the inclusion of each cue should have a significant effect on the performance of the model, because its value would be a predictor of the vowel, increasing the model's accuracy when compared to using the spectral or temporal data alone. A significant result would suggest that the variable in question is the cause of the differences between models.

To compare production and perception, a new model was fitted. This model included an interaction between *duration* and *condition* (i.e. production or perception) in order to investigate whether the condition changed the relationship between the temporal cue and dependent variable. The hypothesis suggests that the condition will not effect the contrasts. For this to be supported, there will be no significance for the interaction term.

4.7 Do heritage bilingual children maintain contrasts in both languages in production and perception?

This section tests whether the contrast was maintained reliably by the participants, and on which dimensions. Before the individual cues are addressed, the analysis will establish the existence of a reliable contrast irrespective of individual cues.

A nested model comparison was conducted using ANOVA, and comprised subsetted models for each of the acoustic cues. A nested ANOVA model comparison was carried out against the null model in order to assess the reliability of the contrast under each condition. Firstly, a null model was specified using the `lme4` package in R. The null model was specified as a generalized linear model with `vowel` as the dependent variable, no fixed effects, and `subject` as a random effect. Secondly, a generalized linear model comprising the same dependent variable, `F2-F1` and `duration` as fixed effects, and `subject` as a random effect. A significant result from the likelihood-ratio test suggests that the contrast was maintained reliably for that condition, given that the comparison model classifies the dependent variable more effectively. Both acoustic cues are tested within this first model comparison in order to establish whether the contrast in maintained on any dimension.

Table 4.2: ANOVA results of a null generalized linear regression model compared to model with F2-F1 and duration cues as fixed effects

Population	Language	Condition	χ^2	p
Heritage Bilingual	English	Production	937.433	<0.001
		Perception	898.717	<0.001
	Polish	Production	705.033	<0.001
		Perception	773.944	<0.001
Monolingual	English	Production	584.340	<0.001
		Perception	598.189	<0.001
	Polish	Production	626.953	<0.001
		Perception	539.181	<0.001

Table 4.2 shows the results of the nested model comparison between a null model and a model including both acoustic cues as fixed effects, with the vowel as the response variable. Whilst the research question above refers specifically to bilinguals, the monolingual effects are included for reference. A significant result suggests that the model with fixed effects is more effective at predicting the vowel than the null model, and that the contrasts are maintained. The analysis produced significant results for all conditions ($p < 0.001$), supporting the hypothesis that contrasts would be maintained. The heritage speakers were able to maintain the contrasts in both languages.

Figure 4.2: Violin plots of acoustic cues in production
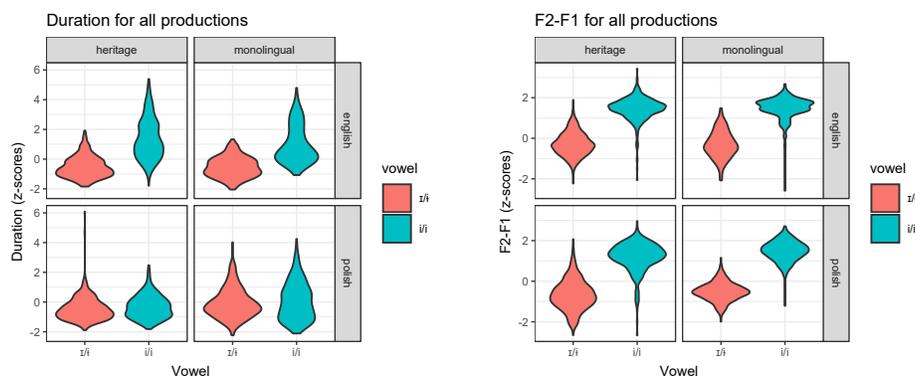
Figure 4.2 depicts the distribution and ranges of duration and formants for the production condition. There are evident differences between both spectral and temporal cues in English. The plots show that the effect is seen in a longer /i/ vowel in English. In Polish, there is mainly a difference in formants according to the plots.

Table 4.3 shows the results of the interaction model comparison. It suggests that there is a significant difference across conditions for both languages for heritage bilinguals (English: $\chi^2 = 554.15$, $p < 0.001$)(Polish: $\chi^2 = 9.976$, $p = 0.019$). For the monolinguals, only the English contrast showed significant effects (English: $\chi^2 = 305.37$, $p < 0.001$)(Polish: $\chi^2 = 5.1881$, $p = 0.159$).

Table 4.3: ANOVA of null generalized linear regression model compared to a model with a duration*condition interaction as fixed effects

Population	Language	χ^2	p
Heritage Bilingual	English	554.150	<0.001
	Polish	9.976	0.019
Monolingual	English	305.37	<0.001
	Polish	5.1881	0.159

To test the degree to which production and perception matched for maintaining contrasts, a null model was fitted to the data, and compared to a model including an interaction between duration and condition, in order to investigate whether the temporal cue differed across conditions. The same null model was used for comparison to a generalized linear model comprising `vowel` as the dependent variable, `duration` and `condition` as interacting fixed effects, and `subject` as a random effect. A significant result here suggests that the effect of duration varied between the production and perception conditions.

Figure 4.3: Comparison of English contrast production and perception in heritage bilinguals, with 50% decision boundaries and 95% confidence ellipses in production
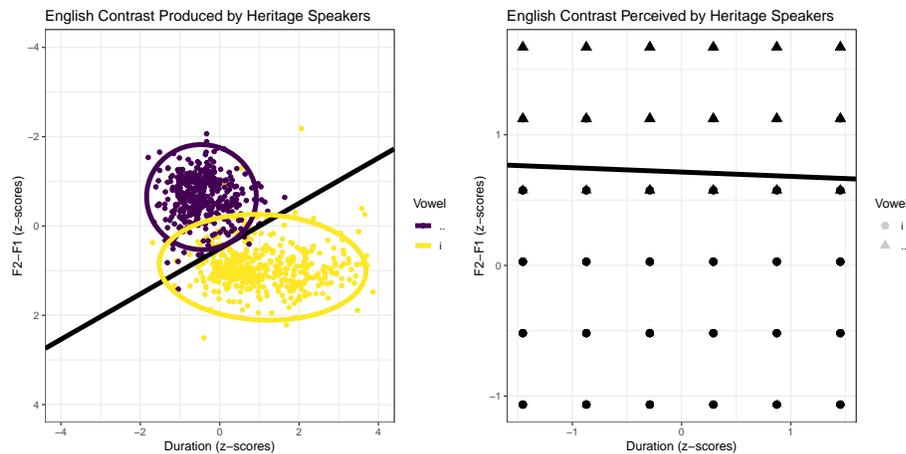
Figures 4.3 and 4.4 show graphical depictions of heritage speaker productions and perceptions of each within-language contrast, juxtaposed for visual comparison. The plots include a point for each token as well as an ellipse with 95% confidence intervals. The straight lines depict decision boundaries, representing the predicted boundary separating the vowels acoustically, calculated using generalized linear models, and predicting an outcome of 0.5 for a the binary decision task, thereby estimating the contrast boundary.

Figure 4.4: Comparison of Polsih contrast production and perception in heritage bilinguals, with 50% decision boundaries and 95% confidence ellipses in production
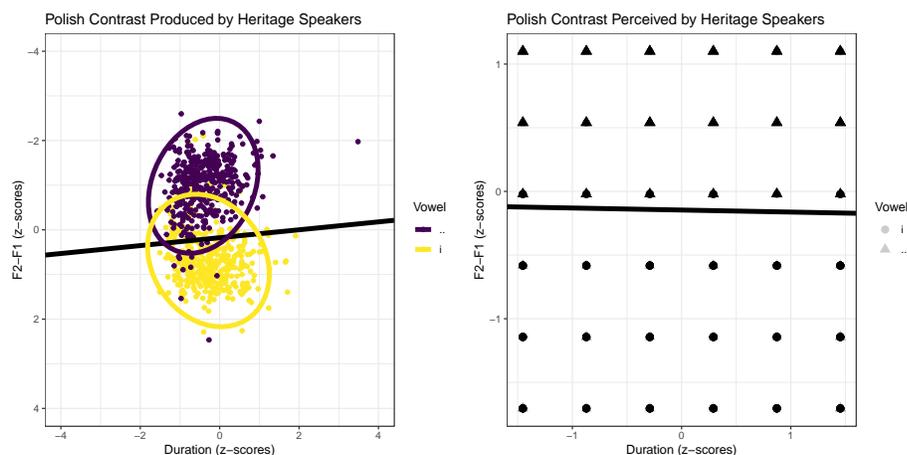
There are noticeable gradients for the English productions, which suggest that the temporal cue is present for this contrast. This suggests that the difference between production and perception is found in the presence of a temporal cue in production only. However, this is not the reflected in perception, supporting the results of the model comparison.

4.8 Do heritage bilingual children produce and perceive Polish differently to monolingual Polish children?

Table 4.4 shows the results of a nested model comparison for each cue and condition, with the vowel as the dependent variable. The comparison is a likelihood ratio test involving the full model, including both duration and F1-F2 as fixed effects, and a nested model in which one cue was removed. If a significant difference is present, then the removed cue has a significant effect on the contrast. For the spectral cue, this is significant in all cases ($p < 0.001$), suggesting that the spectral cue is contrastive across all conditions.

Table 4.4: ANOVA results of a generalized linear mixed model with both acoustic cues as fixed effects, compared to nested models, each with one cue removed

Population	Language	Condition	Cue	χ^2	p
Heritage Bilingual	Polish	Production	Spectral	697.685	<0.001
			Temporal	1.789	0.181
		Perception	Spectral	773.860	<0.001
			Temporal	0.253	0.614
Monolingual	Polish	Production	Spectral	624.066	<0.001
			Temporal	0.993	0.319
		Perception	Spectral	539.054	<0.001
			Temporal	0.434	0.510

The model comparison in Table 4.4 does not show significant weight of the duration cue in the Polish conditions (heritage bilingual: $\chi^2 = 1.7889$, $p = 0.181$)(monolingual: $\chi^2 = 0.9931$, $p = 0.319$). The result suggests that, in both production and perception, the temporal cue was not present in either population. This suggests that the spectral cue alone was sufficient to maintain the contrasts in this task, and that the duration feature of the English contrast has not been assimilated in the heritage bilinguals.

A null model was fitted to the data, and compared to a model including an interaction between duration and condition, in order to investigate whether the temporal cue differed across populations.

Table 4.5: ANOVA of null generalized linear mixed regression model compared to one with duration*population as a fixed effect

Condition	Language	χ^2	p
Production	Polish	10.602	0.014
Perception	Polish	0.411	0.938

Table 4.5 shows the results of a model comparison between the null model and one with an interaction between duration and population. The table shows a significant effect of population in the production condition ($\chi^2 = 10.602$, $p = 0.014$), but not in perception ($\chi^2 = 0.4112$, $p = 0.938$), suggesting that the the heritage bilinguals do not match the monolinguals.

Figure 4.5 shows a side-by-side comparison of heritage bilingual and monolingual Polish productions. In combination with Figure 4.2, it depicts a narrower temporal range in productions, which is the opposite effect to that seen in the English plots, and demonstrates the direction of the effect shown in Table 4.5. This is to be compared to Figure 4.7, which depicts a broader temporal range in productions, and may explain the significant result of the models in Table 4.5.

Figure 4.6 shows a comparison of heritage bilingual and monolingual perception results for Polish. The quantitative analysis suggested no significant difference between these results. This is corroborated by the similarity of the plots and the lack of significance in the models in Table 4.5.

Figure 4.5: Comparison of heritage bilingual and monolingual Polish production, with 50% decision boundaries and 95% confidence ellipses for each vowel
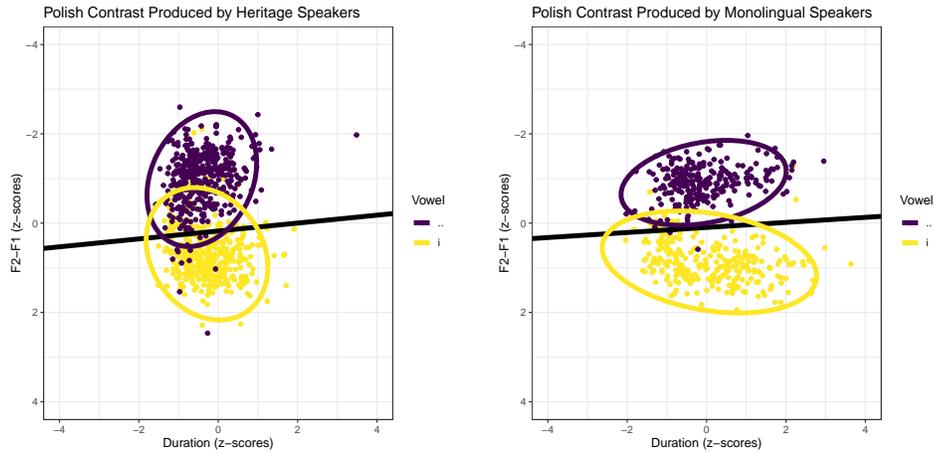
Figure 4.6: Comparison of heritage bilingual and monolingual Polish perception with 50% decision boundaries
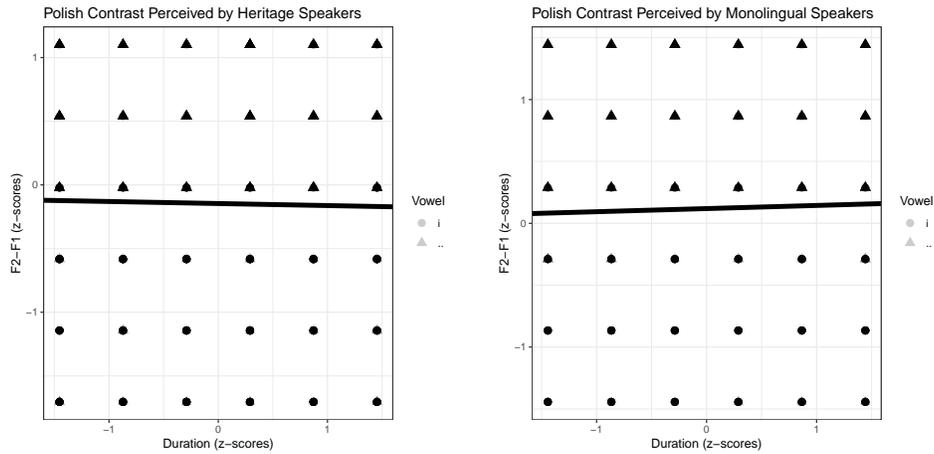
4.8.1 Do heritage bilingual children produce and perceive English differently to monolingual English children?

