Nasal coarticulation in Lombard speech

Justin J. H. Lo

Linguistics and English Language, Lancaster University, Lancaster LA1 4YL, United Kingdom

j.h.lo@lancaster.ac.uk

Highlights

- Nasal coarticulation in Southern British English in quiet and noisy speech is investigated.
- Coarticulatory vowel nasality falls amid a compressed range of nasalisation in Lombard speech.
- Use of nasal coarticulation is maintained in Lombard speech, more so in anticipatory contexts.
- Nasal coarticulation is not targeted for adaptation as part of the Lombard response.

Abstract

Speaking in noisy environments entails a multitude of adaptations to speech production. Such modifications are expected to reduce gestural overlap between neighbouring sounds in order to enhance their distinctiveness, yet evidence for reduced coarticulation has been ambiguous. Nasal coarticulation in particular presents an unusual case, as it has been suggested to increase instead in certain clear speech conditions. The current study presents an experiment aimed at investigating how use of nasal coarticulation varies in quiet and noisy speech conditions. Speakers of Southern British English were recorded using a nasometer in an interactive reading task and produced monosyllabic target words with vowels bound by combinations of stop and nasal consonants. Use of nasal coarticulation was compared by means of a normalised measure that takes into account the speaker- and vowel-specific range of nasalisation available in each condition. In two noisy conditions where the interlocutor was either visible or not, vowel nasality in coarticulatory contexts was found to decrease in a way that closely tracked the compressed range between oral and nasal baselines. Speakers thus maintained their use of nasal coarticulation in Lombard speech, especially in the anticipatory direction. These findings suggest that the spreading of the velum lowering gesture from nasal consonants to neighbouring vowels is not targeted for adaptation in Lombard speech. They further reaffirm that enhancing acoustic distinctiveness and maintaining coarticulation are joint, compatible goals in the production of hyperarticulated speech.

Keywords: Lombard effect; Coarticulation; Nasality; Nasometry; Speech production

1 Introduction

Speakers adapt their linguistic-phonetic behaviour in different communicative contexts to meet the communicative needs befitting the situation. On top of stylistic variation according to audience (Bell, 1984), topic (Love & Walker, 2013) and formality (Labov, 1972), speakers respond to challenges in the speaking and listening environment, making fine phonetic adjustments accordingly in their speech production. Speaking in a noisy environment triggers the Lombard effect (Lombard, 1911), causing modulations in speech articulation to counter the fall in signal-to-noise ratio. Adaptations in both spatial and temporal dimensions of articulatory gestures impact how they interact with neighbouring gestures and in turn implicate how coarticulation is expressed. The present paper investigates the use of coarticulation in Lombard speech, an aspect of speech production which has so far been little explored within research into the Lombard effect. In particular, the study presented here focuses on how nasal coarticulation varies in speech produced in quiet and noisy environments.

1.1 Lombard speech and clear speech

Noisy environments induce speakers to raise their voice and exert greater vocal effort. As part of the vocal response, speakers not only increase their overall intensity in speech, but also raise their fundamental frequency (f0) (e.g., Jessen et al., 2005; Junqua, 1993; Van Summers et al., 1988). These are accompanied by a corresponding expansion of the intensity range (Ibrahim et al., 2022) and f0 range (Garnier & Henrich, 2014; Marcoux & Ernestus, 2019). Changes in the frequency domain affect the whole spectrum, with a shift of acoustic energy to higher frequencies and an accordingly lower spectral tilt (e.g., Garnier & Henrich, 2014; Junqua, 1996). In the temporal dimension, vowels are articulated with longer duration, but there is no consistent lengthening of consonants, which in some cases may even be shortened (Bořil & Pollák, 2005a; Castellanos et al., 1996; Garnier & Henrich, 2014; Lu & Cooke, 2008). These effects are well-attested and cross-linguistically robust (e.g., Bořil & Pollák, 2005a; Ibrahim et al., 2022; Kleczkowski et al., 2017; Le & Tang, 2023), but researchers have also identified considerable variation in the way individual voices respond to noise (e.g., Garnier et al., 2018; Junqua, 1993; Junqua et al., 1999).

On a fundamental level, the Lombard effect involves an automatic reaction to the impact of noise on self-monitoring of auditory feedback (Garnier et al., 2010). A higher level of noise induces a stronger Lombard response (Lu & Cooke, 2008; Van Summers et al., 1988), but the magnitude of the effect is also to some extent mediated by the frequency profile of the noise, specifically the relative loudness within 2-4 kHz that forms part of the frequency region the ear is most sensitive to (Garnier & Henrich, 2014; Lu & Cooke, 2009; Stowe & Golob, 2013). But the Lombard effect is not a purely passive reaction. Rather, it is subject to the influence of communicative intent, as the characteristics of Lombard speech emerge more strongly in interactive settings than in non-interactive settings (Boontham et al., 2016; Garnier et al., 2010, 2018). Speakers also appear to respond to the availability of visual information and adopt subtly different strategies of adaptation when their interlocutor is visible or not (Garnier et al., 2018).

Communicative orientation, while secondary to increasing vocal effort in Lombard speech, takes up a more prominent role in clear speech, produced when there is a heightened demand for clarity. Speech styles collectively investigated under the umbrella term of "clear speech" typically involve situational or listener factors other than noise: They can involve actual or imagined listeners, and in interaction the speakers themselves do not necessarily suffer from the same communicative barriers that give rise to difficulties in their interlocutors' perception. Commonly, when no actual interaction is involved in the study design, clear speech is elicited by instructing participants to imagine directing their speech at hard-of-hearing listeners or second-language speakers of the language, while more casual speech originates from speech directed at friends or family.

Lombard speech and clear speech, although underpinned by distinct primary motivations and environmental factors (Godoy et al., 2014), share a number of similar properties (see Cooke et al., 2014 for a detailed review). Like Lombard speech, clear speech is characterised by increases in intensity, f0, f0 range and vowel duration (Ferguson & Kewley-Port, 2007; Han et al., 2021; Picheny et al., 1986; Smiljanić & Bradlow, 2005). Speaking rate tends to decrease in clear speech (Ferguson et al., 2024; Smiljanić & Bradlow, 2005), although rate can be manipulated independently of the need for intelligibility (Cohn & Zellou, 2023; Krause & Braida, 2002). Lombard speech and clear speech further converge in the use of hyperarticulation, whereby articulatory movements attain greater magnitudes and higher velocity relative to quiet speech (Garnier et al., 2018; Kim & Davis, 2014; Redford et al., 2014; Šimko et al., 2016; L. Y. W. Tang et al., 2015). In Lombard speech at least, more visible articulators, such as the jaw and the lips, tend to undergo more pronounced adaptations than less visible articulators, such as the tongue dorsum (Šimko et al., 2016). On the whole, hyperarticulation serves to enhance acoustic distinctiveness so that speakers realise more canonical, idealised versions of each sound. Such adjustments are commonly explained under the H&H Theory (Lindblom, 1990), which posits that speakers reduce articulatory simplification to increase signal informativity when they detect or infer a need on the listeners' part for more explicit information. Indeed, when it comes to perception, visual cues that accompany articulatory adaptations in hyperarticulation have been shown to boost intelligibility (e.g., Alexanderson & Beskow, 2014; Fitzpatrick et al., 2015; Gagné et al., 2002; Garnier, 2023).

The concurrence of hyperarticulation and increasing vocal effort can result in different acoustic forms in Lombard speech and clear speech due to their different primary aims. This is readily observed in changes to the vowel space. In Lombard speech, the first formant (F1) of vowels is consistently raised (e.g., Bond et al., 1989; Lu & Cooke, 2008; Van Summers et al., 1988) due to a wider jaw opening. Meanwhile, there is little evidence of a global shift of the second formant (F2). Some studies have reported a general upward movement of F2 (Bořil & Pollák, 2005b; Junqua, 1993), while others have produced weak evidence of an overall decrease (Alghamdi et al., 2018; Lu & Cooke, 2008; Pisoni et al., 1985). Yet another set of studies have found vowel- and speaker-specific effects in the direction of change for F2 (Bond et al., 1989; Garnier et al., 2006; Ibrahim et al., 2022; Kirchhuebel, 2010; Van Summers et al., 1988). The findings for higher formants, though less investigated, are similarly mixed (Davis et al., 2006; Godoy et al., 2013; Junqua, 1993). Overall, there is no clear effect on vowel space area, with studies variably reporting expansion (P. Tang et al., 2017), reduction (Gully et al., 2019; Le & Tang, 2023) or an absence of difference (Kim & Davis, 2014). Clear speech, on the other hand, is typically characterised by an expanded vowel space (e.g., Ferguson & Kewley-Port, 2002; Han et al., 2021; Smiljanić & Bradlow, 2005): F1 of high vowels can, in many cases, be maintained or decrease while F1 of low vowels increase, such that F1 range expands without a wholesale lowering of the vowel space; F2 range expands in a similar fashion, as it generally increases for front vowels and may be maintained or decrease for back vowels.

1.2 Coarticulation

The spatial and temporal adaptations in articulation are tightly entwined in the dynamics of Lombard speech and clear speech, as articulatory gestures of neighbouring sounds overlap and interact with one another. Such coarticulatory effects give rise to contextual variability of segments that can render them acoustically less distinctive from similar sounds in the phonological inventory, and be more detrimental to accurate perception when communicative barriers are present. As speakers are motivated to shift production from the more hypoarticulated to the more hyperarticulated end of the speech continuum, a consequence of the adaptations above predicted by the H&H framework is that individual segments should be more resistant to gestural overlap from nearby sounds. At the same time, contextual variability due to coarticulation is not simply noise in the signal but can be informative to perception. There is now abundant evidence that listeners pick up on cues from anticipatory coarticulation and use them in real time to aid processing of upcoming speech materials (e.g., Beddor et al., 2013; Beddor & Krakow, 1999; Salverda et al., 2014). Carryover coarticulation, while lacking a predictive role, is influential in resolving perceptual ambiguity posed by other cues (e.g., Beddor et al., 2002; Mann, 1980) and can also assist listeners in recovering preceding sounds whose signals are absent or compromised (e.g., Ostreicher & Sharf, 1976; West, 1999; cf. Howson et al., 2021). As such, altering the coarticulatory content in hyperarticulated speech may go towards counteracting the perceptual benefits brought about by other adaptations.

1.2.1 Coarticulation: Reduced or maintained?

Few studies have targeted coarticulatory behaviour in hyperarticulated speech. Nicolaidais (2012), the only one of these to investigate coarticulation in noise-induced Lombard speech, found limited support for decreased use of coarticulation in the case of intervocalic consonants in Greek. Speakers in this electropalatographic study tended to produce shorter consonants in noise, with less lingual contact in the palatal region but increased, more anterior contact in the alveolar region. This was interpreted to be hyperarticulation of the tongue tip, so as to compensate for the effect of a wider jaw opening on the tongue dorsum. Contextual variability between /iCi/ and /aCa/ was reduced when produced in noise, especially in the palatal region, suggesting that speakers were aiming for more canonical targets, presumably to maintain distinction between consonants. Among the four recorded speakers, there was much individual variation permeating most parts of the results, which limited the strength of the conclusions that could be drawn.

In terms of clear speech, evidence for reduced coarticulation came from Moon and Lindblom (1994). Front vowels in English embedded in /wVl/ frames were found to undergo significantly less F2 undershoot in clear speech than in citation form. In other words, vowels in clear speech more closely approximated their canonical F2 targets in spite of the low F2 of their surrounding contexts. The differences between clear speech and citation form could not be solely attributed to longer vowel duration in the former. Rather, the additional finding of higher velocity of F2 movement in clear speech suggests that the articulatory gestures follow different underlying organisations.

The same conclusion was likewise reached by Guo and Smiljanic (2023), who investigated local coarticulation between two neighbouring segments using whole-spectrum measures of spectral distance and overlap. Drawing from the LUCID corpus (Baker & Hazan, 2010), this study focused on speech collected from speakers of Southern British English in a collaborative find-the-difference task, where communicative barriers were imposed on their interlocutor, alongside read sentences directed at two imagined audiences. Compared to the quiet, no-barrier condition, coarticulation was reduced to varying degrees in interactive conditions where speakers had to cope with their interlocutor’s difficulty. Whereas clear read speech was comparable to these challenging interactions in terms of coarticulation, speakers neither increased nor decreased the extent of coarticulation vis-à-vis the quiet condition when reading as if to a friend, where there was no heightened demand for clarity.

