Laboratory Phonology
Journal of the Association for Laboratory Phonology

Havenhill, J. (2024). Audiovisual enhancement of vowel contrast: Production and perception of the COT-CAUGHT contrast in Chicago. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 15(1), pp. 1–77. DOI: https://doi.org/10.16995/labphon.11002

Open Library of Humanities

Audiovisual enhancement of vowel contrast: Production and perception of the COT-CAUGHT contrast in Chicago

Jonathan Havenhill, Department of Linguistics, the University of Hong Kong, HK, jhavenhill@hku.hk

This paper considers whether vowel systems are organized not only around principles of auditory-acoustic dispersion, but also around non-auditory perceptual factors, specifically vision. Three experiments examine variability in the production and perception of the COT-CAUGHT contrast among speakers from Chicago, where /ɑ/ (COT) and /ɔ/ (CAUGHT) have been influenced by the spread and reversal of the Northern Cities Shift. Dynamic acoustic and articulatory analysis shows that acoustic strength of the contrast is greatest for speakers with NCS-fronted COT, which is distinguished from CAUGHT by both tongue position and lip rounding. In hyperarticulated speech, and among younger speakers whose COT-CAUGHT contrast is acoustically weak due to retraction of COT, COT and CAUGHT tend to be distinguished through lip rounding alone. An audiovisual perception experiment demonstrates that visible lip gestures enhance perceptibility of the COT-CAUGHT contrast, such that visibly round variants of CAUGHT are perceptually more robust than unround variants. It is argued that articulatory strategies which are both auditorily and visually distinct may be preferred to those that are distinct in the auditory domain alone. Implications are considered for theories of hyperarticulation/clear speech, sound change, and the advancement of low back vowel merger in North American English.

1. Introduction

It has long been argued that phonological systems are organized around principles of acoustic and auditory dispersion. This organizing principle is reflected in typological generalizations, such as the observation that back vowels are typically round while front vowels are typically unround (Maddieson, 1984). Tongue backness and lip rounding have synergistic effects in lengthening the front cavity of the vocal tract and lowering F2, so this configuration maximizes the acoustic distance between front and back vowels (Stevens et al., 1986). Across languages, the most frequently occurring vowel system is the five vowel system comprising [i e a o u], not only because it adheres to the principle of maximal acoustic dispersion, but also because its moderate size provides for a relatively uncrowded vowel space (de Boer, 2001; Liljencrants & Lindblom, 1972; Lindblom, 1986). Auditory-acoustic dispersion also plays an important role in diachronic sound change. Its effects are observed in the phenomenon of chain shifts, for example, in which a change to the quality of one vowel triggers a cascading series of changes in which vowels are "pushed" in order to maintain sufficient acoustic distance from their neighbors or are "pulled" into undesirable acoustic gaps (Labov, 1994; Martinet, 1955). Auditory enhancement can further be found across a broad array of phonological contrasts beyond vowels. For instance, covariation of acoustic voicing correlates, including closure duration, F0, and vowel length, is argued to maximize the acoustic strength (and therefore auditory perceptibility) of laryngeal contrasts (Kingston & Diehl, 1994; Kingston et al., 2008).

The hypothesis that cross-linguistically common vowel inventories emerge due to a preference for articulatory dispersion is explicitly rejected by Diehl and Kluender (1989) in proposing the Auditory Enhancement Hypothesis. They correctly note that the vowel pairs /i/-/u/ and /y/-/ɯ/ are equally dispersed in articulatory terms, yet only /i/-/u/ also exhibits maximal acoustic dispersion. Because /y, ɯ/ occupy a narrower range of F2 values than the more peripheral vowels /i, u/, vowel systems like /y, ɯ, a/ would be acoustically (and thus auditorily) suboptimal. Despite making full use of articulatory height, backness, and rounding distinctions, such systems are unattested because they fail to maximize acoustic dispersion. A number of theoretical frameworks have been proposed to account for the cross-linguistic preference to maximize acoustic distance between contrastive sounds, including H&H Theory (Lindblom, 1990), Dispersion Theory (Flemming, 2004), General Auditory Theory (Lotto & Kluender, 1998), and Acoustic Enhancement Theory (Stevens et al., 1986). Such theories aim to predict sound systems on the basis of auditory-acoustic distance, while also incorporating competing factors like articulatory effort, maximizing the number of possible contrasts, feature economy, and other constraints.

An emphasis on auditory perceptibility is sensible, given that sound is arguably the primary (and often only) medium by which spoken language is transmitted. Nevertheless, there is a large body of evidence that speech perception is influenced by non-auditory perceptual

modes, among them proprioception, haptic or tactile feedback, and vision (e.g., Fowler & Dekle, 1991; Gick & Derrick, 2009; Mayer et al., 2013; Nasir & Ostry, 2006). Non-auditory cues can have considerable impact on the perception of speech sounds, even to the extent of overriding auditory cues. This is demonstrated by the well-known McGurk effect (McGurk & MacDonald, 1976), in which incongruous audiovisual stimuli undergo perceptual fusion (e.g., identification of auditory [ba] with visual [ga] as [da]). It is notable, however, that McGurk and MacDonald did not observe fusion for stimuli with auditory [ga] and visual [ba] or [pa]. In such cases, listeners either perceive both sounds, [bga], or a percept resembling the visual cue, [ba], affirming that listeners are attuned to visual labial cues when present. The integration of visual and other non-auditory cues has been a key source of evidence for gestural models of perception (Fowler, 1986; Liberman & Mattingly, 1985), given that they provide listeners with direct, unambiguous evidence of the speaker's articulation. Fowler (1996), in arguing for the direct realist theory of speech perception (Fowler, 1986; Gibson, 1979), proposes that "listeners perceive gestures, and some gestures are specified optically as well as acoustically" (p. 1733). Nevertheless, it has not widely been considered whether this influence is reflected in the organization of phonological systems.

1.1. Audiovisual perception in language variation and change

Under listener-oriented models of sound change, perceptibility is argued to be one of the key drivers of phonetic change and thereby phonological typology (Blevins, 2004, 2006; Ohala, 1981, 1993). Acoustic ambiguity arises not only due to environmental noise that hinders perception, but also as a result of coarticulation, interspeaker variation (both physiological and sociolinguistic), and other factors. Ohala (1993) argues that listeners are typically able to correct for variability in the acoustic signal when its sources are predictable. A wealth of experimental evidence supports this conclusion and shows that listeners shift their perceptual boundaries according to phonological context (Mann & Repp, 1980, 1981) or their interlocutor (Hay, Nolan, & Drager, 2006; Hay, Warren, & Drager, 2006; Johnson et al., 1999). If the underlying source of variation becomes no longer predictable, such as if a coarticulation-triggering segment is itself misperceived or lenited, misperception-based models of change posit that listeners will re-interpret coarticulatory variability to be an inherent quality of the affected segment. Listeners who have re-interpreted a segment's underlying form adopt production targets which differ from their predecessors, thereby initiating sound change.

This mechanism of change is clearly demonstrated by the development of nasal and nasalized vowels in languages like French and English, among many others (Beddor et al., 1986; Krakow et al., 1988). In English, coarticulatory vowel nasalization is pervasive and occurs due to anticipatory velar lowering before nasal codas. Nasalization of the preceding vowel alters its acoustic quality, particularly F1, which listeners may misattribute to the vowel's oral articulation, namely a

difference in tongue height. Krakow et al. show that American English listeners typically correct for this effect and accurately perceive the height of nasalized vowels, but only in contexts with nasal consonants (e.g., [bṼn] but not [bṼd]). In French, where vowel nasality is phonologically contrastive, the nasal vowel system exhibits a smaller number of height distinctions than the oral vowel system and the nasal vowels differ in height relative to the oral vowels from which they developed (Beddor et al., 1986).