Table 4.6 shows the results of a nested model comparison for each cue and condition. The comparison compared the full model, with F2-F1 and duration as fixed effects, to a nested model in which one cue was removed, in order to test for a significant difference. If a significant difference is present, then the removed cue has a significant effect on the contrast. For the spectral cue, this is significant in all cases ($p < 0.001$), suggesting that the spectral cue is contrastive

across all conditions.

Table 4.6: ANOVA results of a full model with both cues compared to nested models, each with one cue removed

Population	Language	Condition	Cue	χ^2	p
Heritage Bilingual	English	Production	Spectral	416.124	<0.001
			Temporal	96.630	<0.001
		Perception	Spectral	898.317	<0.001
			Temporal	1.812	0.178
Monolingual	English	Production	Spectral	285.511	<0.001
			Temporal	59.249	<0.001
		Perception	Spectral	598.114	<0.001
			Temporal	0.271	0.602

The model comparison in Table 4.6 shows a significant effect of the addition of duration in both English conditions (heritage bilingual: $\chi^2 = 96.63$, $p < 0.001$)(monolingual: $\chi^2 = 59.249$, $p < 0.001$). This suggests that English production of the contrast does use the temporal cue for both monolinguals and heritage bilinguals.

Table 4.7: ANOVA of null model compared to one with duration*population as a fixed effect

Condition	Language	χ^2	p
Production	English	820.770	<0.001
Perception	English	1.416	0.702

Table 4.7 shows the results of a model comparison between the null model without duration and one with an interaction between duration and population. The table shows a significant effect of population in the production conditions ($\chi^2 = 820.770$, $p =< 0.001$), but not in perception ($\chi^2 = 1.416$, $p = 0.702$).

Figure 4.7 shows a side-by-side comparison of heritage bilingual and monolingual English productions. The temporal cue is plotted on the x axis, and the spectral cue on the y axis. As above, ellipses with a 95% confidence interval are plotted for each vowel, and a decision boundary that predicts relative cue weighting has been calculated.

Figure 4.8 shows a comparison of heritage bilingual and monolingual perception results for English. The quantitative analysis suggested no significant difference between these results. This is corroborated by the similarity of the plots and the lack of significance in the models in Table 4.7.

Figure 4.7: Comparison of heritage bilingual and monolingual English productions, with 50% decision boundaries and 95% confidence ellipses
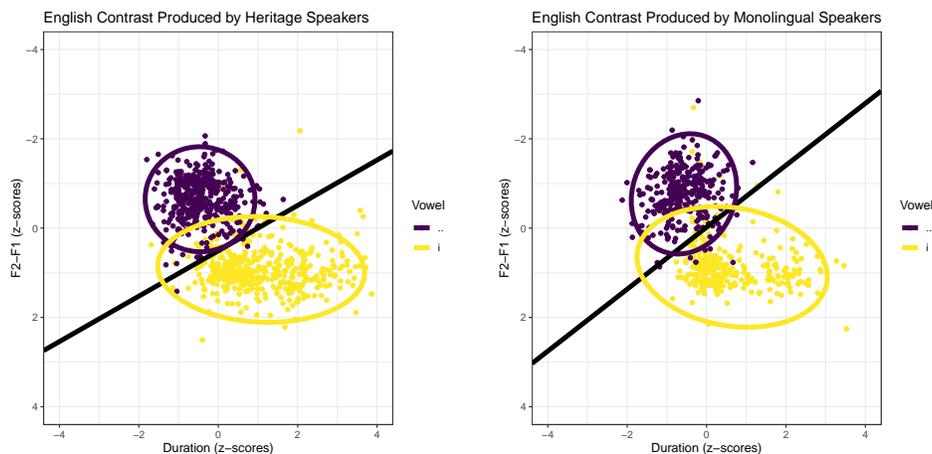
Figure 4.8: Perception results across populations and languages, with 50% decision boundaries
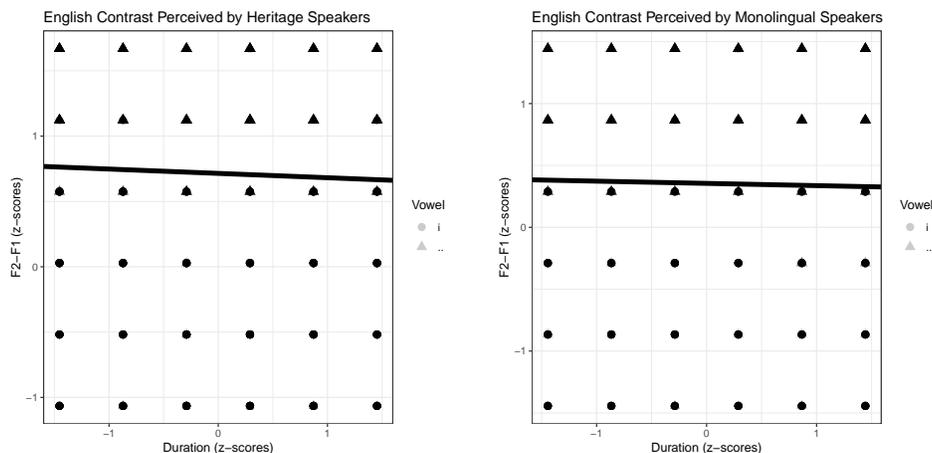
4.9 Discussion

The objective of this study was to appraise the learning of a cross-linguistically similar pair of vowel contrasts in childhood heritage bilingualism. The analysis focused on firstly establishing the spectral and temporal factors in the contrasts, and then assessing cue weighting and cross-linguistic influence. This discussion will address the research questions before offering a brief comparison of the findings to knowledge of sequential and simultaneous bilinguals in order to inform future directions of research.

The results suggest that heritage bilinguals reliably maintain the cross-linguistically similar contrasts in both production and perception. These contrasts are similar to monolinguals in the way that they use the same cues, with duration being present only in production of the English contrast.

Cross-linguistic influence is present in the relative cue weight for the contrasts. The heritage bilinguals relied less on the temporal cue in English production than English monolinguals. As a result, more weight was given to the spectral cue. For the Polish contrast, there was a significant result for higher temporal weight employed by heritage bilinguals. Figure **??** suggests that this is a minor effect.

For perception, the contrasts employed by the heritage speakers were monolingual-like, with no significant difference in cue-weight as seen in Table 5.7.

4.9.1 Do heritage bilingual children maintain contrasts in both languages in production and perception?

The analysis found that the heritage bilinguals were able to maintain both contrasts reliably, in both production and perception. This result affirms the hypothesis that the heritage bilinguals will have reliable contrasts in both languages. In production, the English contrast involved both spectral and temporal cues and the Polish contrast involved only the spectral cue. In perception, both contrasts involved only the spectral cue. The maintenance of both contrasts is expected, even if the contrast is represented differently (Kajouj and Kager, 2019). Having regular practice and exposure to both languages from adults and peers, it's likely that there is sufficient input for the children to recognize and produce both contrasts reliably. These results build upon the previous research on heritage bilingual production by exemplifying that a high degree of proficiency can be present at a relatively early age (Chang et al., 2011). This evidence is contrary to historical emphasis on a critical period that ends in the first few years of life, and supports the more flexible effect of age as discussed by Mayberry and Kleunder (2018), as it would be typical for a heritage language bilingual to begin learning the majority, societal language upon attending school. Future research may be encouraged to collect information about media exposure to each language during the child's early years in order to investigate the importance of such exposure.

4.9.2 Do heritage bilingual children produce and perceive Polish differently to monolingual Polish children?

The heritage bilinguals showed a broader spectral distribution for the Polish contrast when compared to monolinguals, and a significant but apparently minor temporal effect. This suggests the presence of mutual influence between the languages within the heritage bilingual. The broader spectral distribution may be explained by an assimilatory effect. The English contrast also has a broader spectral distribution than the Polish contrast, possibly due to the existence of a secondary cue. This presence can be seen in the Polish contrast as produced

by the heritage bilinguals. These effects combined suggest that there is transfer from English to Polish.

The perception results showed no significant difference between heritage bilinguals and Polish monolinguals. This suggests a monolingual-like perceptual contrast. The hypothesis that cross-linguistic influence will cause the heritage bilinguals to differ to monolinguals is partly affirmed, with the presence of a temporal effect in production which is not reflected in perception.

4.9.3 Do heritage bilingual children produce and perceive English differently to monolingual English children?

The analysis found a significant difference in temporal cue weight between the heritage bilinguals and the English monolinguals, in which the heritage speakers showed less use of the temporal cue, suggesting an assimilatory effect. For perception, the relative cue weighting for spectral and temporal cues is a matter of interest, given that the temporal cue was only present in production. However, this discrepancy was repeated in the monolingual sample. Bohn and Flege (1991) found that the temporal cue was not used in English perception by native speakers, and this result is consistent with that conclusion. The heritage speakers exhibited marginally more reliance on the spectral cue when perceiving the English contrast, but the minor difference was not numerically significant. This behavior aligns the heritage bilinguals with English monolinguals in their perception of the tense-lax contrast. A possible outcome that did not occur was an over-reliance on the duration cue, which, as previously mentioned, has been observed in Polish L1, English L2 individuals by Koźbiał (2015). Koźbiał (2015) also noted that many of the observed sequential bilinguals did not deviate from this pattern regardless of their exposure or their time practicing. These results combined suggest that heritage bilinguals do not simply transfer and maintain phonological relationships, which had been seen in sequential bilinguals. Instead, new relationships are readily available and influence each other. These results affirm the hypothesis that cross-linguistic influence will be observable for the heritage bilinguals. The cross-linguistic influence is bi-directional in this case, considering the this and the previous section. However, it is not symmetrical. The Polish contrast differs from monolinguals spectrally while the English contrast differs from monolinguals temporally.

4.10 Differences in Production and Perception

There were differences between both production and perception in the heritage bilinguals and the English monolinguals. A few explanations exist for this. The different goals of production and perception may be a factor in this result. When producing speech, the children would be encouraged to use a pronunciation that maximizes clarity, and this would be achieved by incorporating all cues possible. A perception system has more opportunity to be somewhat utilitarian, making use of deduction such as knowing which language is being perceived. In previous

work, Shultz et al. (2012) found little to no correlation between production and perception for an English contrast, and noted that secondary cues may have "practical irrelevance" under certain conditions. In other words, the primary cue and context are often sufficient. Boersma and Chládková (2011) showed that some multidimensional boundaries, in this case using both F1 and F2, were reduced to one dimension in perception. The tendency of the mind to prefer simpler perceptual configurations offers an intuitive explanation to these effects (Chater and Vitányi, 2003). In order to reduce cognitive load, the simpler perceptual task will be chosen where possible. Thus, the use of the temporal cue may only be apparent in cases for which the spectral cue does not suffice, such as in noise or when perceiving accented speech.

4.11   Assessment of Cross-linguistic Influence

Whilst the presence of cues matched between heritage bilinguals and monolinguals, the analysis showed that there were significant differences in the duration of productions between heritage bilinguals and monolinguals. This occurred in production for both languages. When producing the Polish contrast, the heritage bilinguals showed more overlap in the spectral qualities of each vowel when compared to the Polish monolinguals, and a broader range, which is more consistent with the English production. In the violin plots, this effect is made apparent, and it suggests that the spectral qualities of the vowels show converging cross-linguistic influence. As the heritage bilinguals age, their relative practice with the heritage language reduces as their use of the majority language increases.The difference suggests that cross-linguistic influence is present, and is causing differences in the weight of the temporal cue. However, these differences are not as substantial as those observed in sequential bilinguals, who have been observed to over-rely on the temporal cue (Koźbiał, 2015).

In combination with the above, a dissimilation effect is present temporally, and the categories occupy a smaller acoustic space than those of monolinguals. Such an effect is likely the result of crowding in a shared phonological space, causing the temporal factor to play a higher role in differentiating the vowels. This can be observed particularly in the temporal distribution of the Polish contrast, for which the heritage bilinguals showed a smaller, more concentrated distribution closer to that of the English /ɪ/. The temporal representation of the Polish contrast (which appears to be considered phonologically "short" by the heritage bilinguals) has accommodated the presence of a long English sound, /i/, by narrowing its distribution. The combination of assimilation and dissimilation represents a complex mutual exchange between the heritage and majority languages.

The discussion of the results indicates that the status of heritage bilinguals is closer to that of simultaneous bilinguals than sequential bilinguals. This is seen in the mutual transfer between languages, as opposed to the transfer of one phonology to the other. Models such as the SLM and PAM-L2 may not be well suited to modeling early bilingualism due to the propensity for mutual

influence between languages. Instead, future research and policy should reflect the interaction between two languages in a shared phonological space during acquisition.

Chapter 5

Mandarin Heritage Speakers' Production and Perception of a Cross-linguistically Similar Vowel Contrast

5.1 Introduction

The previous study on Polish-English bilinguals provided evidence for a shared phonological environment for both the heritage language and the majority language. An implication of this is the mutual influence of categories and contrasts between each language. Specifically, the the similar contrasts of English /i/-/ɪ/ and Polish /i/-/ɨ/ exhibited a degree of assimilation across temporal cues. However, in order to make more robust claims about cross-linguistic influence and the interactions between the heritage and majority languages, a second linguistic scenario must be investigated.

A historical flaw of studies and projects that sought to model the phonetics and phonology of bilingualism is that their scope is regularly, and necessarily in many cases, limited to one specific scenario. The limitation of these historical studies arises from factors such as difficulty in finding participants that are comparable in age and circumstance, but who speak different languages, particularly for childhood studies. Therefore, the second experiment that comprises this thesis mirrors the design of the first. The objective of this is to enable more confident generalisations of the results should they reflect the findings of the first experiment. On the other hand, if the results are not replicated, then it is a suggestion that researchers should hesitate to make broad claims from

empirical data that does not include multiple scenarios.