In contrast, Matthies et al. (2001) did not find an effect of clear speech on anticipatory lip rounding and tongue retraction in [iC(CC)u] sequences. Despite clear indications of increased vocal effort, neither articulatory measures based on electromagnetic articulography nor acoustic measures provided evidence for decreased coarticulation in clear speech. This behaviour of maintaining coarticulation applies to both spatial and temporal dimensions: There was no reduction in how early lip rounding or tongue retraction began, nor was there any increase in the magnitude of gestural movement traversed by the relevant articulator.

Another study that found evidence of coarticulatory maintenance is Bradlow (2002), who investigated C-to-V coarticulation by American English monolinguals and Spanish–English bilinguals, with a particular focus on contextual variability in F2 dynamics between /bu/ and /du/. In clear speech, /u/ was hyperarticulated in both environments with a lower F2 target (i.e., more backing and/or rounding). Importantly, regardless of the language spoken or language background, the distance between /bu/ and /du/ was sustained across speech styles, even up to the end portion of the vowel.

As these studies illustrate, the issue of coarticulation in clear speech has been approached through different types of phenomena, and the evidence available thus far for reduction or maintenance remains mixed. The overall picture becomes even more ambiguous when nasal coarticulation, where sounds in the vicinity of nasal consonants become partially nasalised due to an overlap with the lowering of the velum, is taken into consideration.

1.2.2 The case of nasal coarticulation

In their investigation of anticipatory nasalisation in clear speech in American English, Scarborough and Zellou (2013) elicited read speech in multiple scenarios, two of which aimed at an imagined interlocutor and the other two not directed at specific listeners. Participants further completed a worksheet fill-in task in interaction with a real listener. Although other established parameters of clear speech patterned as expected across these conditions, nasal coarticulation presented mixed findings. Notably, whereas imagined styles of clear speech were produced with lower nasality than the baseline reading condition, speech directed at real interlocutors tended to show greater vowel nasality, despite similarly exhibiting longer vowel duration and expanded vowel space that are emblematic of hyperarticulated speech.

In a similar vein, Cohn and Zellou (2023) examined two distinctive cases of clear speech distinguished by speech rate and showed that speakers made divergent adjustments to coarticulatory nasality. In fast-clear speech imitative of auctioneer style, there was a shallower rise of acoustic nasality over the course of the pre-nasal vowel than in casual speech. Slow-clear speech directed at an imagined hard-of-hearing listener, on the other hand, was produced with overall greater nasality than casual speech.

Relatedly, Zellou et al. (2023) examined face-masked speech and found vowel-specific enhancement of the distinction between pre-nasal and pre-oral environments, achieved by increasing nasalisation in the former and decreasing nasalisation in the latter. Such modifications, which were interpreted as increased use of nasal coarticulation, were shown to be perceptually advantageous, as listeners who heard target words with the relevant vowels were able to make use of the extended cue to more accurately identify the upcoming coda in its absence.

Taken together, this set of studies would seem to suggest a preference for greater, rather than less, nasal coarticulation in pre-nasal vowels in certain styles of hyperarticulated speech. Nonetheless, the findings therein are not completely in harmony with one another. Both Scarborough and Zellou (2013) and Cohn and Zellou (2023), for example, elicited speech directed at an imagined hard-of-hearing listener but found opposite results. The latter attributed such a difference to the specific recording environment, such that the effect of style was effectively overridden by the level to which participants felt at ease in their home or a laboratory environment. The factors that govern the variation of nasal coarticulation in different communicative contexts are thus far from resolved.

1.2.3 Open questions

The broader question remains why nasal coarticulation appears to stand out from other types of coarticulatory phenomena in its possibility to increase in hyperarticulated speech. One possible reason is the role of directionality. Studies on nasal coarticulation, as well as Matthies et al. (2001), focus on anticipatory coarticulation, whereas both Nicolaidis (2012) and Guo and Smiljanic (2023) do not distinguish the direction of coarticulation in their investigations. As anticipatory coarticulation has a perceptually predictive role in a way that carryover coarticulation does not, it may be the case that coarticulatory adjustments in Lombard speech and clear speech depend on the direction of coarticulation. Investigating nasal coarticulation in the two directions separately may shed light on this possibility.

Another potential source of divergence is the style of speech elicited through a variety of tasks. Most research on coarticulation in clear speech encompasses a mix of read speech, directed at either real or imagined listeners, and relatively spontaneous speech in interactive scenarios. Scarborough and Zellou (2013) postulate a distinction between real and imagined audience, but there appears to be no consistency to the effect of such differentiation, either for nasal coarticulation (cf. Cohn & Zellou, 2023) or more generally (Guo & Smiljanic, 2023). Studies looking at (co)articulation in noise-induced Lombard speech, on the other hand, have traditionally made less use of interactive tasks, relying instead on read sentences or words with no designated addressee (e.g., Nicolaidis, 2012; Šimko et al., 2016). Given that Lombard speech, like clear speech, involves an element of listener orientation, it is possible that variation in task interactivity may contribute to different outcomes for coarticulatory behaviour.

A further possibility lies in the variety of English investigated in studies of nasal coarticulation. In American English, anticipatory vowel nasalisation is widely documented to be temporally extensive (Moll & Daniloff, 1971; Pouplier et al., 2024) and is argued to have undergone phonologisation (Solé, 1995). If that is the case, the pre-nasal vowel acts not as the target but the source of nasal coarticulation (Pouplier et al., 2024). Increasing vowel nasality in clear speech may be seen not simply as strengthening a coarticulatory cue to the upcoming nasal, but as enhancing information to the allophonic environment of the vowel itself. Indeed, Zellou and Scarborough (2019) pursue this line of reasoning to explain the effect of neighbourhood density on the realisation of /æ/ in Western US English: Nasalisation in pre-nasal /æ/ is exaggerated in words with more similar-sounding competitors, thus serving to enhance its contrast with its pre-oral counterpart alongside other spectral and temporal cues.

In summary, it remains unclear how speakers alter their use of coarticulation, especially nasal coarticulation, in Lombard speech and clear speech. Though already of much interest in clear speech, nasal coarticulation is yet to be explored in Lombard speech. There is also no indication as to whether the potential for enhanced nasal coarticulation is restricted to the anticipatory direction or extends to the carryover direction. By exploring these issues, this study seeks to contribute to our understanding of the within-speaker variability of coarticulation and its role in speech production in noise.

1.3 Current study

The study presented here investigates the effects of Lombard speech on the production of nasal coarticulation by speakers of Southern British English. It forms part of a larger project that is aimed at exploring the within-speaker variability of coarticulation in different forensically relevant conditions and, consequently, the potential for coarticulatory phenomena to carry speaker-specific information. To this end, speakers engaged in a communicative task with a real interlocutor in a quiet environment and two noisy conditions including both face-to-face and simulated phone conversations. The experiment is thus designed to control for audience and task types, focusing on the impact of noise in an interactive setting where the interlocutor is visible or not. The choice of British English, in which nasal coarticulation is generally accepted as not phonologised (see also Cunha et al., 2024), is intended to sidestep the confounding role that a phonologised nasal target on the vowel may play in American English. Changes in vowel nasality in British English may thus be interpreted not as influencing an underlying target but as coarticulatory effects proper. While the analysis here focuses on comparing each noisy condition with the quiet condition, rather than between the noisy conditions, speakers may be expected to produce a greater degree of speech modifications in the simulated phone condition where their interlocutor is not in view (Fitzpatrick et al., 2015), although the impact of interlocutor visibility on the magnitude of adaptations can be highly idiosyncratic (Garnier et al., 2018).

In addition to anticipatory nasal coarticulation, which has been the focus in previous research, the current study also looks at nasal coarticulation in the carryover direction. If speakers are motivated to enhance coarticulatory cues in Lombard speech for perceptual benefits, the use of nasal coarticulation should be impacted in different ways in the two directions. Specifically, speakers would be expected to make greater use of nasal coarticulation in the anticipatory, but not carryover, direction.

Another point of departure from previous studies is the method used to measure vowel nasality. Instead of A1-P0, the measure used in previous studies, or other relatively robust acoustic correlates of nasality, such as F1 bandwidth and spectral tilt (Styler, 2017), the current study uses nasometry, a quasi-articulatory technique that taps into velic behaviour by recording acoustic radiations from the nasal and oral cavities separately. This not only provides a more direct way of analysing vowel nasalisation than acoustic correlates, but also circumvents well-documented issues suffered by A1-P0, in particular when it comes to high vowels (Carignan, 2021). Using A1-P0, which tracks the difference in amplitude between the first oral formant and the first nasal pole (Chen, 1997), to quantify the degree of nasalisation is itself a complex exercise, as it exhibits substantial between-speaker variability in terms of values as well as ranges (Styler, 2017). Given that different frequency regions undergo varied amplitude modulations in Lombard speech (see, e.g., Garnier & Henrich, 2014), there is also a need to first develop an understanding of how Lombard speech impacts both values and ranges of A1-P0, before its utility for comparing the degree of nasalisation within an individual across loudness levels can be ascertained.

Finally, the current study distinguishes vowel nasality, as measured by nasalance per se, from use of nasal coarticulation in its analysis. Vowel nasality, whether measured by acoustic or nasometric means, is known to vary across vowel categories (e.g., Bell-Berti, 1976; Rochet & Rochet, 1999). There is also much variation between individuals in the nasality of their voices, due in part to morphological differences in the nasal cavity (Stevens, 1972). As such, changes in vowel nasality may not directly translate as changes in coarticulatory behaviour. To account for these differences, use of coarticulation in the current study was measured by normalising vowel nasality in coarticulatory contexts against nasality in maximally oral and maximally nasal contexts (see Section 2.4.3 for details). As it is not yet clear how nasality in different phonetic contexts may be impacted by noise, this study thus pursues twofold analyses of both vowel nasality and use of nasal coarticulation. The former examines how vowel nasality in different phonetic contexts is affected by Lombard speech over the course of the vowel, whereas the latter addresses the primary issue of whether speakers enhance, reduce or maintain nasal coarticulation in Lombard speech. Together, these analyses shed light on the relationship between speakers' vocal modifications and coarticulatory behaviour in their Lombard response.

2 Method

The data and code for the reported analyses, as well as the experimental materials, are publicly available through an Open Science Framework repository (https://osf.io/ynxsj/).

2.1 Speakers

Twelve speakers (11 male, 1 nonbinary)[1] took part in a speech experiment in soundproofed recording booths at University College London. All participants were speakers of Southern British English, aged 18–34, who were born and raised in the South of England. None of the participants reported to speak any other language fluently, or to have any hearing or speech disorder. Participants received payment at the UCL Psychology and Language Sciences standard rate in return for their time.

2.2 Task and materials

Participants completed an interactive reading task, in which the stimuli consisted of 176 target words and 24 further fillers. Target words included 16 monosyllabic words each from 11 vowel categories. Figure 1 shows the categories in a representative vowel space of this variety of British English. Given the range of vowels included in this study, patterns due to vowel height in acoustic properties (e.g., f0; Whalen & Levitt, 1995) and vowel nasality (e.g., Bell-Berti, 1976; Clumeck, 1976) are expected. As vowel-specific patterns are secondary to the aims of the current study, the effect of vowel categories will be taken into account in statistical analyses but will not be discussed in detail.

---

[1] Note that the nonbinary speaker was removed prior to data analysis (see Section 2.4 for details).

Target words were evenly distributed in four phonetic contexts (CVC, CVN, NVC, NVN), where the target vowel was preceded and followed by a stop or a nasal. Only phonologically voiced stops (/b/, /d/, /g/) were used in onset position to avoid aspiration. Words were included in duplicate if there were fewer than four words available for particular combinations of vowels and contexts. A list of words with fricatives in place of stops were randomly drawn up as fillers to make up 200 words in total. These fillers were included when determining the presence of the Lombard effect in each noisy condition (Section 2.4.1), but removed from consideration when examining nasal coarticulation (Sections 2.4.2 and 2.4.3) due to the potential for "spontaneous nasalisation" in such environments (see, e.g., Ohala & Busà, 1995).