In the case of nasal vowels, the articulators involved are presumably (mis)perceived auditorily, as configurations of the tongue, velum, pharynx, and larynx are mostly invisible. It is therefore unsurprising that nasal vowels exhibit a wide array of articulatory configurations, which emerge due to their acoustic complexity. When compared to their oral counterparts, nasal vowels may show differences not only in tongue height, but also in lip aperture, voice quality, and pharyngeal constriction (Carignan, 2014, 2017; Carignan et al., 2015; Chen, 2022; Chen & Havenhill, 2023; Garellek et al., 2016; Shosted et al., 2012). Some of these effects are argued to enhance the acoustic quality of nasalization (Garellek et al., 2016) or to preserve phonological contrasts (Carignan, 2018a; Chen, 2022), although interspeaker articulatory differences may also emerge due to misperception of the speaker's articulatory configuration (Carignan, 2018b). Covert interspeaker articulatory variability is also notably observed for English /ɹ/ and /s/, which may be produced using a range of lingual gestures that yield acoustically similar output (e.g., Bladon & Nolan, 1977; Dart, 1998; Delattre & Freeman, 1968; Espy-Wilson, 2004; Mielke et al., 2016). This type of variation is well predicted by auditory models of speech perception and theories of acoustic enhancement. If speech sounds are optimized for their auditory-acoustic quality, then a speaker's articulatory strategy is somewhat arbitrary as long as it yields the correct acoustic output (Johnson et al., 1993).

By contrast, configurations of the lips can be perceived not only auditorily, but also visually. Given this potential perceptual advantage, it is reasonable to hypothesize that visual perception of lip gestures may inhibit misperception-based change by disambiguating the speech signal. In turn, languages may be more likely to preserve contrasts that are auditorily and visually distinct, as opposed to those that can be distinguished through audition alone. For instance, in comparison with coronals and velars, labial sounds appear to be less likely to undergo changes to their place of articulation: debuccalization or full palatalization of labials are cross-linguistically rare changes (Bateman, 2011; Kochetov, 2011; O'Brien, 2012), and labial-velar segments show a tendency to change into labials rather than velars (Cahill, 1999; Ohala & Lorentz, 1977). However, despite a large body of evidence demonstrating the integration of vision and other non-auditory modalities in speech perception, only a handful of studies have directly considered this hypothesis (Havenhill & Do, 2018; Johnson, 2015; Johnson et al., 2007; King & Chitoran, 2022; McGuire & Babel, 2012). McGuire and Babel (2012) argue that asymmetries in the

misperception of English /θ/ and /f/ may be partly attributable to differences in their visual perceptibility. Whereas /f/ is consistently articulated with visible labiodental constriction, /θ/ may be produced with either a visible interdental constriction or a less-visible dental constriction. They find that /f/ and /θ/ are more reliably discriminated in audiovisual vs. audio-only or video-only conditions, while listener sensitivity in video-only conditions is greater for speakers who produce /θ/ interdentally than for those who produce /θ/ with a non-visible tongue gesture. King and Chitoran (2022) propose that the recent rise in the use of labiodental variants of English /r/ (i.e., [ʋ]) is facilitated by its visually distinct lip configuration. They show that [ʋ] and [w], which have distinct labial configurations (King & Ferragne, 2020), are distinguished equally well in audiovisual and visual-only perception, suggesting that visual cues alone are sufficient for preserving the /r/-/w/ contrast. As labiodental [ʋ] is acoustically similar to /w/ (more so than lingual [ɹ]), they argue that visual cues reinforce the /r/-/w/ contrast, in support of the hypothesis that phonological systems are optimized for audiovisual perceptibility.

1.2. Audiovisual factors in clear and hyperarticulated speech

When sounds have the potential to be produced with varying combinations of visible and non-visible articulatory gestures, audiovisual perception may also influence patterns of within-speaker variability. In clear speaking styles, for example, speakers make a number of modifications to their speech in an apparent attempt to improve intelligibility for the listener. Clear speech involves a number of global changes (applying to all segments), including increased intensity, slower speaking rate, and larger pitch ranges, among others (for reviews see Smiljanić & Bradlow, 2009; Uchanski, 2005). Segment-specific changes are also observed, e.g., higher-amplitude and higher-frequency spectral peaks for stop bursts and sibilants, longer VOT, and expansion of the F1 × F2 vowel space (Bradlow et al., 2003; Cho et al., 2011; Ferguson & Kewley-Port, 2002; Krause & Braida, 2004; Moon & Lindblom, 1994; Picheny et al., 1986). It is not fully resolved to what extent hyperarticulation targets specific phonological contrasts (Buz et al., 2016; Nelson & Wedel, 2017; Schertz, 2013; Tupper et al., 2021; Wedel et al., 2018) as opposed to being the byproduct of greater articulatory effort in general (Ohala, 1994; Wright, 2004). For instance, Bradlow (2002) and Smiljanić and Bradlow (2005) show that similar vowel space expansion occurs in languages with less crowded vowel spaces (Spanish, Croatian) as in those with more crowded vowel spaces (English). While the majority of research on clear speech has focused on acoustic properties, articulatory studies indicate that speakers may increase the velocity, magnitude, and duration of articulatory gestures (Matthies et al., 2001; Perkell et al., 2002). Such adjustments vary on an interspeaker basis, however—Perkell et al. show that some speakers rely on changes to duration or intensity rather than articulatory effort.

Clear speech is usually interpreted as listener-oriented enhancement, often under the framework of H&H Theory (Hyper- and hypo-articulation; Lindblom, 1990). H&H Theory proposes that speakers optimize their speech for perceptibility when they anticipate that the listener's perceptual needs demand it or when their message is contextually less predictable. Perceptual enhancement competes with speaker-oriented goals such as minimizing articulatory effort, so articulatory gestures may be hypoarticulated when not necessary to convey phonological contrasts. The perceptual advantages of audiovisually-transmitted speech are well established (e.g., Anderson et al., 1997; Gagné et al., 1994, 1995, 2002; Macleod & Summerfield, 1987; Sumby & Pollack, 1954), so some articulatory adjustments may occur, at least in part, to increase their visibility. Under especially noisy conditions, speakers might even optimize their speech purely for visual perceptibility, e.g., if they believe that the auditory/acoustic signal has no chance of recovery by the listener. Previous work suggests that speaking strategies are indeed modulated by vision. Anderson et al. found that speakers' utterances were auditorily less intelligible when they were aware that their interlocutor was able to see them. Ménard et al. (2016) tested the clear speech articulation of sighted and congenitally blind speakers. They found that while sighted speakers hyperarticulate lip gestures in clear speech, blind speakers were more likely to hyperarticulate lingual gestures. These results suggest that in order to enhance intelligibility, sighted speakers consider how their speech will be conveyed both optically and acoustically.

At the same time, hyperarticulation is not restricted to clear speech and is not necessarily driven by real-time communicative demands. An adjacent line of research has considered how phonetic variability is mediated by the lexicon (Baese-Berk & Goldrick, 2009; Fricke et al., 2016; Munson & Solomon, 2004; Wright, 2004). In 'contrastive hyperarticulation,' lexical items in dense phonological neighborhoods or that have minimal pair competitors, are produced with more extreme phonetic features along phonologically contrastive dimensions (Nelson & Wedel, 2017; Schertz, 2013; Wedel et al., 2018). For instance, Baese-Berk and Goldrick (2009) show that words with minimal pair neighbors (e.g., *cod* and *god*) exhibit a longer VOT compared to words lacking such a neighbor (e.g., *cop* vs. *[gɑp]). Both perception-oriented and production-oriented accounts for this phenomenon have been proposed. On the one hand, lexical competition may promote hyperarticulation for the purposes of perceptual optimization (Scarborough, 2010; Wright, 2004), as less frequent words and those in denser phonological neighborhoods are more difficult to identify (Luce & Pisoni, 1998). While production-based accounts do not fully dismiss this possibility (Lee & Baese-Berk, 2020), hyperarticulation has also been shown to occur even when potentially competitive lexical items are not relevant to the discourse (Baese-Berk & Goldrick, 2009; N. P. Fox et al., 2015; Wedel & Fatkullin, 2017; Wedel et al., 2018). The finding that hyperarticulation occurs when the speech signal is not likely to be misapprehended calls into question the extent to which

it is driven by the listener's perceptual requirements (or the speaker's estimation thereof). On this basis, Baese-Berk and Goldrick (2009) argue that lexically-conditioned variability is more fully explained by speaker-internal mechanisms of production. Specifically, activation of the production target induces increased (co-)activation of its phonological neighbors. Successful production requires stronger activation of the target than of its competitors, which results in hyperarticulation. Discourse status strengthens this effect—both the target and its competitors are more highly activated when one is contextually relevant, so hyperarticulation is more extreme.