In the interests of a comparable analysis, the second experiment was also planned to examine cross-linguistically similar vowel contrasts. Given the large population of Chinese-born and Chinese heritage residents in the UK, this demographic was chosen due to the perceived reliability of a acquiring a sample, as well as the occurrence of similar monophthongal vowel contrasts that can be compared across similar parameters.

5.2   Chinese Population in the UK

The population of Chinese nationals the among the largest overseas-born populations in the UK, with approximately 198,000 members as of 2019 (The Office for National Statistics, 2019). It is also the second fastest growing of these populations (The Office for National Statistics, 2019). Li (2017) and The UK Federation of Chinese Schools (2020) explain that there is a high number of Chinese complementary schools in the UK that teach Mandarin Chinese. These factors present Mandarin Chinese heritage bilinguals as a useful population for a replication study due to a suitable size from which to sample. It is also useful for informing any relevant pedagogical and medical strategies related to the demographic's speech and language as it continues to grow.

The Chinese population in the United Kingdom has a long history. Understanding this history is essential to understanding the cultural and social effects experienced by bilinguals in the British Chinese community. The first Chinese presence in the in the United Kingdom was in the late 18th Century (Benton and Gomez, 2011). International migration on a large scale from China to the United Kingdom followed this contact in the early 19th Century.

The Chinese ethnicity is heterogeneous, comprising various regions and dialects. Whilst Cantonese was historically the more common Chinese dialect among British Chinese, (Li, 2017) notes the rapidly changing state of this relationship. Migrants from Hong Kong and Southern China are most often Cantonese speakers, whilst those from other areas of the mainland and Taiwan are mostly Mandarin speakers(Li, 2017; Benton and Gomez, 2011). Mostly, a family will speak one dialect at home (Li, 2017). Li (2017) suggests that Mandarin Chinese is becoming the "lingua franca" in this population due to an increase in immigration from the People's Republic of China (PROC), and the elevated language status of Mandarin as the PROC grows in economic power.

Within the Chinese community, as with the Polish and other international communities, there are local schools and activity centres run by community members (UK Federation of Chinese Schools, 2020). The UK Federation of Chinese Schools recognised approximately 80 such institutions, and has over 10,000 students (UK Federation of Chinese Schools, 2020). These centres are often operated with the goal of connecting local families with each other, as well as preserving cultural and linguistic heritage among the youth of the community.

5.3   Chinese and English Monophthongs

There are multiple descriptions of the Standard Chinese vowel system, each attesting various numbers of phonological vowels. Duanmu (2007) and Cheng (2011) describe five phonological vowels, whilst Lee and Zee (2003) describe six. The Mandarin Chinese monophthong inventory, according to the six-phoneme description by Lee and Zee (2003), is depicted below in Figure 5.1.

Figure 5.1: Chinese Monophthongs from Lee and Zee (2003)
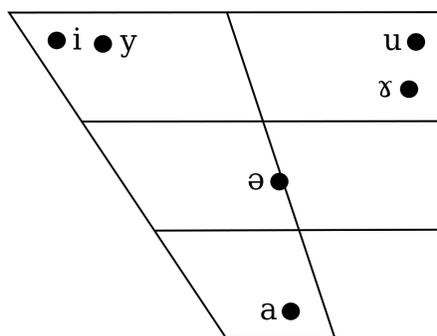
The phonetic realisation of these phonemes may vary considerably between phonetic environments, and this tendency is the source of disagreement in the literature regarding phoneme inventory. Specifically, syllable coda dictates the allophone employed for each phoneme. For example, the phoneme notated as /ə/ may be realised as /e/, /o/, and /ɤ/ in different syllables. Lee and Zee (2003), in the analysis of these categories, describes them as separate phonemes due to the complementary distribution. However, the description by Duanmu (2007); Cheng (2011) considers these various realisations as allophones of /u/. Whilst there is disagreement regarding the number of phonemes in Standard Chinese, both accounts maintain that there is the same set of rhymes, and differ in their distribution of phonemic status.

Mandarin Chinese is a tonal language. The standard Chinese tonal system contains four separate tonal contours which distinguish lexical items. The tonal inventory comprises a high tone, rising tone, dipping tone, and falling tone. These are tabulated in Table 5.1.

Table 5.1: Mandarin Chinese Tones

Tone Number	Tone Label	IPA Transcription	Pinyin Example
1	High	ó	ō
2	Rising	ǒ	ó
3	Dipping	ò	ǒ
4	Falling	ô	ò

In comparison, the monophthongal inventory of British English in considerably larger. Wells and Wells (1982) identifies 11 monophthongal vowels in Received Pronunciation (RP) British English, which are shown with a vowel chart in Figure 5.2.

Figure 5.2: RP English Monophthongs from Roach (2004)
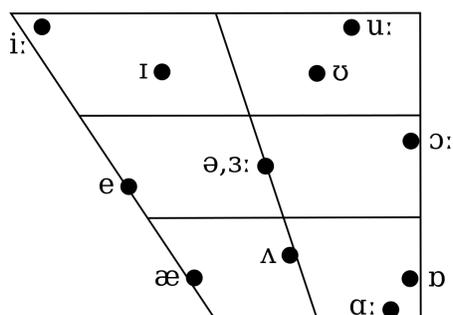
5.4 Similar Vowel Contrast in Chinese and English

The similar scenario occurs as a two-to-two mapping between L1 and L2 sounds (Escudero, 2005; Chang, 2009). Figure 2.1 shows examples presented by Escudero (2005) that depict a visual representation of the two-to-two mapping.

For Mandarin Chinese and English, there are multiple similar scenarios involving both vowels and consonants. However, the allophony present in Mandarin Chinese may complicate a simple phoneme-to-phoneme comparison. Care was taken to accommodate Chinese vowel allophony, and the fact that not all coda and nucleus combinations are present in the Chinese lexicon.

Given the presence of research already existing involving the Mandarin /u/-/y/ contrast by Chang et al. (2011) as produced by heritage speakers of Mandarin Chinese and American English, this contrast was identified as a useful basis for finding a similar contrast for the experiment. Therefore, the second study investigates /u/-/y/ acquired by heritage bilinguals of Mandarin Chinese and British English, which is similar to the English contrast /u/-/ʊ/ (Roach,

2004). The study seeks to evaluate the state of cross-linguistic influence within these speakers in both production and perception scenarios.

These vowel contrasts were chosen to mirror the pair of vowel contrasts chosen in the Polish experiment. The contrasts are canonically produced in the back of the vowel space, which offers the potential for comparison to the previous study, the contrasts in which were in the front of the vowel space. Once again, it may be hypothesised that the duration component of the English tense-lax contrast would influence the Chinese contrast, which does not feature duration.

A further point of interest in this comparison is the observed fronting of /u/ in British English, and how such a trend might influence assimilation patterns for this particular pair of contrasts (Haddican et al., 2013; Cheshire et al., 2008). Whilst /u/ is phonologically backer in terms of historical English and is still transcribed as such, /u/ production in contemporary British English may be acoustically closer to /y/, and a congruent effect has been observed by Lewis et al. (2019). In Lee and Zee (2003), /y/ is described as occurring far closer to /i/ in the Chinese vowel space.

Similarly, previous research has described the English /u/ as relatively advanced when compared to similar phonemes in other languages. This interaction may complicate the cross-linguistic influence, but may also provide evidence for identifying any biases towards abstract phonological knowledge or acoustical similarity.

The hypotheses are the same as the Polish study, and can be seen below.

5.5 Hypotheses

RQ 1. Do heritage bilingual children maintain contrasts in both languages in production and perception? Hypothesis: Yes. Due to regular exposure and practice from a young age, heritage bilinguals will maintain within-language contrasts.

RQ 2. Do heritage bilingual children produce and perceive Chinese differently to monolingual Chinese children? Hypothesis: Cross-linguistic influence will lead to differences between the heritage bilinguals and monolinguals.

RQ 3. Do heritage bilingual children produce and perceive English differently to monolingual English children? Hypothesis: Cross-linguistic influence will lead to differences between the heritage bilinguals and monolinguals.

5.6 Methodology

The experiment was designed to investigate the acoustic cue-weighting of cross-linguistic phonological contrasts employed by the children during language use. The experiment conditions were separated by language and by production and perception. The experiment was separated into an English block including production and perception, which was then followed by a Chinese block. The languages were separated in order to accommodate language modes in the ex-

perimental design, which would allow focused activation of the language-specific contrasts (Grosjean, 2012a, 2001; Antoniou et al., 2011). Further to this, each language block was delivered in the language of its subject for the same language modes effect.

5.6.1 Production

A picture-naming task was presented using a flipbook containing simple cartoon images. Children were asked "What is in the picture?" and shown the images in sequence. In the event that a child could not identify the image, they were given hints that did not reveal the pronunciation, in order to avoid imitation. Images were chosen over orthography to avoid any "sounding out" of the stimuli. The images all depicted lexical nouns that the children were likely to be familiar with in each language. For each vowel, there were 8 words, and each flip-book contained a filler word. For each vowel, there were six words plus a filler word. The stimuli were spoken four times each and recorded using Audacity (Audacity Team, 2019) and a lapel microphone routed through a Sound Devices USBPre2 soundcard, connected to the researcher's laptop at a sampling rate of 44100 Hz.

The Chinese words chosen for the production task all feature the falling tone. This was a deliberate choice made in order to match the tonal contours of English words spoken in an isolated context. This approach was informed by the work of Chang et al. (2011), who identified this approach as effective for cross-linguistic comparisons between Mandarin and English in isolation.

5.6.2 Perception

For the perception condition, within-language minimal pairs were used. For English, the endpoints were <suit> and <soot>. For Chinese, the endpoints were <lu> and <lü>, which are phonemically /lu/ and /ly/. The English endpoints were recorded and spoken by the researcher who is a native speaker of southern standard British English. The Chinese endpoints were recorded by the researcher and spoken by a male native speaker of Mandarin Chinese who is a resident of the UK.

5.6.3 Participants

The participants for the study were 49 children aged 7;0 to 9;6 ($\bar{x} = 8;3$). Within the finals dataset, there were 15 heritage bilinguals of Chinese and English ($\bar{x} = 8;3$, $\sigma = 0;7$), 18 English monolinguals ($\bar{x} = 8;3$, $\sigma = 0;7$), and 15 Chinese monolinguals ($\bar{x} = 9;2$, $\sigma = 0;5$). No children were reported to have any language, motor, or hearing impairment. The details of the sample are outlined below.

The Chinese heritage participants were UK residents and regular pupils of Chinese supplementary schools run by the local Chinese community at weekends. This provided some assurance that the children had regular Chinese input

and practice with both peers and adults. All but one of the participants' parents were Mandarin Chinese L1, English L2 speakers, with some also speaking Cantonese. There were two participating schools in the experiment, both of which were located in the North of England. Mandarin Chinese was the primary home language for all participants, with English being the language used in the day-to-day education and social activities. The research took place in a quiet room on the school premises.

The English monolinguals were recorded at a primary school in the South of England. Chinese monolingual data was collected by a research assistant at a primary school in the Fujian province of mainland China. These data were recorded in a quiet classroom environment using the research assistant's laptop and an external microphone. This experiment was presented with the same procedure as that which was presented to the heritage bilinguals and English monolinguals, written in PsychoPy for convenient sharing (Peirce et al., 2019). All data was collected with informed parental consent using a consent form and information sheet.

5.7   Analysis

For each token, measurements of $F2$, $F1$, $F2 - F1$ (post-calculation) and *duration* were extracted. These were scaled using the `scale` function in R R Core Team (2018). This normalization allowed the data to be compared across speakers, whose anatomy may differ, and improved modeling by giving the variables a consistent scale. Formant frequencies were extracted using Praat at the midpoint of each vowel. The formants were detected in Praat by setting the `Maximum formant (Hz)` to 5000, and the `Number of formants` to 3. The same measurements were recorded for the perception stimuli, so that the results can be compared between conditions.

To compare spectral and temporal cues, formant frequencies were reduced to one dimension by using F2-F1 to represent the spectral cue. This measure was chosen because both contrasts feature these acoustic dimensions. The trajectory between each contrast involves moving the tongue back and lower. Both of these movements lower F2-F1, making it a reasonable method of reducing the dimensions of the spectral cue (Kuhl et al., 1997; Watt and Fabricius, 2002).

This section tests whether the contrast was maintained reliably by the participants, and on which dimensions. Before the individual cues are addressed, the analysis will establish the existence of a reliable contrast irrespective of individual cues.

A nested model comparison was conducted using ANOVA, and comprising subsetted models for each of the acoustic cues. A nested ANOVA model comparison was carried out against the null model in order to assess the reliability of the contrast under each condition. Firstly, a null model was specified using the `lme4` package in R. The null model was specified as a generalized linear model with `vowel` as the dependent variable, no fixed effects, and `subject` as a random effect. Secondly, a generalized linear model comprising the same depen-

dent variable, `F2-F1` and `duration` as fixed effects, and `subject` as a random intercept. A significant result here suggests that the contrast was maintained reliably for that condition, given that the comparison model classifies the dependent variable more effectively. Both acoustic cues are tested within this first model comparison in order to establish whether the contrast in maintained on any dimension.

5.7.1 Do heritage bilingual children maintain contrasts in both languages in production and perception?

Table 5.2: ANOVA results of a null generalized linear regression model compared to model with F2-F1 and duration cues as fixed effects

Population	Language	Condition	χ^2	p
Heritage Bilingual	English	Production	136.46	<0.001
		Perception	198.64	<0.001
	Chinese	Production	82.013	<0.001
		Perception	314.48	<0.001
Monolingual	English	Production	248.7	<0.001
		Perception	175.13	<0.001
	Chinese	Production	345.18	<0.001
		Perception	268.44	<0.001

Table 5.2 shows the results of the nested model comparison between a null model and a model including both acoustic cues as fixed effects, with the vowel as the response variable. Whilst the research question above refers specifically to bilinguals, the monolingual effects are included for reference. A significant result suggests that the model with fixed effects is more effective at predicting the vowel than the null model. Therefore, with such a result the contrasts would be maintained by the participants. The analysis produced significant results for all conditions ($p < 0.001$), supporting the hypothesis that contrasts would be maintained. The heritage speakers were able to maintain the contrasts in both languages, and so were the monolinguals.