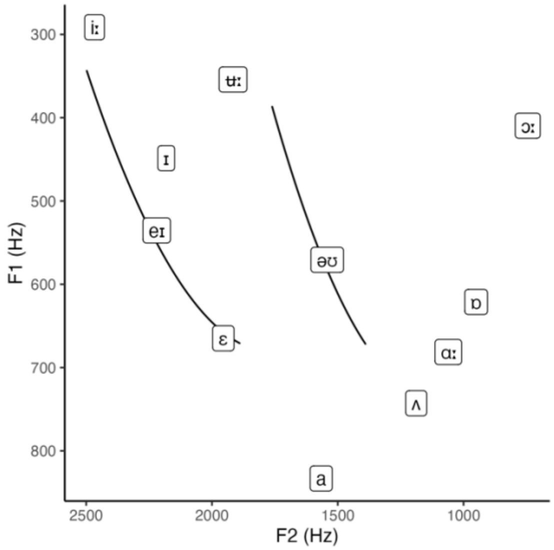

Figure 1. F1–F2 vowel space of speaker 02 in the quiet condition showing all included vowel categories. Labels are positioned at mean values at vowel midpoint. Lines indicate mean diphthong trajectories spanning 20–80% of vowel duration. Formant estimates were extracted using FastTrack (Barreda, 2021).

The stimuli were randomised into groups of four and presented in the format of *Word*$_1$*Word*$_2$*Word*$_3$*Word*$_4$, described to the participants as "codes". Participants were instructed to describe each item (i.e., code) to the experimenter so that the experimenter could write down the codes. They were asked to give their descriptions in sentence form with the target in phrase-final position (e.g., "the first word is *Word*$_1$"), but were explicitly given flexibility as to the exact wording. The overall aim of moving away from presenting the stimuli as isolated words or fixed sentences was to create an interactive setting that elicited connected speech and required both communicative intent and clarity, while ensuring the occurrence of sufficient tokens in each relevant context. In furtherance of this aim, the experimenter provided feedback to the participant throughout, such as to acknowledge target words with backchannelling or to request time for writing before continuing.

2.3 Recording procedure

As part of the larger study, participants completed the task in two recording sessions that were spaced at least two weeks apart. In each session, participants completed the task in three experimental conditions in randomised order, with the stimuli randomised differently in each repetition. Each condition lasted around 12 minutes on average. Participants were given the opportunity to take a short break before the next condition. In the *quiet* condition, no noise was played to the participants, but they wore closed-back headphones[2] to hear the experimenter's speech. In the *Lombard* condition, white noise (lowpass filtered at 22.05 kHz to prevent aliasing) was continuously played to the participants at 70 dB SPL through the headphones until they completed the stimuli. In both of these conditions, the experimenter was seated in an adjacent recording booth and was visible to the participant through a connecting glass. The *phone* condition started with a synthesised ringtone, followed by continuous white noise that was bandpass filtered between 300 and 3400 Hz to simulate a telephonic channel. However, it should be noted that the experimenter's voice was not bandpass filtered, which may have mitigated the participants' percept of the channel. A blackout blind was lowered in this condition such that the experimenter could not be seen through the connecting glass.

Audio recordings were collected using a Glottal Enterprises handheld nasometer, which had two built-in microphones (sampling at 16-bit, 44.1 kHz) separated by an acoustic baffle plate in contact with the speaker's face and recorded nasal and oral output in separate channels (Figure 2).

2.4 Data processing

Tokens with a pronunciation that deviated from the designated phonemes were excluded from analysis, but successful remedial renditions were included. Those with clipping during the target vowel were likewise excluded. All data from the sole nonbinary speaker were excluded at this stage, as a significant portion of their target vowels in the Lombard (32%) and phone conditions (36.5%) were clipped. Among the remaining male speakers, 148 (1.1%) tokens were removed, leaving a total of 13,052 tokens (1,553 of which are fillers). The target vowel in each of the remaining tokens were manually segmented in Praat (Boersma & Weenink, 2022) using acoustic landmarks in the waveforms. Vowel onset was marked by the onset of periodicity after a plosive or fricative onset and by the transition from low-amplitude simple waveform to high-amplitude complex waveform after a nasal

---

[2] Closed-back headphones were necessary to prevent sound leakage being picked up by the microphones on the nasometer. The use of headphones can impact the self-feedback loop and potentially augment the Lombard effect, or induce some degree of increase in vocal effort even when no noise is played (Garnier et al., 2010). Its influence is controlled by keeping this factor constant in all three conditions.

onset. Vowel offset was likewise marked by the end of high-amplitude complex periodic waveform. Where creaky voice (infrequently) occurred in the final portion of the vowel, the end of modal voicing was treated as the vowel offset.

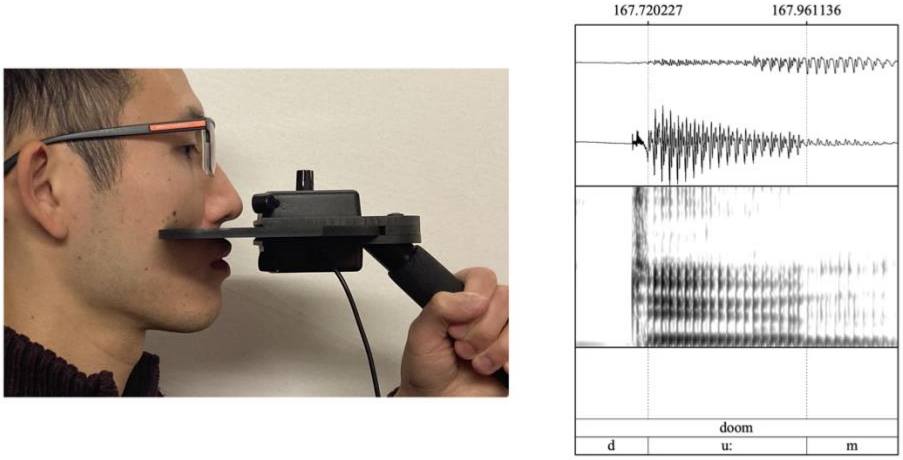

Figure 2. (Left) Right profile view of the nasometer, with an acoustic baffle plate in contact with the speaker's face between the nose and the upper lip. There is one microphone on the device above the plate and another below. (Right) Example recording of the word doom from the nasometer, showing waveforms (top: nasal output; bottom: oral output), spectrogram and segmental boundaries.

2.4.1 Validating Lombard effect

An initial analysis was conducted to verify that the Lombard and phone conditions elicited Lombard speech as designed. The intensity, f0 and duration of the vowel of all target and filler words were extracted from the audio recordings, converted to mono in Praat by averaging the two channels. As described in Section 1.1, these properties are known to be robust indicators of Lombard speech. Vowel duration and mean intensity across the whole vowel were extracted from the segmentation using custom Praat scripts. F0 measurements (in Hz) were extracted at 5 ms interval using the REAPER algorithm (Talkin, 2019) then converted to the psychoacoustic ERB-rate scale using the formula in Equation 1 (Glasberg & Moore, 1990), before they were averaged to obtain the mean f0 of each token.

$$ERB = \frac{\ln(1 + 0.00437f)}{(24.7)(0.00437)} \tag{1}$$

To assess the impact of different conditions, a separate linear mixed effects model was fitted to each of these properties using the *lmerTest* package (Kuznetsova et al., 2017) in R (R Core Team, 2022), which uses Satterthwaite's method for estimating degrees of freedom to obtain *p*-values (α = 0.05). In all models, the primary predictor of interest, condition, was included as a fixed effect, alongside vowel category, session and item number. Condition was treatment-coded with the quiet condition as the reference level.

Vowel category was sum-coded with TRAP /a/ assigned -1 as the reference level and others assigned +1 in their respective contrast. Session was sum-coded with the first and second sessions assigned +1 and -1 respectively. Item number was centred and scaled from [1, 50] to [-0.5, 0.5], so that its estimate represents any overall shift from the beginning to the end of each condition. Vowel duration, centred by the overall mean of each condition, was included as a fixed effect, except for the model for duration itself. Estimates for the effect of condition could then be interpreted against the mean values of the other factors. Each model included by-speaker random intercepts and condition-by-speaker random slopes.

2.4.2 Quantifying nasality dynamics

Vowel nasality was measured using nasalance, or proportional nasal amplitude. At a given timepoint, root-mean-square sound pressure was extracted from the nasal and oral channels in Praat using a 20-ms window. Nasalance was then calculated using the formula $A_N/(A_N + A_O)$, where A_N and A_O represent the RMS value extracted from the nasal and oral channel respectively. A value of 0 for nasalance corresponds to complete orality, while a value of 1 corresponds to a fully nasal sound. Nasalance was extracted from 11 equidistant points over the duration of the vowel to track its dynamics, but the values at each boundary were excluded due to abrupt transitional effects with neighbouring consonants, leaving a 9-point trajectory for subsequent analyses.

Comparison of nasalance dynamics across conditions was accomplished by means of linear mixed-effects modelling, where nasalance trajectories over normalised time were modelled using second-order orthogonal polynomials (cf. growth curve analysis; Mirman, 2014).The fitted model included fixed effects of condition, context and their interaction on the intercept, linear and quadratic terms. Of particular interest here is the effect of the interaction between context and condition on the polynomial terms, which would allow examination of whether experimental condition differentially affected nasalance dynamics depending on the phonetic context. Random effects for vowel and speaker were included in the model, but more complex random effects for speaker on the by-condition level resulted in singular fits that indicated overfitting and were thus not included. Equation (2) shows the model formula, where t_1 and t_2 correspond to the linear and quadratic terms respectively. Condition and context were both treatment-coded, with the quiet condition and CVC context set as reference levels. Model fitting was done using *lmerTest* to obtain parameter-specific *p* values.

$$\text{nasalance} \sim (t_1 + t_2) \times \text{context} \times \text{condition} + (t_1 + t_2|\text{speaker}) + (t_1 + t_2|\text{vowel}) \quad (2)$$

2.4.3 Quantifying nasal coarticulation

As vowel nasality is expected to vary across vowel categories, conditions and individual speakers, the degree of nasalisation in CVN and NVC contexts was quantified within each speaker by vowel and experimental condition, normalised using CVC and NVN as the oral and nasal baselines respectively. To this end, a coarticulatory quotient was derived for each CVN/NVC token, calculated as the ratio of (1) the area enclosed by the CVN/NVC nasalance trajectory and the corresponding mean CVC nasalance trajectory and (2) the area enclosed by the corresponding mean CVC and NVN trajectories. Figure 3 provides an illustration of this derivation, which conceptualises the difference between CVC and NVN as the full available space for coarticulation and use of coarticulation as the proportion of this space taken up by the speaker in his production. A coarticulatory quotient of 0 corresponds to a level of nasalance equal to the CVC context and indicates no use of coarticulation, while coarticulatory quotient of 1 corresponds to a level of nasalance equal to the NVN context and indicates a maximally nasalised vowel. Means were used for CVC and NVN instead of absolute minima/maxima to mitigate the impact of potential individual outliers. Since the oral and nasal baselines are empirically determined using per-speaker means, values of coarticulatory quotient below 0 or above 1 are permissible.

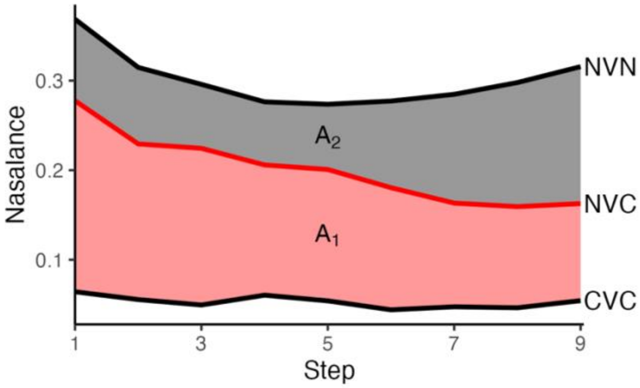

Figure 3. Example NVC nasalance trajectory (centre red line) with corresponding CVC (bottom) and NVN (top) mean trajectories. The coarticulatory quotient is derived as $A_1/(A_1+A_2)$.