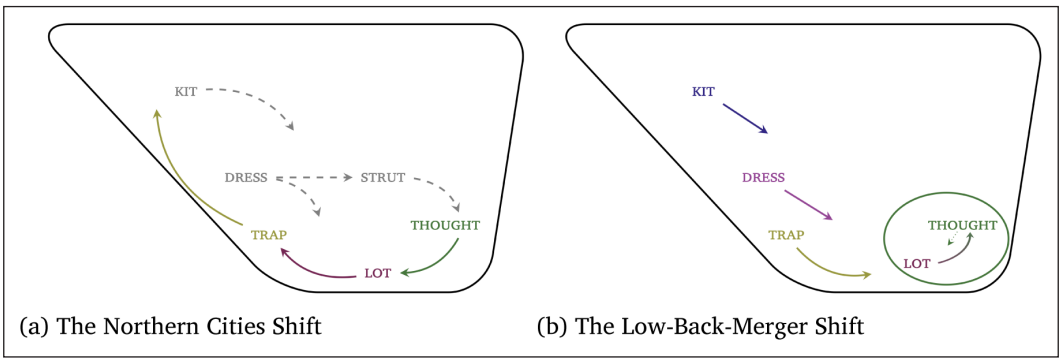
Most work in this area has focused on the acoustic speech signal, often with unidimensional variables like VOT or the Euclidean distance between vowels (e.g., Bradlow, 1996; Clopper & Tamati, 2014; N. P. Fox et al., 2015; Gahl et al., 2012). Some such measures can readily infer increases to articulatory effort; temporal expansion of VOT or coarticulation (Zellou & Chitoran, 2023) plainly correspond to gestures sustained for a longer period. Yet many other acoustic measures, including a vowel's position within the F1 × F2 space, reflect the combined actions of multiple gestures. To take nasal vowels again as an example, hyperarticulation may not only involve adjustments to the duration or magnitude of velar gestures, but also to tongue height and laryngeal setting, all of which influence the perceived quality of F1. For backness/rounding contrasts, conveyed by differences in F2, hyperarticulation may involve changes to the backness of the tongue, protrusion of the lips, and height of the larynx (Lindblom & Sundberg, 1971; Riordan, 1977). While acoustic distance from one vowel to its competitor is a useful proxy for auditory perceptibility, direct observation of speakers' articulatory strategies is necessary to obtain additional insight not only into how speakers maintain and enhance phonological contrasts, but also as to whether contrast is influenced by non-auditory perception.

1.3. This study: Audiovisual enhancement of the COT-CAUGHT contrast

This study examines variability in the strategies used to articulate (and hyperarticulate) vowel contrasts, considering the potential influence of differences in their visibility. Specifically, for vowels distinguished by both backness and rounding, are labial gestures less variable, less likely to be reduced, or more likely to be hyperarticulated than lingual gestures? Three experiments examine inter- and intraspeaker variability of the COT-CAUGHT contrast among speakers from Chicago, considering production in clear and normal speech as well as audiovisual perception. COT and CAUGHT (/ɑ, ɔ/, henceforth LOT and THOUGHT [Wells 1982]) are highly variable and have experienced a range of chain shifts and mergers in different regions throughout the past century. The phonological status of their contrast is marginal in most parts of North America, which makes their articulatory characteristics of theoretical interest, particularly in regions where the contrast has not (yet) collapsed.

Until recently, the Chicago variety of English has been characterized by the Northern Cities Shift (NCS), a vowel chain shift that developed during the late nineteenth and mid twentieth centuries in the Great Lakes region (the 'Inland North') of the United States (Labov et al., 2006). In its earliest stages, the NCS describes the coordinated movement of the vowels TRAP (/æ/), LOT, and THOUGHT, as shown in **Figure 1a**. Labov et al. (2006) propose that the shift began with the raising of TRAP, which can exhibit an F1 as low or lower than KIT (/i/, Labov 1994). Under their proposal, the raising of TRAP creates an opening in the vowel space which LOT moves forward to fill. They find that Northern Cities-shifted speakers typically exhibit a mean F2 for LOT of greater than 1450 Hz, far higher than Peterson and Barney's (1952) finding of a mean F2 for LOT of 1220 Hz for women and 1090 Hz for men. Thomas (2001) and McCarthy (2010) argue that the shift began with the fronting of LOT, for which McCarthy (2010) finds evidence in recordings of Chicago speakers born in the 1890s. In either scenario, the movement of TRAP and LOT is followed by the lowering and fronting of THOUGHT, which adopts the former position of LOT.

These changes have clear motivation under models of acoustic vowel dispersion: Movement of THOUGHT occurs in order to fill an open region of the vowel space caused by the fronting of LOT. However, because an increase in F2 can be the result of any gesture that shortens the front cavity of the vocal tract, acoustic dispersion alone cannot predict the articulatory strategies used to distinguish vowels along this dimension. Both lip rounding and tongue position influence F2 and have the potential to make equivalent changes to the acoustic output. Majors and Gordon (2008) investigated these alternatives through an analysis of video recordings of two speakers from St. Louis, where the NCS is in effect to some extent. They find that THOUGHT is fronted while retaining its rounding, suggesting that THOUGHT-fronting and lowering can be accomplished through a repositioning of the tongue alone. However,



(a) The Northern Cities Shift (b) The Low-Back-Merger Shift

Figure 1: Schematic diagrams of the Northern Cities and Low-Back-Merger Shifts. For NCS, solid line indicates early stages, dashed line indicates later stages. Adapted from Labov et al., 2006 (p. 190) and Becker, 2019.

their analysis did not involve direct observation of tongue position. Using lip video and ultrasound tongue imaging, Havenhill and Do (2018) find that three predicted patterns occur among Metro Detroit speakers. While some speakers produce fronted THOUGHT such that it is distinct from LOT in both tongue position and lip rounding, others contrast THOUGHT from LOT through either tongue position or lip rounding alone. These strategies are not equal with respect to their effects on the acoustic output, however. For speakers who distinguish THOUGHT from LOT by only one articulatory gesture, the vowels are acoustically less distinct than for speakers who contrast them by multiple gestures. Because both single-articulator strategies yield a similar degree of acoustic overlap between LOT and THOUGHT, Havenhill and Do (2018) also tested their visual perceptibility. They found that the absence of visual lip rounding cues for THOUGHT increases the likelihood that listeners will perceive an auditory THOUGHT stimulus as LOT. If visible rounding cues are important to the identification of LOT and THOUGHT, then it may follow that articulatory strategies which would eliminate lip rounding will be dispreferred on audiovisual grounds, even if lingual contrast alone is sufficient for differentiating the vowels in acoustics. If so, speakers may then be less likely to reduce the magnitude of labial gestures than to reduce the magnitude of lingual gestures (in normal or hypoarticulated speech), while they may be more likely to increase the degree of lip rounding in hyperarticulated speech.