Figure 5.3 depicts the distribution and ranges of duration and formants for the production condition. There are evident differences between both spectral and temporal cues in English. The plots show that the effect is seen in a longer /u/ vowel in English. In Chinese, there is mainly a difference in formants according to the plots.

Table 5.3 shows the results of the interaction model comparison. It suggests that there is a significant difference across conditions for both languages for heritage bilinguals (English: $\chi^2 = 306.33$, $p < 0.001$)(Chinese: $\chi^2 = 19.123$, $p = 0.001$). Similarly, for the monolinguals, both contrasts showed significant effects (English: $\chi^2 = 218.39$, $p < 0.001$)(Chinese: $\chi^2 = 5.1881$, $p = 0.159$) from condition. This result suggests that the contrasts are perceived with different cue weight than those with which they are produced.

Figure 5.3: Violin plots of acoustic cues in production
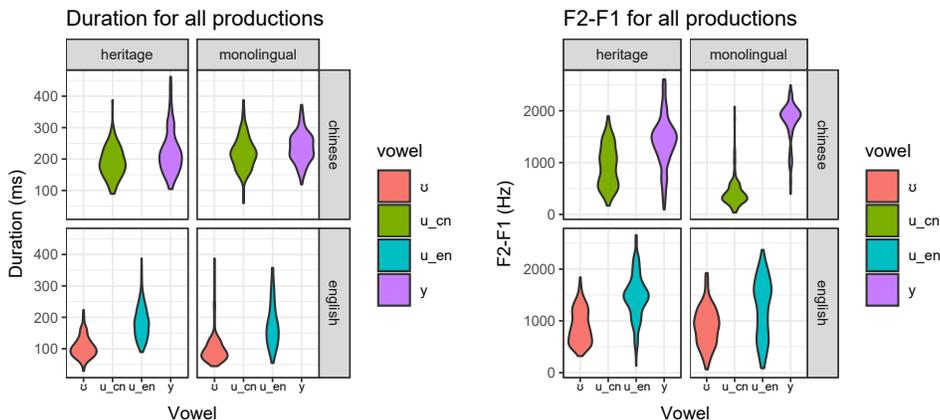
Table 5.3: ANOVA of null generalized linear regression model compared to a model with a duration*condition interaction as fixed effects

Population	Language	χ^2	p
Heritage Bilingual	English	306.33	<0.001
	Chinese	19.123	<0.001
Monolingual	English	218.39	<0.001
	Chinese	36.47	<0.001

To test the degree to which production and perception matched for maintaining contrasts, a null model was fitted to the data, and compared to a model including an interaction between duration and condition, in order to investigate whether the temporal cue differed across conditions. The same null model was used for comparison to a generalized linear model comprising `vowel` as the dependent variable, `duration` and `condition` as interacting fixed effects, and `subject` as a random effect. A significant result here suggests that the effect of duration varied between the production and perception conditions.

Figure 5.4 shows a graphical depiction of heritage speaker productions and perceptions of each language, juxtaposed for visual comparison. The plots include a point for each token as well as an ellipse with 95% confidence intervals. The straight lines depict decision boundaries, representing the predicted boundary separating the vowels acoustically, calculated using generalized linear models, and predicting an outcome of 0.5 for a the binary decision task, thereby estimating the contrast boundary.

Both of the production graphs show steep gradients, which suggests that both acoustic measures involved had considerable cue weight. Whilst there are also gradients apparent on the production graphs, they are much slighter. In

Figure 5.4: Comparison of production and perception in heritage bilinguals, with 50% decision boundaries and 95% confidence ellipses in production
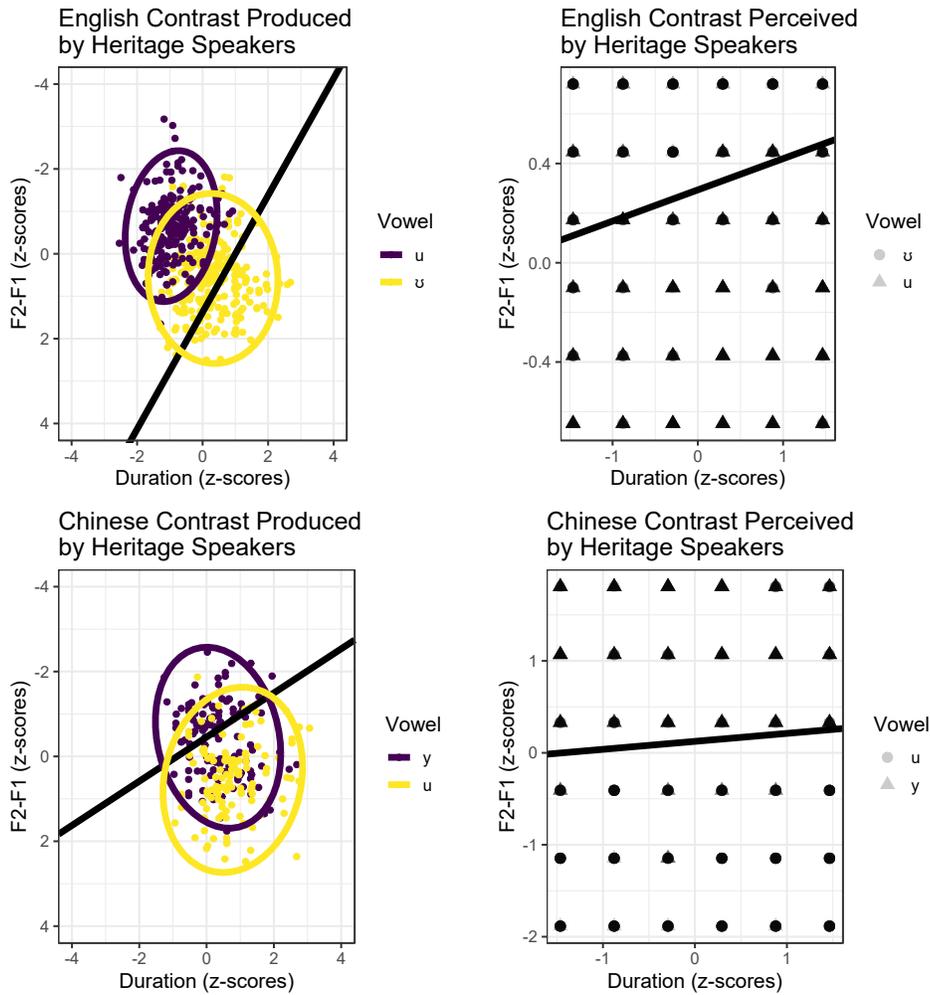
the case of the perception of the Chinese contrast, the gradient is very shallow. The differences between production and perception shown on these graphs are consistent with the results in Table 5.3, which suggest that there are significant differences between production and perception. The visualisation shows that the, for both languages, the temporal cue has less weight in perception.

5.7.2 Do heritage bilingual children produce and perceive Chinese differently to monolingual Chinese children?

Table 5.4 shows the results of a nested model comparison for each cue and condition. The comparison is a likelihood ratio test involving the full model, including both duration and F1-F2 as fixed effects, and a nested model in which one cue was removed. If a significant difference is present, then the removed cue has a significant effect on the contrast. For the spectral cue, this is significant in all cases ($p < 0.001$), suggesting that the spectral cue is contrastive across all conditions.

Table 5.4: ANOVA results of a generalized linear mixed model with both acoustic cues as fixed effects, compared to nested models, each with one cue removed

Population	Language	Condition	Cue	χ^2	p
Heritage Bilingual	Chinese	Production	Spectral	63.6	<0.001
			Temporal	12.337	<0.001
		Perception	Spectral	313.78	<0.001
			Temporal	1.419	<0.234
Monolingual	Chinese	Production	Spectral	327.96	<0.001
			Temporal	1.959	0.162
		Perception	Spectral	267.18	<0.001
			Temporal	2.630	0.105

The model comparison in Table 5.4 does not show significant weight of the duration cue in some of the temporal conditions. Notably, the contrast when perceived by heritage bilinguals did not show a significant temporal weight ($\chi^2 = 1.419$, $p = 0.234$). For the monolingual population, the temporal cue did not have a significant effect in both production and perception ($\chi^2 = 1.959$, $p = 0.162$). This suggests that the spectral cue alone was sufficient to maintain the contrasts in each task, but that the duration differed between populations. The results suggest that heritage children produce the Chinese contrast differently (with a temporal cue).

A null model was fitted to the data, and compared to a model including an interaction between duration and population, in order to investigate whether the temporal cue differed across populations.

Table 5.5: ANOVA of null generalized linear mixed regression model compared to one with duration*population as a fixed effect

Condition	Language	χ^2	p
Production	Chinese	37.643	<0.001
Perception	Chinese	10.595	0.014

Table 5.5 shows the results of a model comparison between the null model and one with an interaction between duration and population. The table shows

a significant effect of population in the production condition ($\chi^2 = 10.602$, $p = 0.014$), but not in perception ($\chi^2 = 0.4112$, $p = 0.938$).

Figure 5.5: Comparison of heritage bilingual and monolingual Chinese production, with 50% decision boundaries and 95% confidence ellipses for each vowel
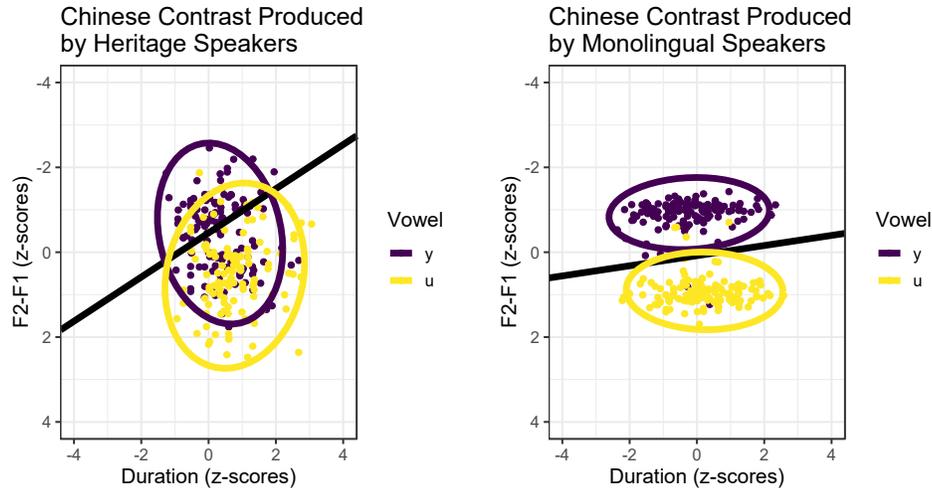
Figure 5.5 shows a side-by-side comparison of heritage bilingual and monolingual Chinese productions. In combination with Figure 5.3, it shows a strong difference between how the contrast in produced by each population. The heritage speakers have a high degree of overlap between their productions, which suggests a degree of a many-to-one assimilation. The Chinese monolinguals have distinct productions that differ on the spectral axis.

Figure 5.6 shows a comparison of heritage bilingual and monolingual perception results or Chinese. The quantitative analysis suggested significant difference between these results. This is corroborated by significance in the models in Table 5.5. The differences between the perceptual boundaries are found in both the temporal cue weight and the spectral position of the boundary, with monolingual speakers having what appears to be a higher, fronter /y/.

5.7.3 Do heritage bilingual children produce and perceive English differently to monolingual English children?

Table 5.6 shows the results of a nested model comparison for each cue and condition. The comparison compared the full model, with F2-F1 and duration as fixed effects, to a nested model in which one cue was removed, in order to test for a significant difference. If a significant difference is present, then the removed cue has a significant effect on the contrast. For the spectral cue, this is significant in all cases ($p < 0.001$ in all cases), suggesting that the spectral cue is contrastive across all conditions. Likewise, the temporal cue is significant in

Figure 5.6: Comparison of heritage bilingual and monolingual Chinese perception with 50% decision boundaries
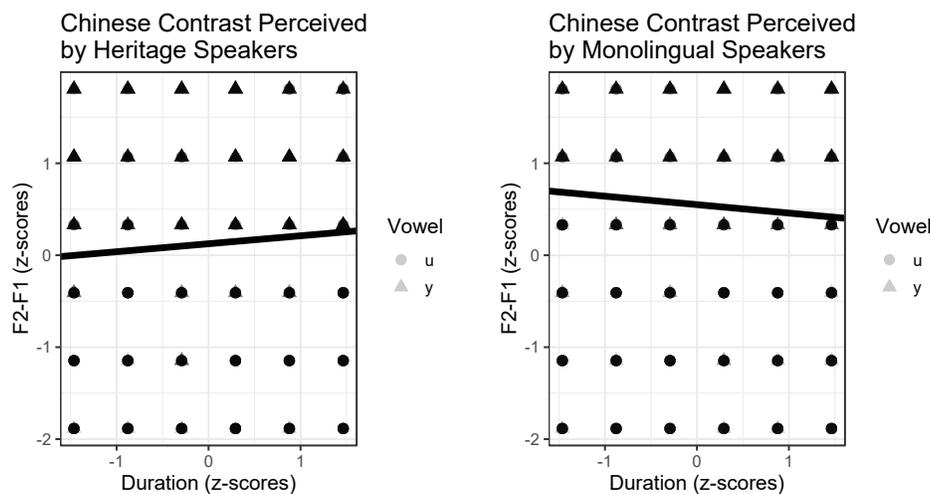
all conditions ($p < 0.001$ in all cases except monolingual perception, for which $p = 0.006$).