This method of normalising nasal coarticulation is analogous to the variant of normalisation applied to A1-P0 in Zellou (2017), who also expressed nasal coarticulation in CVN contexts as a proportion of the speaker's acoustic nasality in CVC and NVN contexts. Note, however, that Zellou (2017) used a static measure of nasality extracted at vowel midpoint and calculated normalised nasal coarticulation on the speaker level rather than for each token. Similarly, the machine-learning approach by Carignan (2021) takes a bundle of acoustic features from CVC and NVN words, which function as the oral and nasal baselines, to empirically circumscribe a speaker's range of nasalisation.

To assess how the use of coarticulation varied in different experimental conditions, separate linear mixed effects models were fitted to the sets of coarticulatory quotients in the anticipatory (CVN) and carryover (NVC) directions, with condition, session, vowel category, item number and vowel duration included as fixed effects. Effects were coded in the same way as in Section 2.4.1. By-speaker random intercepts and condition-by-speaker random slopes were also included. As the carryover model resulted in a singular fit, the predictor with the smallest, statistically insignificant estimated effect was removed to yield a non-singular fit. This reduced model, with item number removed, will be reported instead.

3 Results

3.1 Global effects of Lombard speech

Figure 4 shows the overall distributions of intensity, duration and f0 from the target vowels (including fillers) in each condition. Each panel shows a clear rightward shift of the distribution in either noisy condition in comparison to the quiet condition. The statistical models fitted to each parameter, as shown in Table 1, provided evidence that vowels produced in the noisy conditions had significantly higher intensity, duration and f0 than those in the quiet condition. Target vowels in the Lombard condition were 3.6 dB higher in intensity, 14 ms longer in duration and 0.18 ERB (6.6 Hz) higher in f0. Model estimates indicate slightly greater effect sizes in the phone condition, where vowels were 4.3 dB higher in intensity, 15 ms longer in duration and 0.26 ERB (9.6 Hz) higher in f0 than those in the quiet condition. Overall, then, the analysis of these parameters provides evidence that the Lombard effect was robustly replicated in both Lombard and phone conditions.

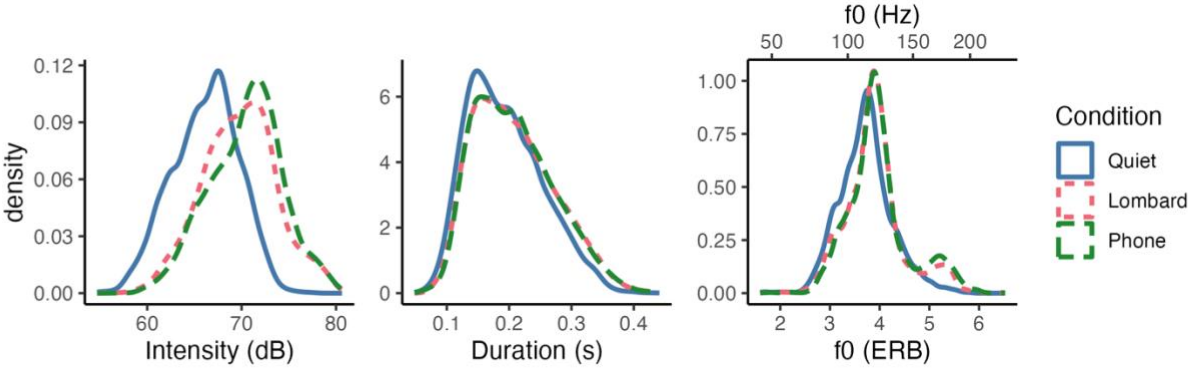

Figure 4. Density distributions of intensity (in dB) (left), duration (in seconds) and f0 (in ERB and Hz) pooled from all speakers and sessions in quiet (blue, solid), Lombard (pink, dashed) and phone (green, long-dashed) conditions.

The fitted models also found significant effects of the other control variables included, as shown in the model summaries (full summaries including vowel effects can be found in

Appendix A). Effects of vowel categories follow well-established patterns of intrinsic properties in relation to vowel height, where more open vowels are associated with higher intensity (e.g., Lehiste, 1976; Ordin, 2011) and lower f0 (Whalen & Levitt, 1995). Unsurprisingly, long monophthongs (BATH, FLEECE, GOOSE, THOUGHT) and diphthongs (FACE, GOAT) in British English were produced with longer duration than short monophthongs. Target vowels in the second session were significantly louder by 0.87 dB, shorter by 4 ms and higher by 0.03 ERB (0.7 Hz) than those in the first session. Within the same condition, target vowels that appeared in later items were significantly louder and higher, with an overall difference of 0.34 dB and 0.10 ERB (3.8 Hz) between the first and last items of each condition, but there was no significant effect of item number on vowel duration. Longer vowels were estimated to have significantly higher intensity and f0: Vowels longer than the per-condition mean by 100 ms were louder by 0.49 dB and higher by 0.02 ERB (0.6 Hz). Compared to the effects of condition, then, the effects of session and item number on the acoustic properties of the target vowels were relatively small. Likewise, the effect of vowel duration, within its range in the current data across all conditions (see Figure 4 middle panel), were considerably smaller than the effect of condition.

Table 1. Abridged summaries of linear mixed-effects models fitted to intensity (in dB), duration (in s) and f0 (in ERB). Reference level for condition is quiet; session is sum-coded; item number is centred and scaled; duration is centred by per-condition mean.

	Estimate	SE	*t*	*P*(>|*t*|)
Intensity				
Intercept	66.155	0.803	82.339	<.0001
Condition (Lombard)	3.631	0.433	8.394	<.0001
Condition (Phone)	4.295	0.482	8.909	<.0001
Session	-0.435	0.199	-21.845	<.0001
Item	0.340	0.068	5.020	<.0001
Duration	4.882	0.494	9.893	<.0001
Duration				
Intercept	0.194	0.007	26.866	<.0001
Condition (Lombard)	0.014	0.003	4.497	.0011
Condition (Phone)	0.015	0.002	8.029	<.0001
Session	0.002	0.000	5.173	<.0001
Item	-0.002	0.001	-1.946	.0516
f0				
Intercept	3.716	0.126	29.584	<.0001
Condition (Lombard)	0.178	0.056	3.180	.0098
Condition (Phone)	0.257	0.077	3.363	.0072
Session	-0.013	0.003	-3.983	<.0001
Item	0.103	0.011	9.294	<.0001
Duration	0.163	0.081	2.019	.0435

3.2 Nasalance dynamics

Turning to the comparison of nasalance dynamics across different conditions and contexts, Table 2 presents an abridged summary of the fitted mixed effects model, where nasalance trajectories were modelled with second-order orthogonal polynomials. Due to the large number of terms in total, Table 2 only includes terms returned as significant. The full model summary is available in Appendix B.

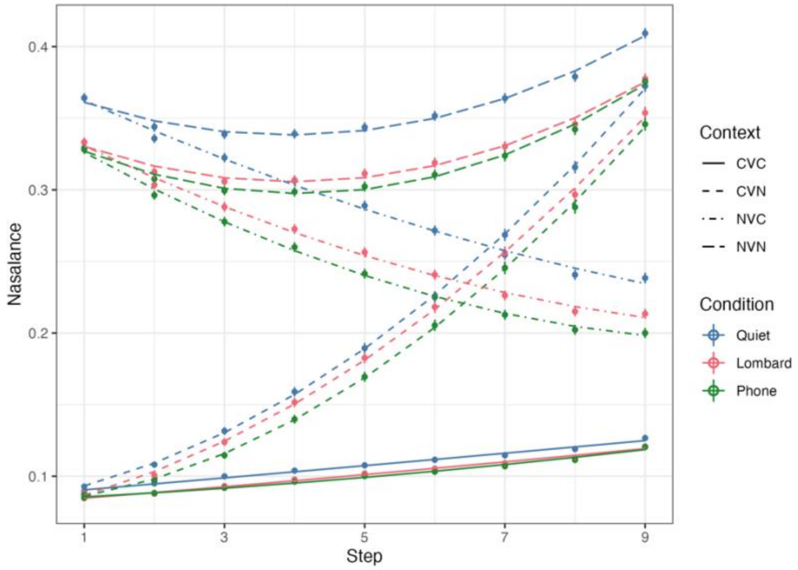

Figure 5. Predicted nasalance over 9 measurement points (with mean and standard error) in CVC (solid), CVN (dashed), NVC (dot-dash) and NVN (long-dashed) contexts in quiet (blue), Lombard (pink) and phone (green) conditions.

The overall effects of context and its interactions with both linear (t_1) and quadratic (t_2) terms in Table 2 provide evidence for expected effects of coarticulation on vowel nasality, also illustrated in Figure 5. In the CVC context, nasality remained generally low with little movement. A significant main effect of t_1 suggests that there was a slow linear rise in nasality over the course of the vowel. For CVN, nasality rose in a non-linear fashion (as confirmed by significant interactions with both t_1 and t_2) from the same level as CVC at the beginning of the vowel to a level almost reaching that of NVN by the end of the vowel. Nasality in NVC fell, also non-linearly, from the same level as NVN, but did not fall as sharply as nasality in CVN rose, such that nasality ended up higher than that in CVC even near the end of the vowel. Finally, for the nasal baseline NVN, nasality remained the highest of all contexts throughout the vowel but did not maintain a flat trajectory. Instead, as indicated by significant interactions with both t_1 and t_2, it dipped in the middle before rising again towards the coda.

Table 2. Abridged summary of linear mixed-effects model fitting nasalance trajectories with second-order orthogonal polynomials. t_1 and t_2 represent linear and quadratic terms. Reference level for context is CVC; reference level for condition is quiet (L: Lombard; P: phone).

	Estimate	SE	t	$P(>\|t\|)$
Intercept	0.108	0.019	5.613	<.0001
t_1	0.033	0.015	2.159	.0457
Context (CVN)	0.099	0.001	86.373	<.0001
Context (NVC)	0.184	0.001	160.601	<.0001
Context (NVN)	0.253	0.001	219.857	<.0001
Condition (L)	-0.006	0.001	-5.610	<.0001
Condition (P)	-0.007	0.001	-6.240	<.0001
t_1 × Context (CVN)	0.236	0.003	68.422	<.0001
t_1 × Context (NVC)	-0.157	0.003	-45.721	<.0001
t_1 × Context (NVN)	0.012	0.003	3.502	.0005
t_2 × Context (CVN)	0.047	0.003	13.598	<.0001
t_2 × Context (NVC)	0.013	0.003	3.710	.0002
t_2 × Context (NVN)	0.047	0.003	13.496	<.0001
Context (CVN) × Condition (L)	-0.004	0.002	-2.415	.0016
Context (NVC) × Condition (L)	-0.024	0.002	-14.714	<.0001
Context (NVN) × Condition (L)	-0.026	0.002	-15.814	<.0001
Context (CVN) × Condition (P)	-0.012	0.002	-7.197	<.0001
Context (NVC) × Condition (P)	-0.035	0.002	-21.562	<.0001
Context (NVN) × Condition (P)	-0.031	0.002	-19.063	<.0001
t_1 × Context (CVN) × Condition (L)	-0.013	0.005	-2.747	.0060
t_1 × Context (CVN) × Condition (P)	-0.018	0.005	-3.707	.0002

It is perhaps notable that nasalance remained at relatively low levels (< 0.4) even in NVN, the context most conducive to nasalisation, in comparison with typical values in similar environments reported in the literature (cf., e.g., Rochet & Rochet, 1999). The example speaker in Carignan (2021), for instance, had nasalance values ranging from 0.496 in /NaN/ to 0.530 in /NiN/. Past research, however, has focused on American English. When the target vowels in the current study are placed in the context of their neighbouring nasals, as is done in Figure 6 for the quiet condition, it can be seen that the nasal consonants maintained a high level of nasalance (> 0.7). The current data thus represent a sharp fall in nasality for vowels in NVN from their surrounding environment. Figure 6 also illustrates the range of variability between vowels, with nasalance in FLEECE maintained at above 0.5. The results here would thus appear to corroborate more qualitative comments in the literature that the extent of nasal coarticulation in British English is, on the whole, much less than that in American English (e.g., Bladon, 1979).