It is not necessarily the case that speakers in traditionally Northern Cities-shifted regions like Chicago will maintain the LOT-THOUGHT contrast. Merger of the low back vowels is increasingly common throughout North America and is proposed to be the catalyst for the Low-Back-Merger Shift (LBMS; Becker, 2019; Labov et al., 2016; Roeder & Gardner, 2013). This supraregional chain shift, which has also been referred to as the Canadian/California Vowel Shift (Clarke et al., 1995; Eckert, 2008), Elsewhere Shift (Stanley, 2020), and Third Dialect Shift (Clarke et al., 1995; Labov, 1991), is illustrated in **Figure 1b**. The shift is proposed to begin with retraction of LOT and its merger with THOUGHT, which preserves THOUGHT as the surviving category (Becker, 2019; Gardner & Roeder, 2022; Labov, 2019). The loss of LOT from the acoustic low vowel space (Thomas, 2019) and/or short vowel system (Labov, 2019) motivates the subsequent retraction and lowering of TRAP, DRESS (/ɛ/), and KIT.

While the LBMS has become widespread throughout much of North America, it remains an open question to what extent it may take hold in regions where the pre-existing vowel systems are less conducive to LOT-THOUGHT merger (D'Onofrio & Benheim, 2020; Nesbitt & Stanford, 2021; Nesbitt et al., 2019). This is true not only for Chicago, but also the South and the Mid-Atlantic, which have likewise resisted the low back merger but have different phonetic implementations of the contrast. Labov (2019) notes that while the NCS distances LOT from THOUGHT via LOT-fronting, speakers in New York City and the South instead differentiate

THOUGHT from LOT by raising or diphthongizing it, respectively (Kendall & Fridland, 2017; Labov, 1966; Labov et al., 2006; Nycz, 2018; Stanley et al., 2021; Thomas, 2001). He proposes that North American English varieties can be taxonomically organized according to whether the vowel class LOT is round or unround, which relates in part to its prior merger with PALM in most places. Despite their many differences, the regions just noted are unified by an unround LOT. They stand apart from the rest of North America, where roundness of LOT encourages its merger into THOUGHT. The phonetic quality of the merged category has been described either as [ɑ]-like or [ɔ]-like, perhaps varying regionally (D'Onofrio et al., 2016; Hall-Lew, 2013; Kendall & Fridland, 2017; Kennedy & Grama, 2012), although its articulatory characteristics often go undescribed, being examined mostly through acoustics. Under Labov's proposal, the rounding distinction between LOT and THOUGHT (or lack thereof) is a key factor in their merger; the articulatory implementation of these vowels is therefore of interest, as it may inform where low back merger is more or less likely to occur.

This study examines the LOT-THOUGHT contrast through an acoustic, articulatory, and perceptual study of Chicago English. Three experiments examine inter- and intra-speaker phonetic variability in the articulation of LOT and THOUGHT, as well as the possible influence of visual cues on perceptibility of the contrast. Experiment 1 focuses on interspeaker variation; that is, whether speakers differ from one another in how they combine lingual and labial gestures to produce LOT and THOUGHT, and to what extent they vary in the strength of the acoustic contrast. Experiment 2 looks at intraspeaker variation, examining how speakers hyperarticulate LOT and THOUGHT in clear speech, whether hyperarticulation increases the acoustic distance between them, and whether lingual or labial gestures are preferentially hyperarticulated. Experiment 3 tests whether visual rounding cues aid in the identification of LOT and THOUGHT through an audiovisual perception experiment.

2. Experiment 1: Interspeaker variability in LOT-THOUGHT production in Chicago

2.1. Methods

Eighteen participants (4 men, 14 women) participated in the study, which was conducted at Northwestern University in Evanston, Illinois. Demographic information is presented in **Table 1**. Participants were natives of the Chicago area, having been born and raised in the region at least through age 18. The age range of participants was 20 to 77 years ($M = 46.7$, $SD = 21.2$). All participants had self-reported normal hearing and speech as well as normal or corrected-to-normal vision. One additional participant (a 19-year-old woman not listed in **Table 1**) also took part in the study but was excluded from analysis because she later reported that she had lived well outside the region for over five years during adolescence.

Speaker	Gender	Birth Year	Ethnicity	Ex-Chi	Areas Lived
CHI001	F	1994	White	0	South Side
CHI002	M	1962	White	1	NW Suburbs
CHI003	F	1963	White	0	Far N. Side, NW Side
CHI004	F	1947	White	0	Far N. Side, North Shore, NW Suburbs
CHI005	F	1941	White	7	West Side, the Loop, South Side
CHI006	F	1952	White	4	Far SE Side, Far SW Side
CHI008	M	1955	White	0	NW Side
CHI009	M	1998	White	0	South Side
CHI010	M	1948	White	0	NW Side
CHI011	F	1997	White	0	North Shore
CHI012	F	1981	Hispanic	6	Near West Side, Western Suburbs
CHI013	F	1995	White	0	Uptown, Far N. Side
CHI014	F	1992	Black	0	South Side
CHI015	F	1953	White	0	Far N. Side
CHI016	F	1961	White	4	South Side, Suburbs
CHI017	F	1955	White	2	North Side, Far N. Side, North Shore
CHI018	F	1991	White	2	Outer Suburbs
CHI019	F	1998	Asian	0	SW Side, South Side

Table 1: Demographic information for study participants. *Areas Lived* indicates the areas of Chicago where the participant has lived; *Ex-Chi* indicates the number of years after age 18 the participant has lived outside the Chicago metropolitan area.

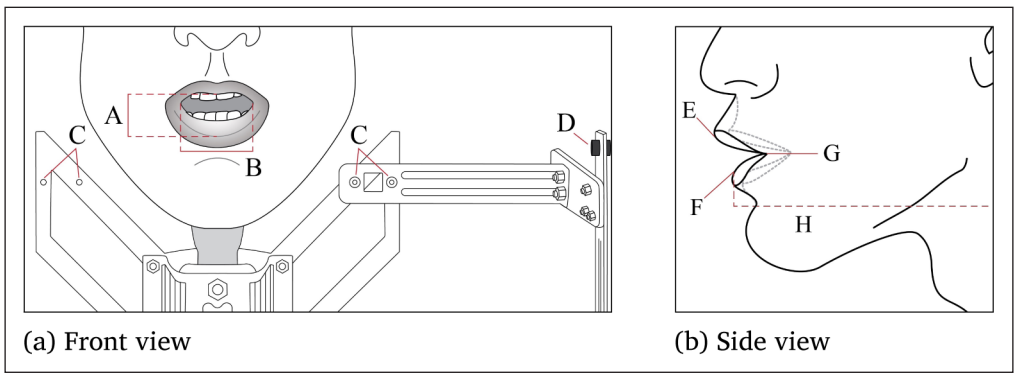
2.1.1. Materials

Prompts included 109 English words containing FLEECE (/i/), GOOSE (/u/), GOAT (/o/), TRAP, LOT, and THOUGHT, listed in supplementary materials. Each vowel appeared in 18 phonological contexts, including words with coronal, velar, and labial onsets and codas, as well as vowel-initial words. To the extent possible, onset and coda consonants were balanced across vowels for voicing and nasality, such that the words for each environment comprise a (near) minimal sextuplet. Words were mostly monosyllabic, although some disyllabic words (with primary stress on the target vowel) were included to fill lexical gaps. Speakers also produced words given as response choices in Experiment 3. These were used to verify membership in the expected vowel class but are not otherwise analyzed.

2.1.2. Procedure

Ultrasound data were captured using an Articulate Instruments ultrasound system with a 20 mm radius 2–4 MHz transducer. Participants were seated with the transducer held in place by a stabilizing headset (Articulate Instruments Ltd., 2008). Side-view lip video was captured by an analog NTSC camera with a 4.75 × 3.55 mm sensor and 0.5 mm f/2.0 lens (55° field of view), mounted to the ultrasound headset (D in **Figure 2a**). Video was digitized at a 640 × 480 pixel resolution using an Imaging Source DFG/USB2pro and deinterlaced to 60 fps. Front-view video was simultaneously recorded at 1920 × 1080 pixels and 120 fps using a Sony DSC RX10-III digital camera, mounted above the display used to present the prompts. Audio was captured with an AKG C544 L headset condenser microphone and continuously recorded at 48 kHz/16-bit by a Marantz PMD661 Mk2. Audio was simultaneously recorded to disk in Articulate Assistant Advanced (AAA; Articulate Instruments Ltd., 2012), including signals from the Articulate Instruments PStretch and SyncBrightUp units that were used to synchronize the acoustic, ultrasound, and side-view video data. Front-view video was synchronized by aligning acoustic landmarks (e.g., bursts) with those present in the recording from the camera's built-in microphone.