Table 5.6: ANOVA results of a full model with both cues compared to nested models, each with one cue removed

Population	Language	Condition	Cue	χ^2	p
Heritage Bilingual	English	Production	Spectral	136.46	<0.001
			Temporal	192.52	<0.001
		Perception	Spectral	190.27	<0.001
			Temporal	13.018	<0.001
Monolingual	English	Production	Spectral	53.261	<0.001
			Temporal	176.65	<0.001
		Perception	Spectral	170.04	<0.001
			Temporal	7.544	0.006

Table 5.7: ANOVA of null model compared to one with duration*population as a fixed effect

Condition	Language	χ^2	p
Production	English	482.4	<0.001
Perception	English	16.153	<0.001

The model comparison in Table 5.6 shows a significant effect of the addi-

tion of duration in both English conditions (heritage bilingual: $\chi^2 = 96.63$, $p < 0.001$)(monolingual: $\chi^2 = 59.249$, $p < 0.001$). This suggests that English production of the contrast does use the temporal cue for both monolinguals and heritage bilinguals.

Figure 5.7: Comparison of heritage bilingual and monolingual English productions, with 50% decision boundaries and 95% confidence ellipses
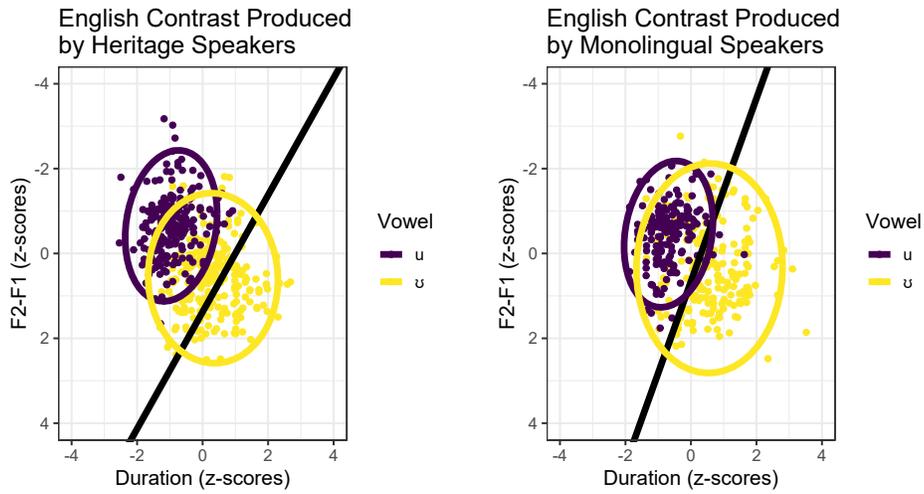

Figure 5.8: Perception results across populations and languages, with 50% decision boundaries
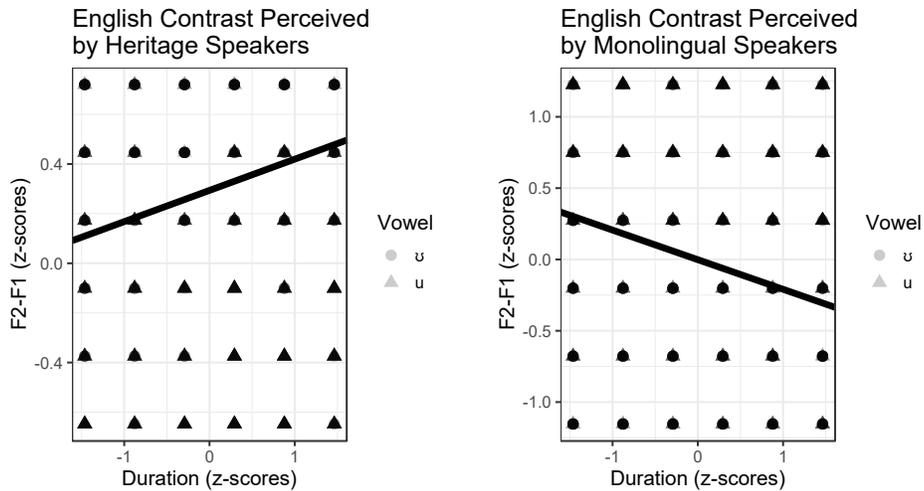
Table 5.7 shows the results of a model comparison between the null model

and one with an interaction between duration and population (heritage and monolingual. The table shows a significant effect of population in both conditions (production: $\chi^2 = 482.4$, $p =< 0.001$, perception: $\chi^2 = 16.153$, $p =< 0.001$). These results suggest that the heritage bilinguals produce and perceive the English contrast with different temporal cue weight when compared to English monolinguals.

Figure 5.7 shows a side-by-side comparison of heritage bilingual and monolingual English productions. The temporal cue is plotted on the x axis, and the spectral cue on the y axis. As above, ellipses with a 95% confidence interval are plotted for each vowel, and a decision boundary that predicts relative cue weighting has been calculated. This graphical depiction enables interpretation of the significant differences between populations. Such differences can be seen most saliently in the increased spectral overlap for the English monolinguals, who appear to rely more on the temporal cue than the heritage bilinguals.

Figure 5.8 shows a comparison of heritage bilingual and monolingual perception results for English. The quantitative analysis suggested a significant difference between these results. The differences suggested can be seen clearly in the graphs, which have opposite slopes much like those in the Chinese contrast comparison above in Figure 5.6.

5.8   Discussion

This discussion will address the research questions before offering a brief comparison of the findings to knowledge of sequential and simultaneous bilinguals. This discussion will discuss the results of this study in a narrow context, and this will inform the following chapter in order to synthesise model-based conclusions.

5.8.1   Do heritage bilingual children maintain contrasts in both languages in production and perception?

The analysis suggests strongly that that the heritage bilinguals were able to maintain both contrasts reliably in both production and perception. This result affirms the hypothesis that the heritage bilinguals have consistent contrasts in both languages. Whilst the English contrast typically involves both spectral and temporal cues, canonically the Chinese contrast involves only the spectral cue (Duanmu, 2007).

For production, the heritage bilinguals exhibited contrasts across both the spectral and temporal dimensions for both the English and the Chinese contrasts. Furthermore, even though the analysis suggests consistent contrasts for both languages across both cues, the formant charts show spectral overlap. The English /u/-/ʊ/ contrast shows this considerably, with a correspondingly steep decision boundary to reflect the importance of the temporal cue in the presence of such overlap. The spectral distribution for English /u/ is broad enough to be a superset of that of /ʊ/. However, there is a distinct peak at an F2-F1 of approximately 1500 Hz. The Chinese contrast by comparison has somewhat

less spectral overlap, though it is still salient. There is a significant temporal component that attributes longer duration to /u/.

For perception, contrasts may be represented differently to production, such as what has been observed in the previous study and those in the literature, with diminished or absent effects of the temporal cue (Kajouj and Kager, 2019; Bohn and Flege, 1991). The perception of the English contrast featured a temporal cue according to the analysis, and this is supported by the gradient of the decision boundary. The boundary suggests a longer temporal cue for /u/, in accordance with the phonological tense-lax contrast. However, the temporal cue appears to be less important in perception than in production. This is consistent with both the previously attested results within this thesis, Kajouj and Kager (2019), and Bohn and Flege (1991). For the Chinese contrast, the temporal cue did not show a significant effect. Though the gradient of the boundary does suggest a minor temporal effect, it cannot be confidently affirmed following the analysis.

In summary, the contrasts for both production and perception are recognised by the heritage children on at least one cue. In other words, there is no many-to-one assimilation process occurring such that one contrast is confused or inconsistently identified. This result is expected and affirms the hypothesis. The participants of the study are still in regular use and exposure to both the heritage language and the majority language in social contexts. Not only this, but they are in structured education for both. However, this is not to say that the heritage bilinguals are monolingual-like.

5.8.2 Do heritage bilingual children produce and perceive Chinese differently to monolingual Chinese children?

In production, the heritage bilinguals showed a broader spectral distribution for the Chinese contrast when compared to monolinguals, and a significant temporal weighting. This is different to the Chinese monolinguals, who did not exhibit a temporal cue. These observations are evidence of cross-linguistic influence, as it appears likely that the temporal cue that is present in the English contrast has been applied to the Chinese contrast. While the asymmetry of the presence of a temporal cue is noticeable, the spectral distributions show significant differences between the populations. The spectral difference is made apparent effectively in the violin plots and the formant charts. For the Chinese monolinguals, there is almost no spectral overlap between the phonemes, and narrow peaks in the distribution. When compared to the heritage bilinguals, the difference is conspicuous.

Chang et al. (2008) demonstrate contrasting Mandarin Chinese /u/-/y/ strongly involves F2, which would resultantly produce a far lower F2-F1 value for /u/ in comparison to /y/. This effect is apparent from the results of the monolingual speakers, for which the distributions for each vowel are easily distinguished. The heritage speakers' productions of Chinese /u/ are varied from F2-F1 values similar to the monolingual productions, to values approaching 2000Hz. For the monolinguals, these F2-F1 values would be more expected for /y/ productions. An explanation for this effect can be explored by compar-

ing the heritage speakers' Mandarin productions to their English productions. For the vowel pairs, the F2-F1 distributions show considerable resemblance. Spectrally, there may exist a one-to-one mapping between the following pairs: Chinese /u/ and /ʊ/; and English /u/ and /y/. This mapping may appear unintuitive when considering IPA phonemic transcription only, but phonetic effects such as GOOSE-fronting offer explanations for this assimilatory pattern (Jansen, 2010).

With the context provided by the previous section - that the heritage bilinguals do produce and pronounce a contrast for these phonemes - a likely explanation is that there is an assimilatory effect. When comparing the heritage bilingual data between languages, there is a clear similarity between the English and Chinese data. Combined with the difference between heritage bilinguals and monolinguals, the results suggest transfer from English to Chinese across both temporal and spectral cues.

In perception, the data analysis suggests a significant difference between the heritage bilinguals and monolinguals. When understood alongside the production differences, this result appears to be consistent with the evidence of transfer from English to Chinese. As a result the heritage bilinguals maintain a contrast in this scenario, but it is unlike that which a native monolingual would maintain. Instead, it appears to have been assimilated to a significant degree by the English contrast. The direction of this influence can be observed by comparing the production data to the perception data and extrapolating. The Chinese monolinguals produced /y/ with a consistently higher F2-F1 when compared to /u/, while the heritage speakers produced /y/ with a lower F2-F1 more reminiscent of their English /u/.

The Chinese contrast, as produced and perceived by heritage speakers, exhibited a significant level of cross-linguistic influence. Specifically, this influence is assimilatory, by which the Chinese phonemes appear to be mapped onto acoustically corresponding English phonemes.

5.8.3 Do heritage bilingual children produce and perceive English differently to monolingual English children?

The analysis found a significant difference between the heritage bilinguals and the English monolinguals across both cues, and in both conditions.

In production, the English monolinguals displayed a higher degree of spectral overlap than the heritage speakers. Correspondingly, the duration cue is assigned more weight for the monolingual speakers. The primary difference that can be observed spectrally is for the /ʊ/ phoneme. For the English monolinguals, the spectral distribution is noticeably broader than that of the heritage bilinguals, with a comparable lower limit but much higher upper limit. As a result, the range of /u/ productions in F2-F1 almost entirely contains that of /ʊ/. However, the /ʊ/ productions across both populations appear to be relatively similar. Therefore, it seems that there is cross-linguistic influence present, primarily for the English /u/ phoneme. If this is understood with the prior knowledge that Chinese /y/ has been assimilated to some degree by a

fronted English /u/. This is consistent with Chang et al. (2011), who observed a similar pattern for Chinese /y/ and English /u/ mutual influence in heritage bilinguals.

With the consideration of the phonological status of each contrast. The specific interaction between Chinese /y/ and English /u/ suggests that there is a preference for phonetic categories in cross-linguistic influence among the heritage bilinguals, as opposed to abstract phonological contrast. This distinction is reflected in models of bilingual acquisition such as the Speech Learning Model (SLM) (Flege, 1995) and Perceptual Assimilation Model-L2 (PAM-L2) (Best and Tyler, 2007b). The SLM is based on advanced learners and considers phonetic categories as loci of cross-linguistic influence for those learners, whereas the PAM-L2 considers phonological contrasts as such, but is based on naive listeners. In this case, it appears that the heritage bilinguals may show cross-linguistic influence more similar to advanced learners than naive or new ones.

In perception, the heritage speakers exhibited more reliance on the spectral cue when perceiving the English contrast, while there appeared to be a minor effect for monolinguals. These results combined suggest that heritage bilinguals do not simply transfer and maintain phonological relationships, which had been seen in sequential bilinguals. Instead, new relationships are readily available and influence each other. These results affirm the hypothesis that cross-linguistic influence will be observable for the heritage bilinguals.

5.9 Differences in Production and Perception

There were differences between both production and perception in the heritage bilinguals and the English monolinguals. In these cases, the temporal cue is weaker in perception that it is in production, consistent with the Polish-English experiment, in which there was no temporal cue present in perception. For the back vowels considered in this experiment, there existed a stronger degree of spectral overlap when compared to the front vowels observed in the Polish experiment. Therefore, it is intuitive that there may be some presence of the temporal cue even in perception. However, it remains much weaker than in production, and provides evidence for a degree of linearity in this relationship between cues.

A few explanations exist for these interactions. The different goals of production and perception may be a factor in this result. When producing speech, the children would be encouraged to use a pronunciation that maximizes clarity, and this would be achieved by incorporating all cues possible. A perception system has more opportunity to be somewhat utilitarian, making use of deduction such as knowing which language is being perceived. In previous work, Shultz et al. (2012) found little to no correlation between production and perception for an English contrast, and noted that secondary cues may have "practical irrelevance" under certain conditions. In other words, the primary cue and context are often sufficient. Boersma and Chládková (2011) showed that some multidi-

mensional boundaries, in this case using both F1 and F2, were reduced to one dimension in perception. The tendency of the mind to prefer simpler perceptual configurations offers an intuitive explanation to these effects (Chater and Vitányi, 2003). In order to reduce cognitive load, the simpler perceptual task will be chosen where possible. Thus, the use of the temporal cue may only be apparent in cases for which the spectral cue does not suffice, such as in noise or when perceiving accented speech.

5.10   Assessment of Cross-linguistic Influence

The analysis showed that there were significant differences between heritage bilinguals and monolinguals. This occurred in production and perception for both languages. When producing the Chinese contrast, the heritage bilinguals showed more overlap in the range of F2-F1 of each category when compared to the Chinese monolinguals, and a broader range for both categories, which is closer consistent with the English contrast. This suggests that there is influence from English, causing the Chinese contrast to assimilate.