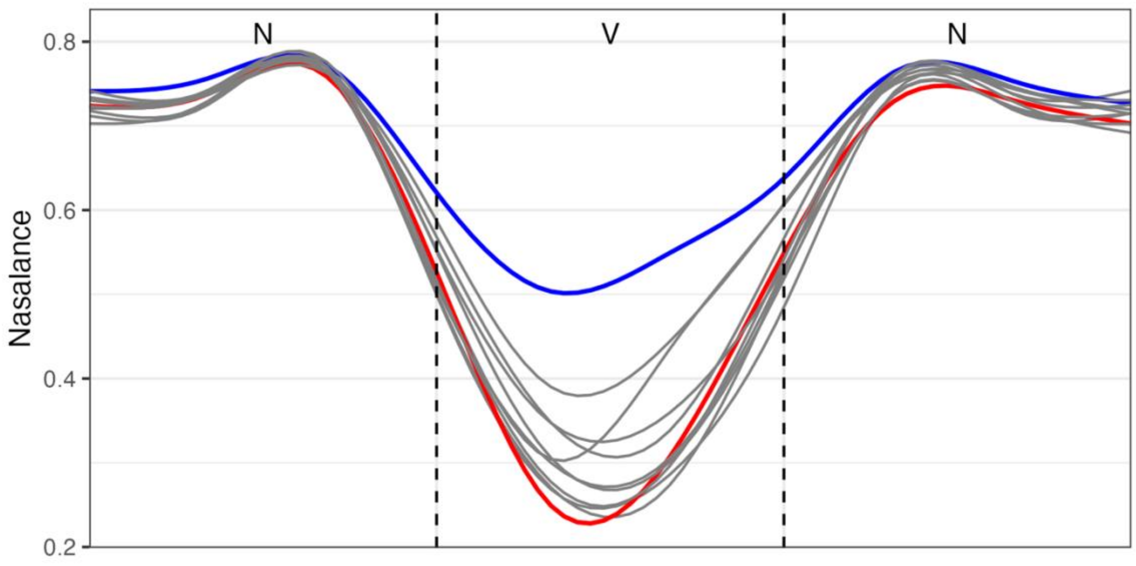

Figure 6. GAM-smoothed, time-normalised trajectories of nasalance in NVN contexts in the quiet condition by vowel category, with vowels reaching the highest (blue: FLEECE) and lowest (red: GOAT) values of nasalance highlighted.

As for condition, the principal factor of interest in this study, a main effect of condition indicates significantly lower nasality in the Lombard and phone conditions than in the quiet condition for the CVC context. Note, however, the small magnitude of the decrease: A fall from 0.108 in the quiet condition by 0.006 in the Lombard condition and 0.007 in the phone condition respectively. There was a significant interaction between context and condition, but no higher-order interactions with either linear or quadratic term, except with the linear term in the CVN context. In other words, in the noisy conditions, nasalance decreased to varying extents for each context but did so in a largely parallel fashion: Compared to the oral baseline (CVC), nasality decreased more in the other contexts, with much sharper differences in NVC and NVN. Meanwhile, the significant three-way interaction between the linear term, the CVN context and condition indicates that anticipatory nasality rose less sharply in the noisy conditions than in the quiet condition.

3.3 Coarticulatory quotient

Figure 7 shows the distributions of coarticulatory quotients in CVN and NVC contexts. For CVN, the coarticulatory quotients were clustered towards the 0-end in all three conditions, with the majority of tokens reporting a coarticulatory quotient below 0.5, indicating a small degree of anticipatory nasal coarticulation. For NVC, the distributions were concentrated between 0.5 and 1 in all three conditions, meaning that carryover nasal coarticulation was extensive. This normalised measure thus reliably mirrors the findings from the dynamic analysis in Section 3.2.

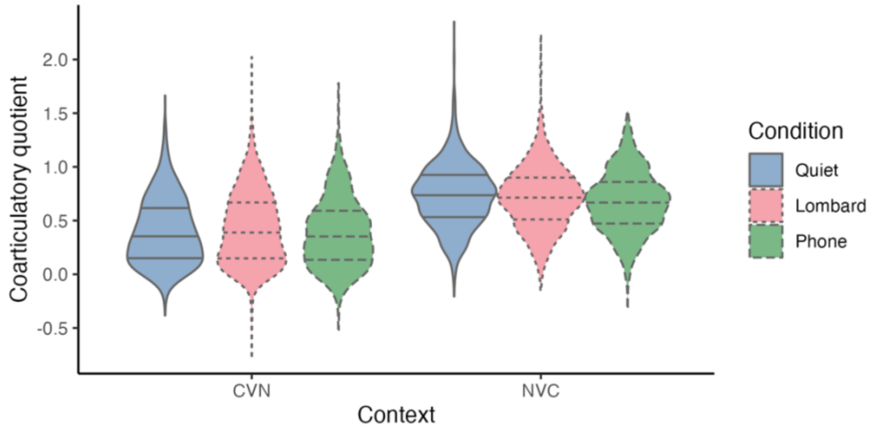

Figure 7. Violin plot of coarticulatory quotient in quiet (blue solid), Lombard (pink dashed) and phone (green long-dashed) conditions in CVN and NVC contexts. Horizontal lines indicate 25th, 50th and 75th percentiles.

Visual inspection of Figure 7 suggests no clear differences across all three conditions in the CVN context. As for NVC, coarticulatory quotient remains high in all conditions, although there appears to be a downward tendency from the quiet condition to the phone condition. The fitted models, summarised in Table 3, corroborated these observations. For CVN, neither Lombard nor phone condition was significantly different from the quiet condition. For NVC, the Lombard condition was not significantly different from the quiet condition, but the phone condition had significantly lower coarticulatory quotients than the quiet condition. The use of anticipatory nasal coarticulation was thus maintained regardless of noise, whereas the use of carryover nasal coarticulation was maintained in the Lombard condition but reduced in the phone condition.

Among the other factors included (vowel effects are available in Appendix C), Table 3 shows that duration, centred by the mean of each condition, had a significant effect on the coarticulatory quotient in both CVN and NVC contexts. As evidenced by the positive coefficient in the CVN model and the negative coefficient in the NVC model, longer vowels in each condition were associated with greater nasal coarticulation in the anticipatory direction and less nasal coarticulation in the carryover direction.

Table 3. Abridged summaries of linear mixed-effects models fitted to coarticulatory quotients in CVN and NVC contexts. Reference level for condition is quiet; session is sum-coded; item number is centred and scaled; duration is centred by per-condition mean. Item number was removed from the NVC model to attain non-singular fit.

	Estimate	SE	*t*	*P*(>|*t*|)
CVN				
Intercept	0.388	0.050	7.690	<.0001
Condition (Lombard)	0.029	0.038	0.770	.4594
Condition (Phone)	-0.008	0.035	-0.232	.8215
Session	-0.005	0.005	-1.057	.2907

Item	-0.012	0.017	-0.732	.4644
Duration	1.198	0.179	6.706	<.0001
NVC				
Intercept	0.724	0.029	24.902	<.0001
Condition (Lombard)	-0.018	0.011	-1.605	.1086
Condition (Phone)	-0.059	0.011	-5.159	<.0001
Session	0.003	0.005	0.554	.5795
Duration	-1.152	0.103	-11.232	<.0001

4 Discussion

The current study set out to examine the effects of Lombard speech on the use of nasal coarticulation among speakers of Southern British English. Lombard speech was elicited through exposing participants to either full-spectrum or band-filtered white noise while they interacted with a real interlocutor. In both cases, the effect was successfully triggered. The overall degree of Lombard effect was relatively small compared to typical studies that have generally used higher levels of noise, but comparable in magnitude to studies in which participants were exposed to the same noise level and type (Letowski et al., 1993; Ng, 2021). Although a direct comparison of the two noisy conditions is not the focus here, the combination of vowel intensity, duration and f0 would suggest that the phone condition elicited Lombard speech to a stronger degree than the face-to-face Lombard condition. Two factors may have contributed to the discrepancy in the speakers' response. The first is the modality of the interaction, as there is no visual information about the interlocutor in the phone condition. In this respect, the findings here are in agreement with previous research that also found speakers to produce speech with higher intensity and f0 when their interlocutor was obstructed from view (Fitzpatrick et al., 2015). The other factor is the profile of the noise played to the participant, as the acoustic energy of the noise, while equalised to 70 dB SPL in both conditions, is limited to between 300 and 3400 Hz in the phone condition. As such, there is a higher concentration of noise energy in the 2-4 kHz frequency region, the greater sensitivity of the ear to which has been shown to lead to a stronger Lombard response (Garnier & Henrich, 2014; Stowe & Golob, 2013).

4.1 Vowel nasality in Lombard speech

In this study, nasal coarticulation itself was investigated by means of nasometry and measured using nasalance. Vowel nasality was found to be generally lower in both Lombard and phone conditions, with a numerically greater decrease observed in the phone condition. Taken together with the differential Lombard response concluded above, these findings indicate that speakers increase the intensity of their oral output more than they do for the nasal output, leading to greater orality in Lombard speech. The results here are in line with those in Scarborough and Zellou (2013), where speakers hyperarticulate and reduce vowel nasality in speech directed at an imaginary hard-of-hearing listener. Replicating this finding in the current study, with a task using a real interlocutor, provides support for Guo and Smiljanic's (2023) conclusion that the presence or absence of a real interlocutor is immaterial to variation in coarticulatory behaviour.

The findings here would seem, at first glance, to contradict the condition with real interlocutors in Scarbourough and Zellou (2013), where speakers were found to hyperarticulate yet increase anticipatory vowel nasality in comparison to normal read speech. Here, instead, nasality in CVN rose less sharply in the Lombard and phone conditions and indeed exhibited an overall downward shift even greater than that in CVC. The different points of reference in the two studies, however, mean that the direction of change cannot be straightforwardly interpreted. The interactions with real listeners in Scarborough and Zellou (2013) did not, in fact, include any communicative barriers to overcome on the part of either participant or their interlocutor. No vowel lengthening was induced, and the degree of hyperarticulation was relatively small. If it indeed does not matter whether the addressee is real or imagined, then the real listener condition would actually be most comparable to the quiet condition in the current study, while the condition most comparable to Lombard speech in Scarborough and Zellou (2013) would arguably be the one directed at the imagined hard-of-hearing listener, which elicited speech with the most vowel lengthening and hyperarticulation. As all conditions in the current study are interactive, it cannot adjudicate on nasal coarticulation relative to read speech in Scarborough and Zellou (2013), where speakers were not provided with specific addressees. Once read speech is discounted from comparison, the two studies converge in the finding that more hyperarticulated listener-oriented speech is associated with lower vowel nasality, at least in the anticipatory direction.

This study additionally looked at vowel nasality in the NVC context in Lombard speech, which has not been specifically investigated in previous studies. In NVC, as well as NVN, contexts, nasality was lowered in Lombard and phone conditions, but this was done in parallel such that the trajectories retained the same shapes across conditions. The drop in nasality in these post-nasal contexts was also noticeably more substantial than in post-oral contexts. This difference is not unexpected: Nasality was much higher throughout the vowel in NVC and NVN than in CVC and the initial portions of CVN. CVC, being the oral baseline, reflects only inherent nasality of any given vowel category, and so nasality is already near floor level and could not have decreased to the same extent as NVN. There is, by contrast, much more room for nasality to move downwards in NVN and NVC. The result is an asymmetric shift in the oral and nasal baselines, thus effectively compressing the available range of nasalisation in Lombard speech.

4.2 Use of nasal coarticulation in Lombard speech

The use of nasal coarticulation in different conditions was compared by means of the coarticulatory quotient, devised to account for baseline differences between vowel categories and speakers. The analysis of coarticulatory quotient found no significant differences across the three conditions in the anticipatory direction. In the carryover direction, only the phone condition was found to have exhibited less nasal coarticulation than the quiet condition, whereas the Lombard condition was comparable to quiet speech. These results suggest that speakers did not vary their use of anticipatory nasal coarticulation in Lombard speech, and only minimally varied their use of carryover nasal coarticulation. Thus, despite the fact that speakers produced lower vowel nasality in the noisy conditions, the present study found no support for the outcome reached in Scarbrough and Zellou (2013) that clear speech results in less use of coarticulation.