Participants repeated the wordlist with words embedded in the carrier phrase "say ___ again," repeated three times in succession. Prompts were presented with AAA in uniquely pseudorandomized orders. No words containing the same vowel appeared in successive order, nor did words containing either LOT or THOUGHT. Prompts advanced automatically at a pace based on the participant's natural speech rate, established during three practice trials. The duration of each trial was typically 5–7 seconds, including stimulus presentation, recording of the synchronization signal, and the speaker's utterance. A palate trace was captured at the start of recording (Stone, 2005). In addition, the occlusal plane was imaged using a tongue depressor held against the tongue



(a) Front view (b) Side view

Figure 2: Measurement points for front- and side-view lip video data. Illustrations traced from still images of THOUGHT and LOT (side-view, dashed) by CHI008. THOUGHT is protruded and out-rounded, with the interior of the lower lip visible in the front view; LOT is spread. Upper portions of the ultrasound headset not shown.

surface between the upper and lower teeth (Scobbie et al., 2012). Total duration of the production task (both experiments 1 and 2) was approximately 25 minutes.

2.1.3. Analysis

Acoustic recordings were segmented using FAVE-align v1.2.2 (Rosenfelder et al., 2015) and manually corrected. Dynamic formant measurements were taken at 2 ms intervals within each target vowel using the FastTrack plugin for Praat (Barreda, 2021; Boersma & Weenink, 2023). Measurements were normalized with the Lobanov z-score method (Lobanov, 1971) using the `phonTools` package for R (Barreda, 2015; R Core Team, 2024). Normalized formants were re-scaled to Hertz units based on the overall mean and standard deviation of F1 ($\mu = 650.7$, $\sigma = 150$) and F2 ($\mu = 1595.5$, $\sigma = 435.2$) for the *Atlas of North American English* (ANAE), as reported by Dinkin (2022). This ensures that the reported formant values are as consistent as possible with the *ANAE* benchmarks for participation in the NCS and also compatible with Dinkin's (2022) Lobanov-adjusted equivalents. Acoustic statistical models were fit with the original z-scores.

Tongue splines were fit with DeepLabCut (Mathis et al., 2018; Nath et al., 2019) using a MobileNet1.0-based model in AAA v221.3.7 (Wrench & Balch-Tomes, 2022). Eleven points along the upper midsagittal tongue surface were labeled in addition to two points at the base of the mandible and at the mental spine where the short tendon attaches to the genioglossus. The latter point is used as a measure of jaw height in Experiment 2, further described there. Points corresponding to the tongue surface were transformed into polar coordinates following Mielke (2015). Radial coordinates (distance from the virtual origin of the transducer) were z-score normalized across all tokens for each speaker. Angular coordinates were normalized through rotation so that the occlusal plane was horizontal (Scobbie et al., 2012). The median of the highest point of the tongue for /i/ was identified and defined as 0° vertical. Angular coordinates anterior to this reference (the tongue front) were scaled proportionally to a 0.4 radian maximum. Posterior coordinates (tongue body to the epiglottic vallecula) were scaled proportionally to a minimum of -0.8 radians.

Lip gestures were tracked using ResNet50-based models (He et al., 2016; Insafutdinov et al., 2016) in DeepLabCut (version 2.3.2) for the front-view and side-view video. Twenty training frames for each speaker were manually labeled, as illustrated in **Figure 2**. For the front-view video (**Figure 2a**), points included the left and right oral commissures (B), the upper and lower lips at the midsagittal oral mucosa (A), and four stable points on the headset (C). For the side-view video (**Figure 2b**), points were the upper and lower lips at the oral mucosa (E, F) and the oral commissure (G). First-pass training of each network was performed for 800,000 iterations. Twenty outlier frames for each speaker were then labeled and added to the training sets. These augmented data sets were used to train new networks for another 800,000 iterations.

Vertical lip openness and horizontal lip spread were quantified as the distance between the upper and lower points and the oral commissures, respectively. Lip aperture was calculated from these values with the formula $A = \pi * \frac{\text{openness}}{2} * \frac{\text{spread}}{2}$, i.e., the area of an ellipse. Because the distance from the speaker to the RX10 camera was variable, measurements were scaled relative to the fixed distance between points tracked on the headset. Lip protrusion was defined as the horizontal distance of the lower lip coordinate (F) from the posterior edge of the side-view video frame (H, **Figure 2b**). Measurements were z-score normalized.

Acoustic contrast between LOT and THOUGHT was quantified using Pillai scores. A Pillai score is a test statistic from multivariate ANOVA (MANOVA), which serves as an indicator both of the distance between the vowel means and the overlap of their distributions in F1 × F2 space (Hay, Nolan, & Drager, 2006; Nycz & Hall-Lew, 2014). Scores are between 0 (identical distributions) and 1 (distinct distributions). In addition to vowel category (LOT vs. THOUGHT), MANOVAs included a combined term for the onset and coda consonants, in order to account for phonological context. Pillai scores were calculated with measurements sampled at 35% of the vowel's duration (Kendall & Fridland, 2017). The potential for the vowels to be distinguished by differences in their dynamic formant trajectories was also examined by sampling measurements at five equidistant points within the 20–80% interval of the vowel (R. A. Fox & Jacewicz, 2009; Fung & Lee, 2019; Jacewicz & Fox, 2020; Jibson, 2021). These measurements were used to calculate a separate, time-varying Pillai score at each time point.

Dynamic acoustic and articulatory trajectories were analyzed using generalized additive mixed models (GAMMs; Wood 2017) calculated in R with `mgcv::bam()` (R Core Team, 2024; Wood, 2011). By-speaker models were fit for each of the dependent variables F1, F2, lip protrusion, lip aperture, tongue body retraction, and tongue dorsum height. Models included a parametric term for vowel category (as an ordered factor) plus reference/difference smooths (Wieling, 2018) for vowel category over normalized time. Random reference-difference smooths were fit for word (Sóskuthy, 2021) and the models included a first-order autoregression error model. An additional set of temporospatial GAMMs were fit for the ultrasound data in a similar fashion, but with two-dimensional tensor product smooths that model tongue position according to time and location along the midsagittal tongue surface (Carignan et al., 2020). Separate reference-difference random smooths by word were fit for time and tongue position.

2.2. Acoustic Results

Speakers are first assessed in terms of their participation in the NCS using benchmarks from the *ANAE* (Labov et al., 2006), of which two are applicable to the present study. AE1 identifies speakers with advanced TRAP-raising (F1 below 700 Hz), while O2 identifies speakers with advanced LOT-fronting (F2 above 1450 Hz). The fronting of GOOSE is also considered given its potential (socioindexical) association with non-NCS vowel systems (Clopper et al., 2019;

D'Onofrio & Benheim, 2020). For present purposes, speakers are categorized by whether GOOSE F2 is front of center (U2 > 1595.5 Hz). **Figure 3** provides normalized by-speaker formant means with AE1, O2, and U2 indicated for reference. Measurements were sampled at the maximum F2 (TRAP) or F1 (all other vowels), the inflection points used by the *ANAE*.