However, the transfer was not unidirectional as evidenced by the significant differences between the English monolinguals and the heritage bilinguals. When referencing the visualisations, the heritage bilinguals appear much closer to the tendencies of English monolinguals in terms of categorical production, suggesting that Chinese categories are being assimilated by English similar categories.

The discussion of the results indicates that the status of heritage bilinguals should not be considered in the same framework as a sequential learner or a naive learner. This is seen in the mutual transfer between languages, as opposed to the transfer of one phonology to the other, and the presence of categorical influence as opposed to total contrast assimilation. Models such as the SLM and PAM-L2 may not be well suited to modeling early bilingualism due to the propensity for mutual influence between languages, but the SLM may provide useful states of categorical influence when considering a majority and heritage language.

Chapter 6

Discussion

The discussion chapter will synthesise the results from the two parallel studies in the project alongside the existing literature. First, a brief quantitative summary will be provided, which will contextualise and contrast the data from each study, with reference to the tables and figures provided in their respective chapters. After this, a full discussion will be presented, addressing the primary theoretical areas that the project has involved.

6.1 What degree of cross-linguistic influence is present in young heritage bilinguals?

The results of the MPhil project, suggest that there is a fairly equal level of influence between each language used by the speakers. This conclusion is derived from considering the proximity of the bilingual participants' mean tokens to those of the monolingual participants. In terms of the formant frequencies, and duration, these tokens were at an approximate midpoint between the monolingual samples in many cases. This result corroborates the conclusions of Dubiel and Guilfoyle (0), who suggests that the point at which the majority language might become preferred is between the ages of 8 and 11;5 years. This could be explained by the gradual change in the relative use of each language as detailed by Montrul and Foote (2014).

Variation is normal and broad among bilinguals, and it is likely that each heritage bilingual undergoes this process at a different rate, and with varying results. Even among the participants of this study, whose relative exposure to English and Polish is relatively consistent, there exists individual variation. Whether this is due to language asymmetry or preference is hard to discern without knowing more information about each child's usage and exposure to each language in considerably more detail. However, this identifies an opening for future research to investigate the more precise effects on language asymmetry. With scales such as Dunn and Tree (2009) in mind, longitudinal work would advance understanding of this interaction.

Language asymmetry is not a static factor in perception and production and is mediated in accordance with the immediate environment of a bilingual.Grosjean (2001) describes the acute shifts in language usage and exposure in the framework of language modes. In a bilingual environment, an individual would be in a corresponding bilingual mode, and speech perception would activate both languages. Therefore, language asymmetry effects might be mitigated when compared to a monolingual mode in which, theoretically, only one language is activated.

There is sufficient cross-linguistic influence that heritage bilinguals differ from monolingual peers in a statistical sense. The cross-linguistic influence was not of a high enough magnitude to, from the researcher's perspective, constitute any difference in proficiency. The observation that the influence was mostly present on different cues is notable. For example, the heritage bilinguals' production of the English contrast appeared to place more weight on the duration cue compared to the English monolinguals. Though the *degree* of cross-linguistic influence appeared small, its presence favoured cues that were asymmetrical between languages. In the previously mentioned case, emphasising the presence of duration in English, which is unique in the sense that it is absent in Polish. Future research may explore this further to examine to what degree human listeners perceive such cross-linguistic influence, and compare it to statistical results such as this.

6.2 Do production and perception align during heritage language acquisition?

Flege (1995) and Best and Tyler (2007a) suggested that second language phonetic acquisition begins with perception when detailing their models of novel category and contrast perception in the SLM and PAM-L2 respectively. In order for a speaker to develop their ability to produce a sound, they must gain the ability to perceive the sound as distinct from their existing phonetic inventory. This occurs over time with repeated exposure to a novel sound, which then enables the speaker to imitate and learn to produce the sound more accurately.

One question that may be raised by researching heritage bilinguals as opposed to the sequential bilinguals on whom the SLM and PAM-L2 are based is whether heritage bilinguals follow this pattern also. A potential complicating factor to this is that the early exposure to both the majority language and the heritage language may preserve articulatory patterns practiced through babbling as protophones (Buder et al., 2013) which may not otherwise be present in older individuals.

The heritage bilingual data in this project shows asymmetry in the contrast boundaries for both perception and production conditions. This asymmetry occurs in both the vowel quality and vowel duration cues. Some variation between the conditions is expected due to the natural methodological differences in testing perception and production. However, the bilingual groups exhibited

a stronger difference between perception and production when compared to the monolingual groups, suggesting that bilingualism is a factor in the asymmetry. For these variations, the bilingual groups' perception contrast boundaries and cue weights were closer to those of the monolinguals than the equivalent production boundaries and cue weights. In this way, the data presented by the thesis study appears to corroborate these perception-leading tenets of the SLM and the PAM as perception leads production as the more monolingual-like condition. Other studies also appear to corroborate (Mack, 1989; Amengual and Chamorro, 2015; McCarthy et al., 2014; Lenneberg, 1964).

De Jong et al. (2009) identify the different mechanics of perception and production progress. Evidence suggests that they are linked, but this is a complex relationship that is further confounded by each process operating with different mechanics De Jong et al. (2009). In terms of cue-weighting, the evidence in the literature and this study suggests that perception and production do not align in this case. This is alluded to in the models of bilingualism such as the SLM-r and PAM-L2.

6.3 Can the existing models of bilingual phonological acquisition predict these outcomes?

6.3.1 Age and Experience

An important revision in the revised Speech Learning Model (SLM-r), explained by Flege and Bohn (2021), is a focus on individual differences between bilinguals. This study showed a small amount of variation in the behaviour of different words in different phonetic contexts. This supports the evidence that variation between bilingual speakers is not exclusive to the segmental level, and that particular words or phrases are affected differently due to contextual use and exposure.

Similarly, the SLM-r removes the age of immersion as a mechanism within the model. Heritage bilinguals, owing to their early exposure to two languages, demonstrate mitigated age of acquisition effects compared to sequential bilinguals. These differences in bilingual acquisition and development are results of early input of the LH. *Input* refers to the combination of *exposure* (processes of interpreting language such as hearing and reading) and *practice* (processes of communicating with language such as speaking and writing). It is possible that age of immersion was, in some cases, confused having more experience with a particular language in prior research. In the context of this research, cross-linguistic influence affects both heritage language and majority language phonetics and phonology bidirectionally. Given that there is mutual cross-linguistic influence between the heritage language and majority language, it follows that the effects of age, or experience, are not limited to affecting the development of one language. Given that neither heritage language group was precisely monolingual-like despite having experience with the heritage language naturalistically since birth, neither age nor experience seems to be the only factor in

native, monolingual closeness.

The age of majority language introduction affects the retainment because of the relative inflexibility of already-established memory networks Montrul (2008); Au et al. (2008). Knightly et al. (2003) and Chang (2016) provide evidence that, as adults, heritage bilinguals (or those exposed to a heritage language during youth) maintain knowledge of the heritage language until much later in life even if they cease use of the heritage language in childhood. The results from the thesis data appear to be consistent with the literature on the trajectories of heritage-majority language balance with regards to age. An explanation could be that after some years of exposure prior to learning the majority language around 3-4 years old, the heritage language remains influential on majority language phonetics as the balance of exposure shifts from heritage language to majority language. The fact that there is variation in the retainment of the heritage language suggest that this is the case, as continued heritage language use with family members still allows further majority language exposure. However, there appears to be a limit to this given that even much later heritage learners or re-learners have a learning advantage (Chang, 2016).

For any process of linguistic acquisition, the quantity and frequency of input the ability to listen to, engage in, and practice producing the target language is crucial to the development of functional abilities in the language (Gathercole, 2002). For example, Gathercole (2002) observed that quantity of input was a sufficient condition for predicting the ability of heritage bilingual children to judge morphosyntax in Spanish and English. When compared to monolingual English speakers, those with higher exposure to English were significantly more likely to judge sentence grammaticality correctly, and more like monolinguals, than those with less exposure.

The existence of input and exposure in bilingualism as a factor in bilingual acquisition is widely discussed across multiple linguistic domains. However, there is comparatively less investigation into the role of input on phonetics and phonology than domains such as grammar and syntax. Not only this, but the mechanisms of and details of the role of input remain a matter of debate. Different models of bilingual phonology represent the acquisition, maintenance, and the long-term evolution of phonetic and phonological boundaries with various quantitative components that operationalise input. For example, the L2LP (Escudero, 2005) models the development of contrasts using the constraint-based framework of optimality theory, whereas the SLM (Flege, 1995) represents this development as statistical category-matching like that which is described in exemplar theory (Ashby and Maddox, 2005).

Whilst the precise psychological expression of input is unclear, a common tenet of the bilingual phonological models discussed previously is that there is a direct correlation between input and monolingual-like function (sometimes referred to as competence). The evidence that exists for phonetic input quantity in heritage bilingual corroborates these conclusions. However, heritage bilinguals are particularly inclined to have unique input profiles compared to others. This is because imbalance of input between each language is not just one of quantity.

75

One of the defining traits of heritage bilingualism is the asymmetry between the languages they use in terms of context, social status, and the balance of exposure and practice. Heritage bilinguals are like to use engage with family members in one language, and friends and education with another. Grosjean (2008) described the Complementary Principle, which suggest that a developing bilingual's vocabulary is distributed between the contexts in which they use each language, as well as language modes, which suggest that a bilingual's accent may shift towards one language or another depending on their linguistic environment.

A heritage language may be associated with social meaning such as prestige or obscurity, which may as a result affect attitudes towards the heritage language. A balance between exposure and practice cannot be presumed either. For example, many heritage bilinguals report a better ability to understand the heritage language as opposed to speaking it, and attribute this to a pattern of hearing the heritage language from parents but responding in the majority language).

The observations made from the data in this thesis from Polish and Chinese heritage bilinguals in the UK demonstrate variation in the phonetic perception and production of contrasts across dimensions both within and between individuals. This variation is evidence possibly related to imbalance between exposure and practice. Motor pattern learning and planning. Perceptual category assimilation.

Heritage speakers in thesis study have consistent, patterned input including at home, school, supplementary school, and with friends, and show bidirectional cross-linguistic influence. Therefore, this input mechanism does not appear to be context-dependent. Otherwise, there might be much more bias towards one language in terms of influence.

6.3.2   One System or Two?

The popular models of bilingual acquisition addressed in this thesis (namely, the SLM, PAM-L2, and L2LP) are built upon the foundational tenet that there exists one shared phonetic system within an individual's psychology. This is supported by evidence of cross-linguistic influence, the existence of which suggests that phonetic information is shared between linguistic systems such that exposure to one language can influence the usage of another. This is especially observable in situations in which exposure to an L2 can influence the phonetic perception and production of an L1. However, much research into cross-linguistic influence is conducted on sequential bilinguals, and it is from data from sequential bilinguals that the SLM, PAM-L2, and L2LP are derived.

Heritage bilinguals may differ in their expression of cross-linguistic influence due to the unique ways in which they learn and are exposed to each language they speak. Gonzales and Lotto (2013) observed Spanish-English simultaneous bilinguals and compared them to English monolinguals in speech perception and suggests evidence for a separation of phonetic systems in bilinguals that are exposed to multiple languages from a young age. Intuitively, these conclusions are not mutually exclusive. The existence of one system, or two separated ones,

may be contingent on the nature of an individual's language exposure. Growing up with multiple languages during the phases of life with the highest neuroplasticity, and while perceptual and articulatory systems are still developing, may fundamentally influence the organisation of these systems. Whilst this project can not address this question directly, future research may seek to investigate this further by including heritage bilinguals as well as children in immersion programs. However, a further issue to answering these potential research questions is the ability to define what a separated system is, and to what degree such separation represents partitioning of the same broader system as opposed to two different ones.

Grosjean (2001) considers the use of different language settings as contextual expressions of the same system. Language modes are defined by Grosjean (2001) as the degree to which one language is "active" as individual's language processing at any given time. The language mode that an individual is in is influenced by the surrounding linguistic environment. Which language(s) an individual expects to produce and perceive is the primary motivator for shifting in language modes. However, it is not the only factor, and other influences such as social prestige, recent language usage, and individual preference may also. Language mode selection is a subconscious process that can be thought of as attuning the language use to the linguistic environment in order to maximise the user's ability to communicate and understand. More specifically, the language mode hypothesis suggests that environmental factors cause individuals to change their perceptual phonemic boundaries or production targets to suit the needs of the situation. This can occur to such a degree that the same acoustic information can be perceived as different phonemes (Magnuson and Nusbaum, 2007).

From this project, the are key result relevant to the discussion of the existence of multiple separated language systems or one partitioned system is in the phonemic boundaries observed in the data. Firstly, there is mutual cross-linguistic influence between both languages for each bilingual sample. This cross-linguistic influence is mostly in the form of convergence, in which the production or perception of a phoneme in one language becomes more acoustically similar to one in the other language. There are two explanations for this result, depending on the initial assumption of whether there are separate systems for each language. In one interpretation, this is evidence for the systematic organisation that Grosjean (2001) describes and is evidence for a shared phonological organisation due to the degree to which one language may influence the other. In a more separated structure, this may not be an expected result, because the systems for each language would not affect each other. The second explanation, if there are two separation "monolingual systems" is that the participants were in a mixed language environment, and that both language systems were active at once. Therefore, there would be mixed results.

However, if the second explanation were the case, the result of a medium point being selected for perceptual boundaries and production targets is an unintuitive one. It would be more likely that such competing systems would lead to results in which there would be bimodal distributions, in which sometimes

one language takes priority and sometimes the other, even between tokens of the same language. Instead, the results showed clear differences when each language was spoken and perceived, but that each one was affected by convergent cross-linguistic influence. The evidence presented here supports the hypothesis that there is a shared phonological system in bilinguals, and that heritage bilinguals follow this pattern.