While Scarbrough and Zellou (2013) and the current study focused on clear speech and Lombard speech respectively, the different conclusions are unlikely due to stylistic differences as elicited by the different tasks, given the clear alignment in the preceding acoustic analyses. The critical difference lies in how use of nasal coarticulation is measured in relation to vowel nasality itself. Recall that, in the present study, the coarticulatory quotient measures the ratio of the difference in nasalance between CVN/NVC and CVC contexts and the overall difference between CVC and NVN. In other words, the increased nasality in CVN relative to CVC is viewed in the light of the whole range of nasal coarticulation that is available to the speaker. In doing so, this analysis takes up the position set out by Styler (2017) that “even for within-speaker comparisons, researchers should attend to both the differences in baseline and range” (p. 2480) of measurements of nasality. Even though Styler’s argument is aimed at acoustic measurements like A1-P0, the range of vowel nasalisation available to the speaker as measured by nasalance *was* indeed found to be condition-specific and susceptible to compression in Lombard speech. The argument here, then, is that speakers decrease coarticulatory vowel nasality in Lombard speech in a way that closely corresponds to this range, effectively maintaining their coarticulatory behaviour, especially in the anticipatory direction.

While not directly reported before, the finding that the range of vowel nasality shifts based on task and style is apparent, albeit obliquely, from previous studies. One such instance is the data presented in Cohn and Zellou (2023), who centred their A1-P0 measurements using each speaker’s mean across CVC, CVN and NVN words but did not scale them. In addition to their reported effects of style on the overall degree and rate of change of nasality, their Figure 2 further suggests that fast-clear speech compresses the range of anticipatory nasality traversed by the vowel in CVN words, resulting in a shallower fall of (speaker-centred) A1-P0 than casual speech, while slow-clear speech has a potentially expanded A1-P0 range and accordingly steeper change in nasality. Granted, as vowel nasality in the CVC and NVN words were not reported, the specific relationship between the fall in A1-P0 in CVN words and the corresponding range between CVC and NVN words remains a speculative one.

4.3 Implications and limitations

The present results may be interpreted as suggesting that velum movement is not targeted for hyperarticulation or adaptation during pre- and post-nasal vowels in Lombard speech. Instead, the lowered vowel nasality that is found in all contexts is likely the by-product of other articulatory modifications: A wider jaw opening and possibly a more open lip aperture, both common in Lombard speech (Garnier et al., 2018; Šimko et al., 2016), would have the effect of lowering oral impedance and lead to a proportionally greater increase in oral intensity as overall intensity rises. Meanwhile, the preservation of the time course of nasality, as evidenced by the parallel movements of nasalance trajectories between quiet and noisy conditions, provides support for the notion that the time course of the velum gesture itself is not altered. The current evidence based on nasometry, however, remains indirect, as it relies on measuring acoustic output. Future research using techniques that can directly probe the movement of the velum, such as magnetic resonance imaging, may help elucidate its state during Lombard speech.

These findings are also consistent with the overarching principles of the H&H framework, wherein speakers would aim to conserve articulatory effort to the extent permitted by the communicative context to relay sufficient information to the listener. In Lombard speech, the speaker makes a host of effortful adjustments to mandibular, labial and laryngeal activities that boost intelligibility. Countering the lowering of vowel nasality brought on by these modifications would have required a greater degree of velum lowering to be made and maintained during vowel production. Yet, if the underlying aim is to signal speech content with more explicit information, the perceptual benefits of doing so are not entirely clear. Whereas coarticulatory cues may be deemed to be enhanced, the acoustic consequences of raising vowel nasality in coarticulatory contexts increase contextual variability of the vowel and risk endangering its distinctiveness. This ambivalence speaks to the contrary driving forces that speakers experiencing the Lombard effect are considered to face: (1) to reduce coarticulation so as to limit contextual variability and maximise segmental distinctiveness, versus (2) to enhance coarticulation so as to accentuate perceptual cues. With regard to nasal coarticulation, the balance between the two is sustained by reducing vowel nasality while preserving the shape of its progression over time, such that listeners are not disadvantaged by the removal of coarticulatory information.

The idea that compressing, or indeed expanding, the spatial dimension of nasal coarticulation is not to the detriment of the listener rests on the assumption that listeners are, or at least can be, attuned to stylistic modifications to the range of nasalisation available to the speaker. At present, direct evidence addressing this question is not readily available, although circumstantial evidence from two areas of work is in favour of such a possibility. First, listeners are adept at adapting their perception of coarticulatory information, including that of nasal coarticulation, to individual speakers (Coetzee et al., 2022; Zellou, 2022; Zellou et al., 2017; Zellou & Ferenc Segedin, 2019). Second, listeners are known to be influenced by range information when performing speaker normalisation, such as for the perception of lexical tones in Cantonese (Wong & Diehl, 2003), or the perceptual boundary between voiced and voiceless stops (e.g., Theodore & Monto, 2019; Zhang & Holt, 2018). Future research on the perception of nasal coarticulation that explicitly extends beyond coarticulatory timing to test the role of the spatial dimension will be able to shed light on how listeners interpret within- and between-speaker variation in ranges of nasal coarticulation.

An account of maintenance of nasal coarticulation aligns with the conclusion in Bradlow (2002) that maintaining coarticulation and enhancing distinctiveness are joint, compatible goals of clear speech production. In the context of C-to-V coarticulation in Bradlow (2002), the target F2 of /u/ is lowered, while F2 movement from the preceding /b/ and /d/ to the target is preserved. In the context of nasal coarticulation in British English, lowering overall vowel nasality contributes to the goal of enhancing acoustic distinctiveness, amidst other modifications for Lombard speech that are well known to achieve such an end (e.g., Garnier, 2023; Kim & Davis, 2014; Perkell et al., 2007), while the movement of nasality towards or from a nasal consonant is maintained. In both cases, then, the compatibility of the two goals derives from the fact that distinctiveness is primarily targeted in the spatial dimension, whereas maintenance of coarticulation relates more to the temporal dimension.

It must be noted that the current study has only tested a single, relatively low noise level in a limited set of scenarios, where the speaker responds to noise in interactive settings. These findings thus do not preclude the potential for use of nasal coarticulation to be modified in other styles of clear speech, where considerations other than noise (e.g., audience) are at play. Speakers were also only exposed to white noise, and so it remains to be determined how other types of noise with different spectral profiles may impact use of nasal coarticulation, particularly in cases where the concentration of spectral energy coincides with the region of low frequencies that is most susceptible to effects of vowel nasalisation (see, e.g., Chen, 1997). Another aspect that merits further investigation is the impact of the Lombard effect in different coarticulatory directions. In the current study, use of nasal coarticulation was arguably more strictly maintained in the anticipatory direction but had greater scope for variation in the carryover direction, where coarticulatory information is not so much predictive as redundant. Signal redundancy is especially useful when the communicative context is challenging, as it increases the chances of successful, accurate perception if other cues are compromised (Assmann & Summerfield, 2004). The difference here between anticipatory and carryover directions may be a reflection of the secondary role of carryover nasal coarticulation, suggesting that the balance between enhancing distinctiveness and maintaining (redundant) coarticulation may be tipped in conditions that warrant a stronger Lombard response. More research is thus also needed to test how use of coarticulation may vary as a function of the magnitude of the Lombard response.

5 Conclusion

The current study is one of the first to examine coarticulatory behaviour in Lombard speech. It also extends the investigation of within-speaker variation in nasal coarticulation from clear speech conditions to speech produced in noisy environments. In Lombard speech, speakers of Southern British English were found to produce vowels with overall lower nasality than in quiet speech. Vowel nasality in CVN and NVC contexts decreased in close correspondence to the compression in the range of nasalisation available in each condition, such that the speakers' use of nasal coarticulation was maintained rather than reduced. This is especially the case for anticipatory coarticulation. The findings here suggest that the Lombard effect may well not incur adaptations to the spreading of the velum lowering gesture, as other well-established modifications already add sufficient information to the signal while preserving temporal information for nasal coarticulation.

Acknowledgements

This work was supported by an International Association for Forensic Phonetics and Acoustics (IAFPA) Research Grant and a TRINITY Internship by the Department of Linguistics and English Language, Lancaster University. Thank you to Christopher Carignan for assistance with nasometry, Gordon Mills for technical assistance, Ellie Judson for assistance with segmentation, and all participants for their time and effort. Thank you also to two anonymous reviewers for their feedback.

References

Alexanderson, S., & Beskow, J. (2014). Animated Lombard speech: Motion capture, facial animation and visual intelligibility of speech produced in adverse conditions. *Computer Speech & Language*, *28*(2), 607–618. https://doi.org/10.1016/j.csl.2013.02.005

Alghamdi, N., Maddock, S., Marxer, R., Barker, J., & Brown, G. J. (2018). A corpus of audio-visual Lombard speech with frontal and profile views. *The Journal of the Acoustical Society of America*, *143*(6), EL523–EL529. https://doi.org/10.1121/1.5042758

Assmann, P., & Summerfield, Q. (2004). The perception of speech under adverse conditions. In *Speech processing in the auditory system* (Vol. 18, pp. 231–308). Springer-Verlag. https://doi.org/10.1007/0-387-21575-1_5

Baker, R., & Hazan, V. (2010). LUCID: A corpus of spontaneous and read clear speech in British English. *Proceedings of the DiSS-LPSS Joint Workshop (DiSS 2010)*, 3–6.

Barreda, S. (2021). Fast Track: Fast (nearly) automatic formant-tracking using Praat. *Linguistics Vanguard*, *7*(1), 20200051. https://doi.org/10.1515/lingvan-2020-0051

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, *30*(4), 591–627. https://doi.org/10.1006/jpho.2002.0177

Beddor, P. S., & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *The Journal of the Acoustical Society of America*, *106*(5), 2868–2887. https://doi.org/10.1121/1.428111

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, *133*(4), 2350–2366. https://doi.org/10.1121/1.4794366

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204. https://doi.org/10.1017/S004740450001037X

Bell-Berti, F. (1976). An electromyographic study of velopharyngeal function in speech. *Journal of Speech and Hearing Research*, *19*(2), 225–240. https://doi.org/10.1044/jshr.1902.225

Bladon, R. A. W. (1979). Some control components of a speech production model. In H. Hollien & P. Hollien (Eds.), *Current issues in linguistic theory* (Vol. 9, pp. 63–70). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.9.11bla

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (Version 6.2.23) [Computer software]. www.praat.org

Bond, Z. S., Moore, T. J., & Gable, B. (1989). Acoustic–phonetic characteristics of speech produced in noise and while wearing an oxygen mask. *The Journal of the Acoustical Society of America*, *85*(2), 907–912. https://doi.org/10.1121/1.397563

Boontham, C., Onsuwan, C., Saimai, T., & Tantibundhit, C. (2016). An analysis of Lombard effect on Thai lexical tones: The role of communicative aspect. In C. Carignan & M. D. Tyler (Eds.), *Proceedings of the Sixteenth Australasian International Conference on Speech Science and Technology* (pp. 149–152). ASSTA.

Bořil, H., & Pollák, P. (2005a). Comparison of three Czech speech databases from the standpoint of Lombard effect appearance. *Proceedings of Applied Spoken Language Interaction in Distributed Environments (ASIDE 2005)*, Paper 28.