Overall mean F1 for TRAP is 693 Hz (SD 89). Eleven speakers satisfy the AE1 criterion, with the most extreme raising seen for CHI003 (female, born 1963). As highlighted in **Figure 3**, her mean TRAP F1 is 606 Hz (SD 43), lower than the mean F1 of GOAT for any speaker (min. 611 Hz) and approaching high front FLEECE. Her LOT and THOUGHT are likewise consistent with the NCS; LOT is fronted well beyond 1450 Hz (mean F2: 1505 Hz) and is clearly distinct from THOUGHT (the mean F2 of which is also high: 1368 Hz). In regions where the NCS is undergoing reversal, younger speakers adopt a continuous or nasal TRAP system (D'Onofrio & Benheim, 2020; Nesbitt, 2023; Wagner et al., 2016) with raising only in pre-nasal contexts. Here, only one TRAP item had a non-oral coda (*dan*) so it was excluded from calculation of the mean. As most speakers thus have relatively raised TRAP in exclusively oral contexts, its position is generally more NCS-like than LBMS-like. CHI001 (female, born 1994) has the lowest and most retracted TRAP (F1: 783 Hz, F2: 1706 Hz) and will be seen to have one of the least distinct low back contrasts.

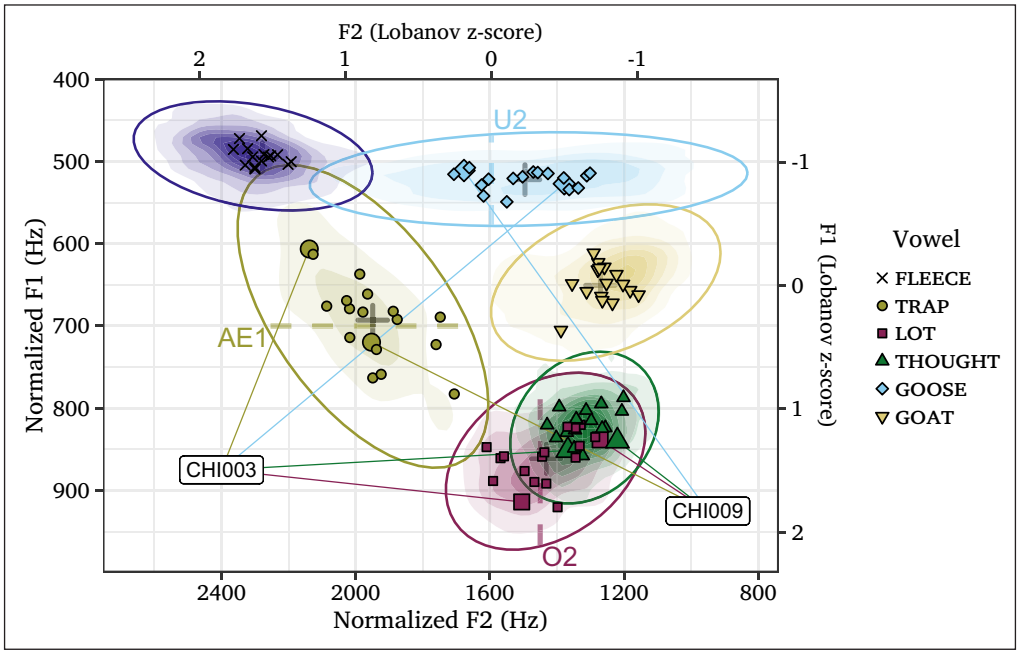


Figure 3: Acoustic measurements for Experiment 1. Points indicate individual vowel category means, cross marks indicate group means. Larger points indicate means for the speakers with **a)** the most raised TRAP (CHI003) and **b)** the least distinct LOT-THOUGHT contrast (CHI009). Distributions are represented by kernel density estimates and 95% confidence interval ellipses. Dashed lines are *ANAE* benchmarks (Labov et al., 2006), equal to the Lobanov-adjusted equivalents (Dinkin, 2022).

The overall mean F2 for LOT is 1432 Hz (SD 141) with the individual means for seven speakers exceeding O2. In contrast to LOT, the F2 of which is highly variable, individual F2 means for THOUGHT are more tightly clustered around the group mean (1316 Hz, SD 109). The F2 of LOT occupies an approximate range of 1250–1650 Hz, nearly encompassing the F2 range of THOUGHT (1200–1450 Hz). The speaker with the least distinct LOT-THOUGHT contrast (in general) is CHI009 (male, born 1998), for whom the mean F2 for LOT and THOUGHT are 1271 Hz (SD 113) and 1220 Hz (SD 112), respectively. As indicated in **Figure 3**, he is also among the speakers with the most advanced GOOSE-fronting (F2: 1677 Hz, SD 325), while his TRAP falls somewhat short of AE1 (F1: 720 Hz, SD 58).

2.2.1. Spectral overlap of LOT and THOUGHT

Figure 4 contains kernel density estimate plots for LOT and THOUGHT by speaker, with Pillai scores and predicted F1 × F2 trajectories within 20–80% of the vowel's duration. The participant with the highest Pillai score is CHI002 (male, born 1962), which at 0.921 ($p < 0.001$) indicates a highly robust acoustic distinction with effectively no overlap of the vowel distributions. He is among the most advanced with respect to LOT-fronting (mean F2: 1568 Hz, SD 61) and produces TRAP with a mean F1 of 714 Hz (SD 39). His LOT and THOUGHT are relatively monophthongal; LOT undergoes almost no F1 or F2 change throughout its duration, while THOUGHT shows a limited increase in F2. Other speakers with high Pillai scores (e.g., CHI003, CHI015, and CHI010) show a more substantial F2 increase for THOUGHT beyond 30–35% of its duration, but produce LOT with flat F1 and F2 trajectories.

As noted above, CHI009 produces the acoustically least distinct LOT-THOUGHT contrast overall and has a Pillai score of 0.098 ($p = 0.019$). **Figure 4** shows that his LOT/THOUGHT distributions almost completely overlap. CHI019's Pillai score is slightly lower at 35% of the vowel's duration, 0.093 ($p = 0.032$), but her vowel distibutions show greater variation over time, as will be examined next. CHI013 has a larger (but still weak) acoustic contrast, with a Pillai score of 0.215 ($p < 0.001$). Scores in this range are typical for regions with established merger (Fridland & Kendall, 2019; Kendall & Fridland, 2017; Nycz & Hall-Lew, 2014; Swan, 2019). As cut-offs for establishing merger through Pillai scores depend on sample size, however, Stanley and Sneller (2023) propose a metric of *e* divided by the average token count per vowel. For this study's sample of 54 tokens per vowel, this yields a limit of 0.0503. While CHI009 and CHI019's Pillai scores are somewhat higher than this value, their overall vowel distributions show extensive overlap.

Pillai scores calculated at multiple points within the vowel intervals are given in **Figure 5**, visualizing how the degree of overlap varies as the result of vowel-internal formant dynamics (Farrington et al., 2018; R. A. Fox & Jacewicz, 2009). For speakers with high Pillai scores, CHI002 the highest among them, scores are stable regardless of the measurement sampling time. From