6.3.3 Methodological Considerations for Developing Models

Modelling the phonetic and phonological development of heritage bilinguals, and bilinguals in a more broad sense, is a complex and challenging task. There is an extremely high number of factors that contribute to phonetic bilingual outcomes, many of which involving other fields of linguistics such as sociolinguistics and syntax. A comprehensive model must consider as many of these factors as possible and their varying effects on the phonetic outcomes. In order to achieve such models, these factors must be explored in experimental research. For many factors this is already the case and research continues to develop theoretical understanding of individual factors. However, there is comparatively less research into the interactions between factors influencing bilingual development, especially for heritage speakers. This is a considerable methodological challenge, so it is clear why there is less research on comprehensive multi-factor modelling and more on modelling single factors. Indeed, it is important to understand individual factors and their outcomes before attempting to understand their interactions and combined effects.

The present thesis study has explored the expansion of modelling methodology by collecting a broad dataset that incorporates a number of vectors for comparison across multiple variables. The study incorporates perception and production conditions in a way that encourages their comparison and compatibility. For both production and perception, monolingual age-matched samples were collected. Finally, the study involved two heritage languages (Polish and Mandarin Chinese) with a shared majority language (British English). In exploring a data set and statistical procedure that combines a variety of comparisons and separate statistics, the challenges of comprehensive modelling methodology are made apparent.

Firstly, the dynamics of comparing production and perception are highlighted. Investigating production and perception with a view to comparing them is complex due to each requiring different methodology. Future research may attempt more varieties of methods in order to discern what is truly different and what is methodological noise. The heritage bilingual data in the project show asymmetry in the contrast boundaries for both perception and production conditions, and both the vowel quality and vowel duration cues. The different goals and mechanisms of production and perception may be a factor in this result. When producing speech, the children would be encouraged to use a pronunciation that maximizes clarity, and this would be achieved by incorporating all cues possible. A perception system has more opportunity to be somewhat

utilitarian, making use of deduction such as knowing which language is being perceived.

De Jong et al. (2009) identify the different mechanics of perception and production progress. Evidence suggests that they are linked, but this is a complex relationship that is further confounded by each process operating with different mechanics and goals De Jong et al. (2009). In previous work, Shultz et al. (2012) found little to no correlation between production and perception for an English contrast and noted that secondary cues may have "practical irrelevance" under certain conditions. In other words, the primary cue and context are often sufficient. Boersma and Chládková (2011) showed that some multidimensional boundaries, in this case using both 19F1 and F2, were reduced to one dimension in perception. The tendency of the mind to prefer simpler perceptual configurations offers an intuitive explanation to these effects (Chater and Vitányi, 2003). In order to reduce cognitive load, the simpler perceptual task will be chosen where possible. Thus, the use of the temporal cue may only be apparent in cases for which the spectral cue does not suffice, such as in noise or when perceiving accented speech. More research into the different goals of production and perception. Would enable more effective methodology for their comparison. Some variation between the conditions is also expected due to the natural methodological differences in testing perception and production. incorporating all cues possible.

However, the bilingual groups exhibited a stronger difference between perception and production when compared to the monolingual groups, suggesting that bilingualism is a factor in the asymmetry. For these variations, the bilingual groups' perception contrast boundaries and cue weights were closer to those of the monolinguals than the equivalent production boundaries and cue weights. In this way, the data presented by the thesis study appears to corroborate these perception-leading tenets of the SLM and the PAM as perception leads production as the more monolingual-like condition. Other studies also appear to corroborate (Mack, 1989; Amengual and Chamorro, 2015; McCarthy et al., 2014; Lenneberg, 1964).

Another methodological challenge for investigations such as this is the limits of the "laboratory" environment that many production and perception experiments take place. Approaches in this environment involve singular word tokens and forced choice contrast identification tasks. Methodology for investigating bilingual development would benefit from case studies that attempt to observe more natural language use, though this comes along with its own difficulties. The intuitive synthesis is that both laboratory and natural environments have benefits for investigating bilingualism.

Diachronic case studies are likely to provide a complement to laboratory work such as the present thesis study and contextualise the individual variation as well as the overall statistically derived outcomes. Such studies can incorporate and inform the understanding of factors such as language attitudes as well as more finely-grained measurements of exposure and practice.

Chapter 7

Conclusion and Contribution

This thesis has provided evidence that young heritage bilinguals exhibit cross-linguistic influence for similar vowel contrasts. The presence of the influence is mostly present in production, which may imply a motor component to the mutual influence such as anterograde interference. This is shown to be a bidirectional process not limited to particular acoustic cues, and may cause assimilation of cues cross-linguistically. Young heritage bilinguals have more in common with simultaneous bilinguals than sequential bilinguals, and the evidence suggests a shared phonological space between languages.

The original contribution made by this thesis is the testing of these research questions across multiple paradigms with consistent methodology. Previous work in the field has tested many languages, conditions and scenarios, but often these are separate studies across many authors, datasets, and age groups. The studies in this thesis stand out due to their testing of these variables simultaneously across age-matched samples. By doing this, questions regarding variability between production and perception can be observed, and further research into the relationship between them in this context can draw from this work. Similarly, testing multiple languages with age matching and monolingual samples from both languages provides the foundations for robust conclusions that can be made across

Bibliography

Altvater-Mackensen, N. and Fikkert, P. (2010). The acquisition of the stop-fricative contrast in perception and production. *Lingua*, 120(8):1898–1909.

Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism *. *Bilingualism: Language and Cognition*, 15(3):517–530.

Amengual, M. (2018). Asymmetrical interlingual influence in the production of Spanish and English laterals as a result of competing activation in bilingual language processing. *Journal of Phonetics*, 69:12–28.

Amengual, M. and Chamorro, P. (2015). The effects of language dominance in the perception and production of the galician mid vowel contrasts. *Phonetica*, 72(4):207–236.

Anderson, J. (2008). Towards an integrated second-language pedagogy for foreign and community/heritage 1 languages in multilingual Britain. *The Language Learning Journal*, 36(1):79–89.

Antoniou, M., Best, C. T., Tyler, M. D., and Kroos, C. (2011). Inter-language interference in vot production by l2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics*, 39(4):558–570.

Arnot, M., Schneider, C., Evans, M., Liu, Y., Welply, O., and Davies-Tutt, D. (2014). School approaches to the education of eal students. *Language development, social integration and achievement. Cambridge: The Bell Educational Trust Ltd.*

Ashby, F. G. and Maddox, W. T. (2005). Human category learning. *Annu. Rev. Psychol.*, 56:149–178.

Au, T. K.-F., Knightly, L. M., Jun, S.-A., and Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science*, 13(3):238–243.

Au, T. K.-F., Oh, J. S., Knightly, L. M., Jun, S.-A., and Romo, L. F. (2008). Salvaging a Childhood Language. *Journal of Memory and Language*, 58(4):998–1011.

Audacity Team (2019). *Audacity: Free Audio Editor and Recorder.*

Baker, C. and Jones, S. P. (1998). *Encyclopedia of bilingualism and bilingual education.* Multilingual Matters.

Benmamoun, E., Montrul, S., and Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, 39(3):129–181.

Benton, G. and Gomez, E. T. (2011). *Chinese in Britain, 1800-present.* Springer.

Berken, J. A., Gracco, V. L., and Klein, D. (2017). Early bilingualism, language attainment, and brain development. *Neuropsychologia*, 98:220–227.

Best, C. (1995). A direct realist view of crosslanuagage speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, pages 171–204. Timonium, New York Press.

Best, C. and Tyler, M. (2007a). Nonnative and second-language speech perception: Commonalities and complementarities. In Munro, M. J. and Bohn, O.-S., editors, *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*, pages 13–34.

Best, C. T. et al. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words*, 167(224):233–277.

Best, C. T. and Tyler, M. D. (2007b). Nonnative and second-language speech perception: Commonalities and complementarities. *Language experience in second language speech learning: In honor of James Emil Flege*, 1334:1–47.

Bialystok, E. and Hakuta, K. (1994). *In other words.* New York: Basic Books.

Bialystok, E. and Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In Birdsong, D., editor, *Second language acquisition and the critical period hypothesis*, pages 161–181. Lawrence Erlbaum, Mahwah, NJ.

Birdsong, D. (2019). Age and the end state of second language acquisition.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10):341–345.

Boersma, P. and Chládková, K. (2011). Asymmetries between speech perception and production reveal phonological structure. In *ICPhS*, pages 328–331.

Bohn, O.-S. and Flege, J. E. (1991). Interlingual Identification and the Role of Foreign Language Experience in L2 Vowel Perception. *Applied Psycholinguistics*, 11(3):303–28.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In Birdsong, D., editor, *Second Language Acquisition and the Critical Period Hypothesis*, pages 133–159. Erlbaum, Mahwah, NJ.

Bosch, L. and Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics*, 39(4):514–526.

Bosch, L. and Sebastián-Gallés, N. (2003). Simultaneous Bilingualism and the Perception of a Language-Specific Vowel Contrast in the First Year of Life. *Lang Speech*, 46(2-3):217–243.

Brown, R. and Berko, J. (1960). Word association and the acquisition of grammar. *Child development*, pages 1–14.

Buder, E. H., Warlaumont, A. S., and Oller, D. K. (2013). An acoustic phonetic catalog of prespeech vocalizations from a developmental perspective. *Comprehensive perspectives on child speech development and disorders: Pathways from linguistic theory to clinical practice*, 4:103–34.

Byun, T. M.-I. (2016). The A-map model: Articulatory reliability in child-specific phonology. *Language*, 92(1):141–178.

Cabo, D. P. Y. and Rothman, J. (2012). The (Il)Logical Problem of Heritage Speaker Bilingualism and Incomplete Acquisition. *Applied Linguistics*, 33(4):450–455.

Chang, C. (2019). *The phonetics of second language learning and bilingualism*, page 427–447.

Chang, C., Haynes, E., Yao, Y., and Rhodes, R. (2008). The phonetic space of phonological categories in heritage speakers of mandarin. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 44, pages 31–45. Chicago Linguistic Society.

Chang, C. B. (2009). Native language phonetic drift in beginning second language acquisition. *The Journal of the Acoustical Society of America*, 126(4):2312–2312.

Chang, C. B. (2016). Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition*, 19(4):791–809.

Chang, C. B., Yao, Y., Haynes, E. F., and Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America*, 129(6):3964–80.

Chater, N. and Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in cognitive sciences*, 7(1):19–22.

Cheng, C.-C. (2011). *A synchronic phonology of Mandarin Chinese*, volume 4. Walter de Gruyter.

Cheshire, J., Fox, S., Kerswill, P., and Torgersen, E. (2008). Linguistic innovators: the english of adolescents in london. *Final report submitted to the ESRC*.

Clarke, C. M. and Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116(6):3647–3658.

Colantoni, L. (2015). *Second language speech: theory and practice.* Cambridge University Press, Cambridge, United Kingdom.

Curdt-Christiansen, X. L. (2020). Educating migrant children in england: language and educational practices in home and school environments. *International Multilingual Research Journal*, 14(2):163–180.

De Houwer, A. (2009). *Bilingual first language acquisition.* Multilingual Matters.

De Houwer, A. (2011). Language input environments and language development in bilingual acquisition. *Applied Linguistics Review*, 2:221–240.

De Jong, K., Hao, Y.-C., and Park, H. (2009). Evidence for featural units in the acquisition of speech production skills: Linguistic structure in foreign accent. *Journal of Phonetics*, 37(4):357–373.

Dekeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22(4):499–533.

Duanmu, S. (2007). *The phonology of standard Chinese.* OUP Oxford.

Dubiel, B. and Guilfoyle, E. (0). Early shifts in the heritage language strength: A comparison of lexical access and accuracy in bilingual and monolingual children. *International Journal of Bilingualism*, 0(0):1367006920936337.

Dunn, A. L. and Tree, J. E. F. (2009). A quick, gradient bilingual dominance scale. *Bilingualism*, 12(3):273.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968):303–306.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization.* Netherlands Graduate School of Linguistics.

Escudero, P. and Boersma, P. (2004). Bridging the gap between l2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26(4):551–585.

Escudero, P. and Boersma, P. (2007). Bridging the Gap between L2 Speech Perception Research and Phonological Theory. *Studies in Second Language Acquisition*, 26(4):551–585.

Fabiano-Smith, L. and Barlow, J. A. (2010). Interaction in bilingual phonological acquisition: evidence from phonetic inventories. *International Journal of Bilingual Education and Bilingualism*, 13(1):81–97. PMID: 20126516.

Fabiano Smith, L. and Goldstein, B. A. (2010). Early-, Middle-, and Late-Developing Sounds in Monolingual and Bilingual Children: An Exploratory Investigation. *American Journal of Speech-Language Pathology*, 19(1):19(1), 66–77.

Flege, J. and Bohn, O.-S. (2021). *The Revised Speech Learning Model (SLM-r)*, pages 3–83.

Flege, J. E. (1995). Second-language Speech Learning : Theory, Findings, and Problems. In Strange, W., editor, *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, pages 229–273. York Press, Timonium, MD.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. *Laboratory phonology*, 9:353–382.

Flege, J. E., Frieda, E. M., and Nozawa, T. (1997). Amount of native-language (l1) use affects the pronunciation of an l2. *Journal of phonetics*, 25(2):169–186.

Flege, J. E., Schirru, C., and MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech communication*, 40(4):467–491.

Flege, J. E., Yeni-Komshian, G. H., and Liu, S. (1999). Age constraints on second-language acquisition. *Journal of memory and language*, 41(1):78–104.

Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., and Hallé, P. (2008). Cross language phonetic influences on the speech of French–English bilinguals. *Journal of Phonetics*, 36(4):649–663.

Gahl, S. and Strand, J. F. (2016). Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. *Journal of Memory and Language*, 89(C):162–178.

García, O. and Wei, L. (2014). Language, bilingualism and education. In *Translanguaging: Language, bilingualism and education*, pages 46–62. Springer.

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., Ortiz, S., and Kuhl, P. K. (2011). Bilingual language learning: An erp study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, 39(4):546–557.

Gathercole, V. (2002). *Monolingual and bilingual acquisition: Learning different treatments of that-trace phenomena in English and Spanish*, pages 220–254.

Genesee, F. (2015). Myths About Early Childhood Bilingualism. *Canadian Psychology/Psychologie canadienne*, 56(1):6–15.

Gertken, L. M., Amengual, M., and Birdsong, D. (2014). Assessing language dominance with the bilingual language profile. *Measuring L2 proficiency: Perspectives from SLA*, pages 208–225.