Bořil, H., & Pollák, P. (2005b). Design and collection of Czech Lombard speech database. *Interspeech 2005*, 1577–1580. https://doi.org/10.21437/Interspeech.2005-461

Bradlow, A. R. (2002). Confluent talker- and listener-oriented forces in clear speech production. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 241–274). Mouton de Gruyter. https://doi.org/10.1515/9783110197105.1.241

Carignan, C. (2021). A practical method of estimating the time-varying degree of vowel nasalization from acoustic features. *The Journal of the Acoustical Society of America*, *149*(2), 911–922. https://doi.org/10.1121/10.0002925

Castellanos, A., Benedí, J.-M., & Casacuberta, F. (1996). An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *Speech Communication*, *20*(1–2), 23–35. https://doi.org/10.1016/S0167-6393(96)00042-8

Chen, M. Y. (1997). Acoustic correlates of English and French nasalized vowels. *The Journal of the Acoustical Society of America*, *102*(4), 2360–2370. https://doi.org/10.1121/1.419620

Clumeck, H. (1976). Patterns of soft palate movements in six languages. *Journal of Phonetics*, *4*(4), 337–351. https://doi.org/10.1016/S0095-4470(19)31260-4

Coetzee, A. W., Beddor, P. S., Styler, W., Tobin, S., Bekker, I., & Wissing, D. (2022). Producing and perceiving socially structured coarticulation: Coarticulatory nasalization in Afrikaans. *Laboratory Phonology*, *13*(1). https://doi.org/10.16995/labphon.6450

Cohn, M., & Zellou, G. (2023). Selective tuning of nasal coarticulation and hyperarticulation across slow-clear, casual, and fast-clear speech styles. *JASA Express Letters*, *3*(12), 125203. https://doi.org/10.1121/10.0023841

Cooke, M., King, S., Garnier, M., & Aubanel, V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech & Language*, *28*(2), 543–571. https://doi.org/10.1016/j.csl.2013.08.003

Cunha, C., Hoole, P., Voit, D., Frahm, J., & Harrington, J. (2024). The physiological basis of the phonologization of vowel nasalization: A real-time MRI analysis of American and Southern British English. *Journal of Phonetics*, *105*, 101329. https://doi.org/10.1016/j.wocn.2024.101329

Davis, C., Kim, J., Grauwinkel, K., & Mixdorff, H. (2006). Lombard speech: Auditory (A), Visual (V) and AV effects. *Speech Prosody 2006*, paper 252-0. https://doi.org/10.21437/SpeechProsody.2006-88

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *112*(1), 259–271. https://doi.org/10.1121/1.1482078

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1241–1255. https://doi.org/10.1044/1092-4388(2007/087)

Ferguson, S. H., Morgan, S. D., & Hunter, E. J. (2024). Within-talker and within-session stability of acoustic characteristics of conversational and clear speaking styles. *The Journal of the Acoustical Society of America*, *155*(1), 44–55. https://doi.org/10.1121/10.0024241

Fitzpatrick, M., Kim, J., & Davis, C. (2015). The effect of seeing the interlocutor on auditory and visual speech production in noise. *Speech Communication*, *74*, 37–51. https://doi.org/10.1016/j.specom.2015.08.001

Gagné, J.-P., Rochette, A.-J., & Charest, M. (2002). Auditory, visual and audiovisual clear speech. *Speech Communication*, *37*(3–4), 213–230. https://doi.org/10.1016/S0167-6393(01)00012-7

Garnier, M. (2023). Audio, Visual and Audiovisual intelligibility of vowels produced in noise. *INTERSPEECH 2023*, 4204–4208. https://doi.org/10.21437/Interspeech.2023-803

Garnier, M., Bailly, L., Dohen, M., Welby, P., & Loevenbruck, H. (2006). An acoustic and articulatory study of Lombard speech: Global effects on the utterance. *Interspeech 2006*, paper 1862-Thu1A3O.6-0. https://doi.org/10.21437/Interspeech.2006-323

Garnier, M., & Henrich, N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech & Language*, *28*(2), 580–597. https://doi.org/10.1016/j.csl.2013.07.005

Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research*, *53*(3), 588–608. https://doi.org/10.1044/1092-4388(2009/08-0138)

Garnier, M., Ménard, L., & Alexandre, B. (2018). Hyper-articulation in Lombard speech: An active communicative strategy to enhance visible speech cues? *The Journal of the Acoustical Society of America*, *144*(2), 1059–1074. https://doi.org/10.1121/1.5051321

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1–2), 103–138. https://doi.org/10.1016/0378-5955(90)90170-T

Godoy, E., Koutsogiannaki, M., & Stylianou, Y. (2013). Assessing the intelligibility impact of vowel space expansion via clear speech-inspired frequency warping. *Interspeech 2013*, 1169–1173. https://doi.org/10.21437/Interspeech.2013-13

Godoy, E., Koutsogiannaki, M., & Stylianou, Y. (2014). Approaching speech intelligibility enhancement with inspiration from Lombard and Clear speaking styles. *Computer Speech & Language*, *28*(2), 629–647. https://doi.org/10.1016/j.csl.2013.09.007

Gully, A. J., Foulkes, P., French, P., Harrison, P., & Hughes, V. (2019). The Lombard effect in MRI noise. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 800–804). Australasian Speech Science and Technology Association Inc.

Guo, Z., & Smiljanic, R. (2023). Speakers coarticulate less in response to both real and imagined communicative challenges: An acoustic analysis of the LUCID corpus. *Journal of Phonetics*, *97*, 101210. https://doi.org/10.1016/j.wocn.2022.101210

Han, H. J., Munson, B., & Schlauch, R. S. (2021). Fundamental frequency range and other acoustic factors that might contribute to the clear-speech benefit. *The Journal of the Acoustical Society of America*, *149*(3), 1685–1698. https://doi.org/10.1121/10.0003564

Howson, P. J., Kallay, J. E., & Redford, M. A. (2021). A psycholinguistic method for measuring coarticulation in child and adult speech. *Behavior Research Methods*, *53*(2), 846–863. https://doi.org/10.3758/s13428-020-01464-7

Ibrahim, O., Yuen, I., Van Os, M., Andreeva, B., & Möbius, B. (2022). The combined effects of contextual predictability and noise on the acoustic realisation of German syllables. *The Journal of the Acoustical Society of America*, *152*(2), 911–920. https://doi.org/10.1121/10.0013413

Jessen, M., Köster, O., & Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech, Language and the Law*, *12*(2), 174–213. https://doi.org/10.1558/sll.2005.12.2.174

Junqua, J.-C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America*, *93*(1), 510–524. https://doi.org/10.1121/1.405631

Junqua, J.-C. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication*, *20*(1–2), 13–22. https://doi.org/10.1016/S0167-6393(96)00041-6

Junqua, J.-C., Fincke, S., & Field, K. (1999). The Lombard effect: A reflex to better communicate with others in noise. *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, 2083–2086 vol.4. https://doi.org/10.1109/ICASSP.1999.758343

Kim, J., & Davis, C. (2014). Comparing the consistency and distinctiveness of speech produced in quiet and in noise. *Computer Speech & Language*, *28*(2), 598–606. https://doi.org/10.1016/j.csl.2013.02.002

Kirchhuebel, C. (2010). The effects of Lombard speech on vowel formant measurements. *The Journal of the Acoustical Society of America*, *128*(4_Supplement), 2395–2395. https://doi.org/10.1121/1.3508588

Kleczkowski, P., Zak, A., & Król-Nowak, A. (2017). Lombard effect in Polish speech and its comparison in English speech. *Archives of Acoustics*, *42*(4), 561–569. https://doi.org/10.1515/aoa-2017-0060

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *The Journal of the Acoustical Society of America*, *112*(5), 2165–2172. https://doi.org/10.1121/1.1509432

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). *lmerTest* Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Le, G., & Tang, Y. (2023). The Lombard effect on the vowel space of Northern Vietnamese. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 1167–1171). Guarant International.

Lehiste, I. (1976). Suprasegmental features of speech. In N. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 225–240). Academic Press.

Letowski, T., Frank, T., & Caravella, J. (1993). Acoustical properties of speech produced in noise presented through supra-aural earphones. *Ear and Hearing*, *14*(5), 332–338.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

Lombard, É. (1911). Le signe de l’élévation de la voix. *Annales Des Maladies de l’oreille, Du Larynx, Du Nez et Du Pharynx*, *37*(2), 101–119.

Love, J., & Walker, A. (2013). Football versus football: Effect of topic on /r/ realization in American and English sports fans. *Language and Speech*, *56*(4), 443–460. https://doi.org/10.1177/0023830912453132

Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America*, *124*(5), 3261–3275. https://doi.org/10.1121/1.2990705

Lu, Y., & Cooke, M. (2009). Speech production modifications produced in the presence of low-pass and high-pass filtered noise. *The Journal of the Acoustical Society of America*, *126*(3), 1495–1499. https://doi.org/10.1121/1.3179668

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, *28*(5), 407–412. https://doi.org/10.3758/BF03204884

Marcoux, K. P., & Ernestus, M. (2019). Differences between native and non-native Lombard speech in terms of pitch range. *Proceedings of the 23rd International Congress on Acoustics: Integrating 4th EAA Euroregio 2019: 9-13 September 2019 in Aachen, Germany*, 5713–5720. https://doi.org/10.18154/RWTH-CONV-239240

Matthies, M., Perrier, P., Perkell, J. S., & Zandipour, M. (2001). Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language, and Hearing Research*, *44*(2), 340–353. https://doi.org/10.1044/1092-4388(2001/028)

Mirman, D. (2014). *Growth curve analysis and visualization using R* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315373218

Moll, K. L., & Daniloff, R. G. (1971). Investigation of the timing of velar movements during speech. *The Journal of the Acoustical Society of America*, *50*(2B), 678–684. https://doi.org/10.1121/1.1912683

Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America*, *96*(1), 40–55. https://doi.org/10.1121/1.410492

Ng, M. L. (2021). The use of the Lombard effect in improving alaryngeal speech. *Journal of Voice*, *35*(1), 18–28. https://doi.org/10.1016/j.jvoice.2019.07.007

Nicolaidis, K. (2012). Consonant production in Greek Lombard speech: An electropalatographic study. *Italian Journal of Linguistics*, *24*(1), 65–101.

Ohala, J. J., & Busà, M. G. (1995). Nasal loss before voiceless fricatives: A perceptually-based sound change. *Rivista Di Linguistica*, *7*(1), 124–144.

Ordin, M. (2011). Palatalization and intrinsic prosodic vowel features in Russian. *Language and Speech*, *54*(4), 547–568. https://doi.org/10.1177/0023830911404962

Ostreicher, H. J., & Sharf, D. J. (1976). Effects of coarticulation on the identification of deleted consonant and vowel sounds. *Journal of Phonetics*, *4*(4), 285–301. https://doi.org/10.1016/S0095-4470(19)31256-2

Perkell, J. S., Denny, M., Lane, H., Guenther, F., Matthies, M. L., Tiede, M., Vick, J., Zandipour, M., & Burton, E. (2007). Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users. *The Journal of the Acoustical Society of America*, *121*(1), 505–518. https://doi.org/10.1121/1.2384848

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech,* *Language, and Hearing Research*, *29*(4), 434–446. https://doi.org/10.1044/jshr.2904.434

Pisoni, D., Bernacki, R., Nusbaum, H., & Yuchtman, M. (1985). Some acoustic-phonetic correlates of speech produced in noise. *ICASSP '85. IEEE International Conference on Acoustics, Speech, and Signal Processing*, *10*, 1581–1584. https://doi.org/10.1109/ICASSP.1985.1168217

Pouplier, M., Rodriquez, F., Lo, J. J. H., Alderton, R., Evans, B. G., Reinisch, E., & Carignan, C. (2024). Language-specific and individual variation in anticipatory nasal coarticulation: A comparative study of American English, French, and German. *Journal of Phonetics*, *107*, 101365. https://doi.org/10.1016/j.wocn.2024.101365

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Redford, M. A., Stine, J. N., & Vatikiotis-Bateson, E. (2014). A question of scope? Direct comparison of clear and in-focus speech production. In S. Fuchs, M. Grice, A. Hermes, L. Lancia, & D. Mücke (Eds.), *Proceedings of the 10th International Seminar on Speech Production (ISSP), 5 – 8 May 2014 Cologne, Germany* (pp. 352–355).

Rochet, A. P., & Rochet, B. L. (1999). Patterns of assimilation nasality in English as a function of vowel height. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 699–702). University of California, Berkeley.