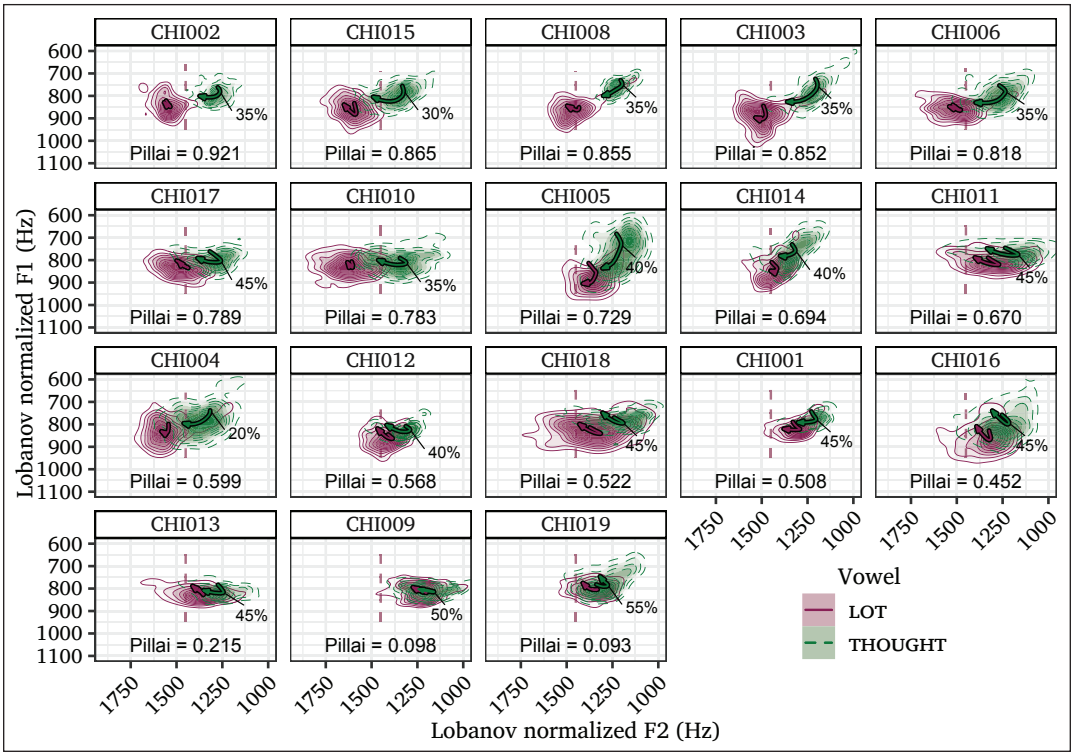


Figure 4: Kernel density estimates for F1 and F2 of LOT and THOUGHT in normal speech task, sampled at five equidistant time points within 20–80% of the vowel's duration. Arrows indicate GAMM-predicted formant trajectories for the same interval, labeled with the time of the predicted F2 minimum for THOUGHT. Vertical line indicates O2 criterion for LOT-fronting (1450 Hz). Speakers arranged by LOT-THOUGHT Pillai score, highest at upper left.

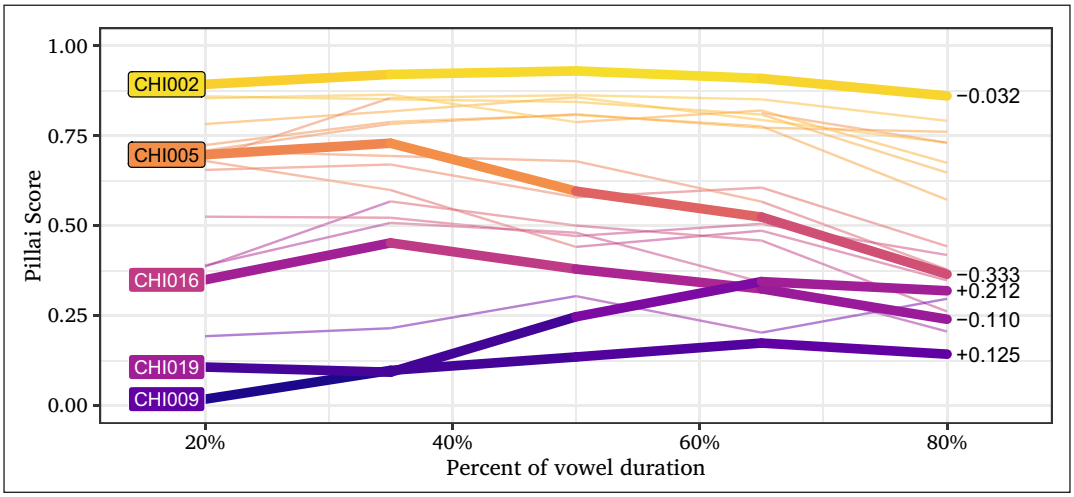


Figure 5: Pillai scores calculated at five equidistant points within the vowel's duration. Speakers highlighted are those with the highest and lowest scores at any timepoint, the greatest within-vowel increase or decrease, and the speaker with the median score.

20–80% of the vowel interval, CHI002's score varies by no more than ±0.035. This contrasts starkly from CHI005, who was observed in **Figure 4** to have one of the most dynamic formant trajectories for THOUGHT. Because her LOT is comparatively less fronted and is also temporally stable, the acoustic overlap between LOT and THOUGHT increases as THOUGHT is lowered and fronted throughout its duration. As a result, her Pillai score is 0.729 when calculated at 35% of the vowel interval, but only 0.365 when calculated at 80%. For speakers with low Pillai scores, scores tend to increase toward the vowel offset, rather than decrease. CHI009 has the lowest overall Pillai score of 0.018 at 20% of vowel duration, increasing to 0.14 toward the vowel offset. CHI019 shows the greatest increase of all speakers, from 0.11 to 0.32. In **Figure 4**, her vowel distributions reveal a number of THOUGHT tokens that are higher and backer than any for LOT, which occur during the offglide.

2.2.2. Dynamic formant trajectories

Table 2 summarizes all by-speaker F1/F2 GAMMs for LOT and THOUGHT, with Pillai scores and *ANAE* benchmarks. Parametric coefficients represent the constant F1/F2 difference (in Lobanov *z*-score units) of LOT from THOUGHT (the intercept), while difference smooths correspond to non-linear differences in F1/F2 trajectory (with THOUGHT as the reference smooth). For most speakers, the constant F1 and F2 differences between LOT and THOUGHT are highly significant ($p < 0.001$). For CHI009, CHI010, and CHI013, the F1 difference between LOT and THOUGHT is not significant, nor is the F2 difference for CHI009. CHI013 and CHI019, on the other hand, show significant intercept differences in F2 (CHI013: 0.191, $p = 0.001$, CHI019: 0.125, $p = 0.008$), while all three exhibit significant non-linear formant differences. Nevertheless, when considered with the small or non-significant intercept differences, the trajectories show little time-varying change.

Among speakers with a Pillai score above 0.8, all exhibit significant differences in both of the formant trajectories. For most speakers with Pillai scores below this level, the smooth for F1 of LOT does not significantly differ from the reference smooth, indicating similarity in the shape of the F1 trajectories. For several other speakers, the difference smooths have an edf (effective degrees of freedom) at or close to 1, indicating a linearly-changing difference over time (e.g., F1 for CHI016), although the difference is not significant. A significant linear difference smooth is observed for CHI005, which accounts for her steadily decreasing Pillai scores in **Figure 5**.

2.2.3. Duration

Besides diphthongization, spectrally similar vowels may potentially show differences in their duration. Several studies indicate that vowels which appear to be merged in F1 × F2 space may instead be maintained through durational contrast (Fridland et al., 2014, e.g., specifically examined LOT and THOUGHT), which has been noted as a factor in the avoidance of contrast collapse in