Gonzales, K. and Lotto, A. J. (2013). A bafri, un pafri: Bilinguals' pseudoword identifications support language-specific phonetic systems. *Psychological science*, 24(11):2135–2142.

Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. *Tutorials in bilingualism: Psycholinguistic perspectives*, pages 225–254.

Grosjean, F. (1998). Transfer and language mode. *Bilingualism: Language and Cognition*, 1(3):175–176.

Grosjean, F. (2001). The bilingual's language modes. *One mind, two languages: Bilingual language processing*, 122.

Grosjean, F. (2008). *Studying bilinguals.* Oxford University Press.

Grosjean, F. (2012a). An attempt to isolate, and then differentiate, transfer and interference. *International Journal of Bilingualism*, 16(1):11–21.

Grosjean, F. (2012b). Bilingual and monolingual language modes. *The encyclopedia of applied linguistics.*

Grosjean, F. (2013). Bilingual and monolingual language modes. *The encyclopedia of applied linguistics*, pages 489–493.

Haddican, B., Foulkes, P., Hughes, V., and Richards, H. (2013). Interaction of social and linguistic constraints on two vowel changes in northern england. *Language Variation and Change*, 25(3):371–403.

Hakuta, K. (1986). *Mirror of Language. The Debate on Bilingualism.* ERIC.

Hazan, V. and Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4):377–396.

Herdina, P., Jessner, U., and Vaid, J. (2004). A dynamic model of multilingualism: perspectives of change in psycholinguistics. *World Englishes*, 23(3):491–492.

Højen, A. and Flege, J. E. (2006). Early learners' discrimination of second-language vowels. *The Journal of the Acoustical Society of America*, 119(5):3072–3084.

Hossein Keshavarz, M. and Ingram, D. (2002). The early phonological development of a farsi-english bilingual child. *International journal of bilingualism*, 6(3):255–269.

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.

James, D., Van Doorn, J., and McLeod, S. (2001). Vowel production in mono-, di-and poly-syllabi words in children 3; 0 to 7; 11 years. In *Speech Pathology Australia National Conference: Evidence and Innovation*, pages 127–136. Speech Pathology Australia.

Jansen, S. (2010). High back vowel fronting in the north-west of england. *Proceedings of Sociophonetics, at the crossroads of speech variation, processing and communication*, pages 29–32.

Jarvis, S. and Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition.* Routledge.

Johnson, J. S. and Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(1):60–99.

Ju, M. and Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, 15(5):314–318.

Kajouj, F. and Kager, R. (2019). Effects of bilingualism on cue weighting: How do bilingual children perceive the dutch [a]-[a:] contrast? *International Journal of Bilingualism*, 23(2):509–524.

Kelleher, A. (2010). Who is a heritage language learner. *Heritage briefs*, pages 1–3.

Knightly, L. M., Jun, S.-A., Oh, J. S., and Au, T. K.-f. (2003). Production benefits of childhood overhearing. *The Journal of the Acoustical Society of America*, 114(1):465–474.

Koźbiał, S. (2015). Phonological error mapping: an English – Polish contrastive study. *Diffusion- The UCLan Journal of Undergraduate Research*, 4(1).

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In *Developmental neurocognition: Speech and face processing in the first year of life*, pages 259–274. Springer.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., and Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277(5326):684–686.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (nlm-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493):979–1000.

Kupisch, T. and Rothman, J. (2018). Terminology matters! why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, 22(5):564–582.

Lee, S. A. S. and Iverson, G. K. (2012). Stop consonant productions of korean–english bilingual children. *Bilingualism: Language and Cognition*, 15(2):275–287.

Lee, W.-S. and Zee, E. (2003). Standard chinese (beijing). *Journal of the International Phonetic Association*, 33(1):109–112.

Leeman, J. and Serafini, E. J. (2016). Sociolinguistics for heritage language educators and students. *Innovative strategies for heritage language teaching: A practical guide for the classroom*, pages 56–79.

Lenneberg, E. H. (1964). New directions in the study of language.

Lenneberg, E. H. (1967). *Biological foundations of language.* Wiley, New York,.

Lewis, S., Mehrabi, A., and de Leeuw, E. (2019). An articulatory-acoustic investigation into goose-fronting in german-english bilinguals residing in london, uk. *Proc. Interspeech 2019*, pages 3544–3548.

Li, X. (2017). The changing nature of chinese complementary schools in the uk. Available at: `http://eprints.nottingham.ac.uk/42429/`.

Mack, M. (1989). Consonant and vowel perception and production: Early english-french bilinguals and english monolinguals. *Perception & Psychophysics*, 46(2):187–200.

Magnuson, J. S. and Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human perception and performance*, 33(2):391.

Mayberry, R. I. and Kleunder, R. (2018). Rethinking the critical period for language: New insights into an old question from american sign language. *Bilingualism: Language and Cognition*, 21(5):886–905.

Mayo, C. and Turk, A. (2004). Adult–child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *The Journal of the Acoustical Society of America*, 115(6):3184–3194.

McCarthy, K. (2015). Speech perception and production in sequential bilingual children: A longitudinal study. *The Journal of the Acoustical Society of America*, 138(3):1894–1894.

Mccarthy, K. M., Evans, B. G., and Mahon, M. (2013). Acquiring a second language in an immigrant community: The production of Sylheti and English stops and vowels by London-Bengali speakers. *Journal of Phonetics*, 41(5):344–358.

McCarthy, K. M., Mahon, M., Rosen, S., and Evans, B. G. (2014). Speech Perception and Production by Sequential Bilingual Children: A Longitudinal Study of Voice Onset Time Acquisition. *Child Development*, 85(5):1965–1980.

Messum, P. and Howard, I. S. (2015). Creating the cognitive form of phonological units: The speech sound correspondence problem in infancy could be solved by mirrored vocal interactions rather than by imitation. *Journal of Phonetics*, 53(C):125–140.

Miller, J. F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., and Francis, D. J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research & Practice*, 21(1):30–43.

Montrul, S. (2011). Multiple interfaces and incomplete acquisition. *Lingua*, 121(4):591–604.

Montrul, S. (2018). Heritage language development: Connecting the dots. *International Journal of Bilingualism*, 22(5):530–546.

Montrul, S. and Foote, R. (2014). Age of acquisition interactions in bilingual lexical access: A study of the weaker language of L2 learners and heritage speakers. *International Journal of Bilingualism*, 18(3):274–303.

Montrul, S. A. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor*, volume 39. John Benjamins Publishing.

Moradi, H. (2014). An investigation through different types of bilinguals and bilingualism. *International Journal of Humanities & Social Science Studies*, 1(2):147–154.

Munro, M. and Mann, V. (2002). Age of immersion as a predictor of foreign accent. *The Journal of the Acoustical Society of America*, 112(5):2388–2388.

Nagy, N. (2018). Linguistic attitudes and contact effects in Toronto's heritage languages: A variationist sociolinguistic investigation. *International Journal of Bilingualism*, 22(4):429–446.

Newman, M. and Wu, A. (2011). "do you sound asian when you speak english?" racial identification and voice in chinese and korean americans' english. *American Speech*, 86(2):152–178.

Nittrouer, S. (2005). Age-related differences in weighting and masking of two cues to word-final stop voicing in noise. *The Journal of the Acoustical Society of America*, 118(2):1072–88.

Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., and Yamada, T. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics*, 39(2):156–167.

Ohde, R. N. and German, S. R. (2011). Formant onsets and formant transitions as developmental cues to vowel perception. *The Journal of the Acoustical Society of America*, 130(3):1628–42.

Paradis, J. and Genesee, F. (1996). Syntactic Acquisition in Bilingual Children: Autonomous or Interdependent? *Studies in Second Language Acquisition*, 18(1):1–25.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior research methods*, 51(1):195–203.

Peyton, J., Carreira, M., Wang, S., and Wiley, T. (2008). Heritage language education in the united states. *Sustaining linguistic diversity: Endangered and minority languages and language varieties*, pages 173–86.

Piske, T., Mackay, I. R. A., and Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29(2):191–215.

Polinsky, M. (2018). Bilingual children and adult heritage speakers: The range of comparison. *International Journal of Bilingualism*, 22(5):547–563.

Polinsky, M. and Kagan, O. (2007). Heritage Languages: In the 'Wild' and in the Classroom. *Language and Linguistics Compass*, 1(5):368–395.

R Core Team (2018). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rao, R. and Ronquest, R. (2015). The heritage spanish phonetic/phonological system: Looking back and moving forward. *Studies in Hispanic and Lusophone Linguistics*, 8(2):403–414.

Raumolin-Brunberg, H. (2009). Lifespan changes in the language of three early modern gentlemen. *The language of daily life in England (1400–1800)*, pages 165–96.

Roach, P. (2004). British english: Received pronunciation. *Journal of the International Phonetic Association*, 34(2):239–245.

Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism*, 13(2):155–163.

Ruiz-Felter, R., Cooperson, S. J., Bedore, L. M., and Peña, E. D. (2016). Influence of current input–output and age of first exposure on phonological acquisition in early bilingual Spanish–English-speaking kindergarteners. *International Journal of Language & Communication Disorders*, 51(4):368–383.

Safford, K. and Drury, R. (2013). The 'problem'of bilingual children in educational settings: Policy and research in england. *Language and Education*, 27(1):70–81.

Sancier, M. L. and Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4):421–436.

Schertz, J., Kang, Y., and Han, S. (2019). Cross-language correspondences in the face of change: Phonetic independence versus convergence in two korean-mandarin bilingual communities. *International Journal of Bilingualism*, 23(1):157–199.

Schmitz, J., Díaz, B., Fernandez Rubio, K., and Sebastian-Galles, N. (2018). Exploring the relationship between speech perception and production across phonological processes, language familiarity, and sensory modalities. *Language, Cognition and Neuroscience*, 33(5):527–546.

Sebastián-Gallés, N. and Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: Is the distributional account all there is to it? *Developmental science*, 12(6):874–887.

Sebastian Galles, N., Echeverria, S., and Bosch, L. (2005). The Influence of Initial Exposure on Lexical Representation: Comparing Early and Simultaneous Bilinguals. *Journal of Memory and Language*, 52(2):240–255.

Shultz, A. A., Francis, A. L., and Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, 132(2):EL95–EL101.

Simonet, M. (2014). Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics*, 43(C):26–37.

Simonet, M. (2016). *The Phonetics and Phonology of Bilingualism.* Oxford University Press.

Smith, M. and Kellerman, E. (1986). Crosslinguistic influence in second language acquisition: An introduction. *Crosslinguistic Influence in Second Language Acquisition, New York: Pergamon.*

Smith, N. V. et al. (1973). *The acquisition of phonology: A case study.* Cambridge University Press.

Smolensky, P. and Prince, A. (1993). Optimality theory: Constraint interaction in generative grammar. *Optimality Theory in phonology*, 3.

So, C. and Best, C. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech*, 53:273–93.

Stangen, I., Kupisch, T., Ergün, A. L. P., and Zielke, M. (2015). Foreign accent in heritage speakers of turkish in germany. *Transfer effects in multilingual language development*, pages 87–108.

Stow, C. and Dodd, B. (2003). Providing an equitable service to bilingual children in the uk: a review. *International Journal of Language & Communication Disorders*, 38(4):351–377.

Strand, S. (2016). Educational outcomes among children with english as an additional language (eal).

Tamburelli, M., Sanoudaki, E., Jones, G., and Sowinska, M. (2015). Acceleration in the bilingual acquisition of phonological structure: Evidence from polish–english bilingual children. *Bilingualism: Language and Cognition*, 18(4):713–725.

The Office for National Statistics (2018). Population of the uk by country of birth and nationality: July 2017 to june 2018. Available at: `https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/bulletins/ukpopulationbycountryofbirthandnationality/july2017tojune2018?:uri=peoplepopulationandcommunity/populationandmigration/internationalmigration/bulletins/ukpopulationbycountryofbirthandnationality/july2017tojune2018`.

The Office for National Statistics (2019). Population of the uk by country of birth and nationality: July 2018 to june 2019. Available at: `https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/bulletins/ukpopulationbycountryofbirthandnationality/july2018tojune2019`.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., and Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, 33(3):263–290.

Turner, J., Netelenbos, N., Rosen, N., and Li, F. (2014). Stop consonant production by french-english bilingual children in southern alberta. *Canadian Acoustics*, 42(3).

Tyler, M. D., Best, C. T., Faber, A., and Levitt, A. G. (2014). Perceptual Assimilation and Discrimination of Non-Native Vowel Contrasts. *Phonetica*, 71(1):4–21.

UK Federation of Chinese Schools (2020). About us. Available at: `http://ukfcs.info/en/who_are_we.php`.

Valdés, G. (2005). Bilingualism, Heritage Language Learners, and SLA Research: Opportunities Lost or Seized? *Modern Language Journal*, 89(3):410–426.

Vihman, M. M. (1996). *Phonological development: the origins of language in the child.* Applied language studies. Blackwell, Cambridge, Mass.

Vihman, M. M. (2017). In search of a learning model. *British Journal of Psychology*, 108(1):40–42.

Watt, D. and Fabricius, A. (2002). Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the f1˜ f2 plane. *Leeds working papers in linguistics and phonetics*, 9(9):159–173.

Wei, L. (2013). Conceptual and methodological issues in bilingualism and multilingualism research. *The handbook of bilingualism and multilingualism*, 2:26–53.

Wells, J. C. and Wells, J. C. (1982). *Accents of English*, volume 1. Cambridge University Press.

Winn, M. (2017). Praat scripts. Available at: `http://www.mattwinn.com/praat.html`.

Yeni-Komshian, G. H., Flege, J. E., and Liu, S. (2000). Pronunciation proficiency in the first and second languages of korean–english bilinguals. *Bilingualism: Language and cognition*, 3(2):131–149.

Yu, B. (2012). Issues in bilingualism and heritage language maintenance: Perspectives of minority-language mothers of children with autism spectrum disorders. *American journal of speech-language pathology / American Speech-Language-Hearing Association*, 22.

Zhou, J., Mai, Z., and Yip, V. (2021). Bidirectional cross-linguistic influence in object realization in cantonese–english bilingual children. *Bilingualism: Language and Cognition*, 24(1):96–110.