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002

Scarborough, R., & Zellou, G. (2013). Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America*, *134*(5), 3793–3807. https://doi.org/10.1121/1.4824120

Šimko, J., Beňuš, Š., & Vainio, M. (2016). Hyperarticulation in Lombard speech: Global coordination of the jaw, lips and the tongue. *The Journal of the Acoustical Society of America*, *139*(1), 151–162. https://doi.org/10.1121/1.4939495

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, *118*(3), 1677–1688. https://doi.org/10.1121/1.2000788

Solé, M.-J. (1995). Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech*, *38*(1), 1–23. https://doi.org/10.1177/002383099503800101

Stevens, K. N. (1972). Sources of Inter- and Intra-Speaker Variability in the Acoustic Properties of Speech Sounds. In A. Rigault & R. Charbonneau (Eds.), *Proceedings of the seventh International Congress of Phonetic Sciences / Actes du Septième Congrès international des sciences phonétiques* (pp. 206–232). De Gruyter. https://doi.org/10.1515/9783110814750-014

Stowe, L. M., & Golob, E. J. (2013). Evidence that the Lombard effect is frequency-specific in humans. *The Journal of the Acoustical Society of America*, *134*(1), 640–647. https://doi.org/10.1121/1.4807645

Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America*, *142*(4), 2469–2482. https://doi.org/10.1121/1.5008854

Talkin, D. (2019). *REAPER: Robust Epoch And Pitch EstimatoR* [Computer software]. https://github.com/google/REAPER

Tang, L. Y. W., Hannah, B., Jongman, A., Sereno, J., Wang, Y., & Hamarneh, G. (2015). Examining visible articulatory features in clear and plain speech. *Speech Communication*, *75*, 1–13. https://doi.org/10.1016/j.specom.2015.09.008

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, *142*(2), 493–503. https://doi.org/10.1121/1.4995998

Theodore, R. M., & Monto, N. R. (2019). Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions. *Psychonomic Bulletin & Review*, *26*(3), 985–992. https://doi.org/10.3758/s13423-018-1551-5

Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, *84*(3), 917–928. https://doi.org/10.1121/1.396660

West, P. (1999). Perception of distributed coarticulatory properties of English /l/ and /r/. *Journal of Phonetics*, *27*(4), 405–426. https://doi.org/10.1006/jpho.1999.0102

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, *23*(3), 349–366. https://doi.org/10.1016/S0095-4470(95)80165-0

Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, *46*(2), 413–421. https://doi.org/10.1044/1092-4388(2003/034)

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, *61*, 13–29. https://doi.org/10.1016/j.wocn.2016.12.002

Zellou, G. (2022). *Coarticulation in phonology* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009082488

Zellou, G., Dahan, D., & Embick, D. (2017). Imitation of coarticulatory vowel nasality across words and time. *Language, Cognition and Neuroscience*, *32*(6), 776–791. https://doi.org/10.1080/23273798.2016.1275710

Zellou, G., & Ferenc Segedin, B. (2019). The distribution of coarticulatory variation influences perceptual adaptation. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 393–397). Australasian Speech Science and Technology Association Inc.

Zellou, G., Pycha, A., & Cohn, M. (2023). The perception of nasal coarticulatory variation in face-masked speech. *The Journal of the Acoustical Society of America*, *153*(2), 1084–1093. https://doi.org/10.1121/10.0017257

Zellou, G., & Scarborough, R. (2019). Neighborhood-conditioned phonetic enhancement of an allophonic vowel split. *The Journal of the Acoustical Society of America*, *145*(6), 3675–3685. https://doi.org/10.1121/1.5113582

Zhang, X., & Holt, L. L. (2018). Simultaneous tracking of coevolving distributional regularities in speech. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(11), 1760–1779. https://doi.org/10.1037/xhp0000569

Appendix A

Table A.1. Summary of linear-mixed effects model fitted to intensity (in dB). Reference level for condition is quiet; reference level for vowel (sum-coded) is TRAP; session is sum-coded; item number is centred and scaled; duration is centred by per-condition mean.

Intensity	Estimate	SE	*t*	$P(>\lvert t\rvert)$
Intercept	66.155	0.803	82.339	<.0001
Condition: Lombard	3.631	0.433	8.394	<.0001
Condition: Phone	4.295	0.482	8.909	<.0001
Vowel: BATH	0.162	0.072	2.238	.0253
Vowel: DRESS	0.152	0.066	2.306	.0211
Vowel: FACE	-1.015	0.065	-15.734	<.0001
Vowel: FLEECE	-1.222	0.063	-19.334	<.0001
Vowel: GOAT	0.234	0.065	3.616	.0003
Vowel: GOOSE	-0.437	0.064	-6.870	<.0001
Vowel: KIT	0.626	0.071	8.863	<.0001
Vowel: LOT	1.177	0.066	17.752	<.0001
Vowel: STRUT	1.113	0.067	16.515	<.0001
Vowel: THOUGHT	-0.300	0.067	-4.499	<.0001
Session	-0.435	0.199	-21.845	<.0001
Item	0.340	0.068	5.020	<.0001
Duration	4.882	0.494	9.893	<.0001

Table A.2. Summary of linear-mixed effects model fitted to vowel duration (in s). Reference level for condition is quiet; reference level for vowel (sum-coded) is TRAP; session is sum-coded; item number is centred and scaled.

Duration	Estimate	SE	*t*	$P(>\lvert t\rvert)$
Intercept	0.194	0.007	26.866	<.0001
Condition (Lombard)	0.014	0.003	4.497	.0011
Condition (Phone)	0.015	0.002	8.029	<.0001
Vowel: BATH	0.064	0.001	55.869	<.0001
Vowel: DRESS	-0.041	0.001	-37.264	<.0001
Vowel: FACE	0.032	0.001	28.486	<.0001
Vowel: FLEECE	0.009	0.001	7.612	<.0001
Vowel: GOAT	0.030	0.001	26.920	<.0001
Vowel: GOOSE	0.018	0.001	16.472	<.0001
Vowel: KIT	-0.067	0.001	-60.144	<.0001
Vowel: LOT	-0.039	0.001	-35.028	<.0001
Vowel: STRUT	-0.051	0.001	-46.143	<.0001
Vowel: THOUGHT	0.048	0.001	42.921	<.0001
Session	0.002	0.000	5.173	<.0001
Item	-0.002	0.001	-1.946	.0516

Table A.3. Summary of linear-mixed effects model fitted to mean f0 (in ERB). Reference level for condition is quiet; reference level for vowel (sum-coded) is TRAP; session is sum-coded; item number is centred and scaled; duration is centred by per-condition mean.

f0	Estimate	SE	*t*	*P*(>|*t*|)
Intercept	3.716	0.126	29.584	<.0001
Condition: Lombard	0.178	0.056	3.180	.0098
Condition: Phone	0.257	0.077	3.363	.0072
Vowel: BATH	-0.094	0.012	-7.923	<.0001
Vowel: DRESS	0.025	0.011	2.363	.0182
Vowel: FACE	-0.037	0.011	-3.559	.0004
Vowel: FLEECE	0.058	0.010	5.581	<.0001
Vowel: GOAT	-0.040	0.011	-3.786	.0002
Vowel: GOOSE	0.053	0.010	5.130	<.0001
Vowel: KIT	0.087	0.012	7.504	<.0001
Vowel: LOT	-0.008	0.011	-0.772	.4404
Vowel: STRUT	0.007	0.011	0.633	.5269
Vowel: THOUGHT	-0.024	0.011	-2.212	.0270
Session	-0.013	0.003	-3.983	<.0001
Item	0.103	0.011	9.294	<.0001
Duration	0.163	0.081	2.019	.0435

Appendix B

Table B.1. Summary of linear mixed-effects model fitting nasalance trajectories with second-order orthogonal polynomials. t_1 and t_2 represent linear and quadratic terms. Reference level for context is CVC; reference level for condition is quiet (L: Lombard; P: phone).

	Estimate	SE	t	$P(>\|t\|)$
(Intercept)	0.108	0.019	5.613	<.0001
t_1	0.033	0.015	2.159	.0457
t_2	0.000	0.010	0.033	.9744
Context (CVN)	0.099	0.001	86.373	<.0001
Context (NVC)	0.184	0.001	160.601	<.0001
Context (NVN)	0.253	0.001	219.857	<.0001
Condition (L)	-0.006	0.001	-5.610	<.0001
Condition (P)	-0.007	0.001	-6.240	<.0001
t_1 × Context (CVN)	0.236	0.003	68.422	<.0001
t_1 × Context (NVC)	-0.157	0.003	-45.721	<.0001
t_1 × Context (NVN)	0.012	0.003	3.502	.0005
t_2 × Context (CVN)	0.047	0.003	13.598	<.0001
t_2 × Context (NVC)	0.013	0.003	3.710	.0002
t_2 × Context (NVN)	0.047	0.003	13.496	<.0001
t_1 × Condition (L)	0.000	0.003	0.068	.9454
t_1 × Condition (P)	-0.001	0.003	-0.264	.7915
t_2 × Condition (L)	0.001	0.003	0.184	.8541
t_2 × Condition (P)	0.003	0.003	0.844	.3986
Context (CVN) × Condition (L)	-0.004	0.002	-2.415	.0016
Context (NVC) × Condition (L)	-0.024	0.002	-14.714	<.0001
Context (NVN) × Condition (L)	-0.026	0.002	-15.814	<.0001
Context (CVN) × Condition (P)	-0.012	0.002	-7.197	<.0001
Context (NVC) × Condition (P)	-0.035	0.002	-21.562	<.0001
Context (NVN) × Condition (P)	-0.031	0.002	-19.063	<.0001
t_1 × Context (CVN) × Condition (L)	-0.013	0.005	-2.747	.0060
t_1 × Context (NVC) × Condition (L)	0.007	0.005	1.521	.1282
t_1 × Context (NVN) × Condition (L)	-0.002	0.005	-0.347	.7285
t_1 × Context (CVN) × Condition (P)	-0.018	0.005	-3.707	.0002
t_1 × Context (NVC) × Condition (P)	0.001	0.005	0.183	.8549
t_1 × Context (NVN) × Condition (P)	0.001	0.005	0.265	.7912
t_2 × Context (CVN) × Condition (L)	-0.006	0.005	-1.320	.1870
t_2 × Context (NVC) × Condition (L)	0.004	0.005	0.914	.3606
t_2 × Context (NVN) × Condition (L)	0.001	0.005	0.174	.8618
t_2 × Context (CVN) × Condition (P)	0.001	0.005	0.134	.8936
t_2 × Context (NVC) × Condition (P)	0.008	0.005	1.671	.0948
t_2 × Context (NVN) × Condition (P)	0.005	0.005	1.022	.3066

Appendix C

Table C.1. Summary of linear-mixed effects model fitted to coarticulatory quotient in CVN context. Reference level for condition is quiet; reference level for vowel (sum-coded) is TRAP; session is sum-coded; item number is centred and scaled; duration is centred by per-condition mean.

CVN	Estimate	SE	*t*	*P*(>|*t*|)
Intercept	0.388	0.050	7.690	<.0001
Condition: Lombard	0.029	0.038	0.770	.4594
Condition: Phone	-0.008	0.035	-0.232	.8215
Vowel: BATH	0.099	0.019	5.185	<.0001
Vowel: DRESS	0.029	0.017	1.686	.0919
Vowel: FACE	0.070	0.017	4.130	<.0001
Vowel: FLEECE	-0.079	0.015	-5.138	<.0001
Vowel: GOAT	-0.090	0.016	-5.536	<.0001
Vowel: GOOSE	-0.137	0.015	-8.872	<.0001
Vowel: KIT	-0.064	0.020	-3.217	.0013
Vowel: LOT	0.018	0.017	1.083	.2788
Vowel: STRUT	0.050	0.018	2.795	.0052
Vowel: THOUGHT	-0.073	0.017	-4.366	<.0001
Session	-0.005	0.005	-1.057	.2907
Item	-0.012	0.017	-0.732	.4644
Duration	1.198	0.179	6.706	<.0001

Table C.2. Summary of linear-mixed effects model fitted to coarticulatory quotient in NVC context. Reference level for condition is quiet; reference level for vowel (sum-coded) is TRAP; session is sum-coded; duration is centred by per-condition mean. Item number was removed to attain non-singular fit.

NVC	Estimate	SE	*t*	*P*(>|*t*|)
Intercept	0.724	0.029	24.902	<.0001
Condition: Lombard	-0.018	0.011	-1.605	.1086
Condition: Phone	-0.059	0.011	-5.159	<.0001
Vowel: BATH	0.189	0.017	11.384	<.0001
Vowel: DRESS	0.003	0.015	0.180	.8575
Vowel: FACE	-0.047	0.015	-3.152	.0016
Vowel: FLEECE	-0.009	0.015	-0.587	.5575
Vowel: GOAT	-0.162	0.015	-10.717	<.0001
Vowel: GOOSE	-0.132	0.015	-8.857	<.0001
Vowel: KIT	-0.059	0.016	-3.681	.0002
Vowel: LOT	-0.068	0.015	-4.449	<.0001
Vowel: STRUT	0.022	0.016	1.432	.1523
Vowel: THOUGHT	0.073	0.016	4.487	<.0001
Session	0.003	0.005	0.554	.5795
Duration	-1.152	0.103	-11.232	<.0001