						Parametric coefficient				Difference smooth			
Speaker	**Pillai**	**AE1**	**O2**	**U2**	**DV**	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
CHI002 (1962, M)	0.921		Y		F1	0.289	0.041	7.06	<0.001	5.78	7.2	4.91	<0.001
					F2	0.544	0.028	19.4	<0.001	7.26	9.14	8.06	<0.001
CHI015 (1953, F)	0.865	Y	Y		F1	0.404	0.0449	9.01	<0.001	9.07	11.5	2.16	0.025
					F2	0.491	0.0366	13.4	<0.001	6.91	8.73	7.22	<0.001
CHI008 (1955, M)	0.855	Y	Y		F1	0.571	0.0625	9.15	<0.001	6	7.37	9.1	<0.001
					F2	0.488	0.0492	9.92	<0.001	5.58	6.96	7.4	<0.001
CHI003 (1963, F)	0.852	Y	Y		F1	0.662	0.0892	7.42	<0.001	7.82	10	3.57	<0.001
					F2	0.53	0.0525	10.1	<0.001	7.22	9.15	12.4	<0.001
CHI006 (1952, F)	0.818	Y	Y		F1	0.403	0.0473	8.52	<0.001	10.1	12.7	4.47	<0.001
					F2	0.458	0.0439	10.4	<0.001	5.47	6.56	7.03	<0.001
CHI017 (1955, F)	0.789	Y			F1	0.185	0.0635	2.92	0.007	**1**	**1**	**3.72**	**0.104**
					F2	0.335	0.0435	7.72	<0.001	7.68	9.75	7.45	<0.001
CHI010 (1948, M)	0.783	Y	Y		F1	**0.0867**	**0.047**	**1.85**	**0.130**	**1**	**1**	**0.469**	**0.964**
					F2	0.511	0.0539	9.48	<0.001	9.1	11.6	8.34	<0.001
CHI005 (1941, F)	0.729			Y	F1	0.675	0.0878	7.69	<0.001	**4.51**	**5.71**	**2.41**	**0.054**
					F2	0.298	0.0291	10.2	<0.001	1	1	8.22	0.008
CHI014 (1992, F)	0.694				F1	0.487	0.077	6.32	<0.001	**1**	**1**	**2.89**	**0.179**
					F2	0.192	0.0255	7.53	<0.001	3.48	4.31	9.21	<0.001
CHI011 (1997, F)	0.67			Y	F1	0.278	0.0356	7.81	<0.001	**3.37**	**4.12**	**0.927**	**0.866**
					F2	0.189	0.0388	4.88	<0.001	4.39	5.67	4.88	<0.001

(Contd.)

						Parametric coefficient				Difference smooth			
Speaker	**Pillai**	**AE1**	**O2**	**U2**	**DV**	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
CHI004 (1947, F)	0.599		Y	Y	F1	0.332	0.103	3.21	0.003	**4.12**	**5.14**	**1.56**	**0.329**
					F2	0.403	0.0547	7.36	<0.001	4.93	6.3	13.4	<0.001
CHI012 (1981, F)	0.568	Y			F1	0.175	0.0416	4.2	<0.001	**3.29**	**3.95**	**1.62**	**0.316**
					F2	0.173	0.027	6.41	<0.001	3.18	3.7	6.68	<0.001
CHI018 (1991, F)	0.522	Y		Y	F1	0.319	0.0634	5.03	<0.001	**4.37**	**5.43**	**2.18**	**0.106**
					F2	0.299	0.0575	5.19	<0.001	5.21	6.62	2.6	0.022
CHI001 (1994, F)	0.508			Y	F1	0.239	0.045	5.32	<0.001	4.77	5.9	3.43	0.004
					F2	0.205	0.026	7.9	<0.001	**1.97**	**2.27**	**1.23**	**0.675**
CHI016 (1961, F)	0.452	Y		Y	F1	0.466	0.137	3.4	0.001	**1**	**1**	**0.00123**	**>0.999**
					F2	0.18	0.0554	3.25	0.002	3.87	4.77	4.18	0.002
CHI013 (1995, F)	0.215	Y		Y	F1	**–0.00616**	**0.0826**	**–0.0746**	**>0.999**	5.34	6.64	4.56	<0.001
					F2	0.191	0.0584	3.27	0.002	3.43	4.15	2.77	0.048
CHI009 (1998, M)	0.0975			Y	F1	**0.0034**	**0.0255**	**0.133**	**>0.999**	3.95	4.75	3.56	0.009
					F2	**0.0657**	**0.0471**	**1.39**	**0.327**	4.21	5.14	2.89	0.027
CHI019 (1998, F)	0.0925	Y		Y	F1	0.153	0.0628	2.44	0.029	6.62	8.41	2.84	0.006
					F2	0.125	0.047	2.65	0.016	5.13	6.51	5.91	<0.001

Table 2: GAMM estimates and binary difference smooths for F1 and F2 of LOT (Lobanov z-score) relative to THOUGHT. Intercept and reference smooth (for THOUGHT) omitted. O2, AE1, and U2 indicate whether speaker meets criteria for LOT-fronting, TRAP-raising, and GOOSE-fronting. P-values are Bonferroni corrected to account for multiple comparisons (parametric + smooth). Non-significant terms given in boldface.

cases of near merger (Labov & Baranowski, 2006; Wade, 2017). However, that possibility is not borne out for Chicagoans. **Figure 6** gives kernel density estimates of the normalized log duration for LOT and THOUGHT by Pillai score. As anticipated, LOT and THOUGHT are longer in syllables with voiced codas as compared to those with voiceless codas.

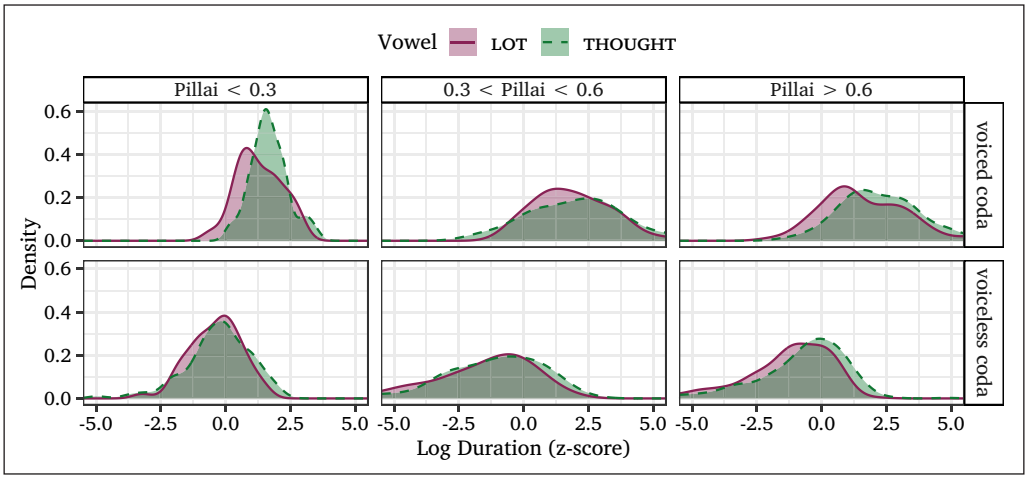


Figure 6: Duration of LOT and THOUGHT by Pillai score in syllables with voiced and voiceless codas.

Linear mixed effects regression confirms that there is no LOT-THOUGHT duration contrast irrespective of the degree of spectral overlap. The model in **Table 3** was fit with fixed effects of coda voicing and vowel category, plus by-speaker random slopes for vowel and by-word random intercepts. The main effect of vowel is not significant and likelihood ratio tests show that dropping this term from the model does not significantly worsen model fit ($\chi^2(1)$, $p = 0.074$), nor does including its interaction with voicing significantly improve model fit ($\chi^2(1)$, $p = 0.723$). While inclusion of Pillai score as a categorical predictor (high/medium/low) does improve model fit ($\chi^2(2)$, $p = 0.019$), an interaction of Pillai score and vowel class does not ($\chi^2(2)$, $p = 0.606$). This result indicates that speakers with lower Pillai scores are not more likely to distinguish the vowels by differences in duration and vice versa.

Predictor	**Estimate**	**SE**	**t value**	**Pr(>|t|)**	
(Intercept)	0.664	0.163	4.074	$p < 0.001$	***
Vowel					
LOT	–0.271	0.154	–1.757	$p = 0.086$	
Coda voicing					
voiced	1.653	0.154	10.698	$p < 0.001$	***

Table 3: Mixed effects linear regression model for duration by vowel class and voicing.

2.2.4. Relationship of vowel formants to Pillai score and age

ANAE benchmarks in **Table 2** reveal that all speakers but one (CHI004) who have a LOT F2 above 1450 Hz also have a Pillai score above 0.7, suggesting a possible relationship between the frontedness of LOT and its overall acoustic contrast with THOUGHT, especially as THOUGHT appeared to be relatively less variable than LOT. Moreover, all speakers who satisfy O2 were born prior to 1965, consistent with possible LOT-retraction in apparent time. Positions of the NCS/LBMS-associated vowels and their relationship to speaker Pillai scores, age, and to each other were assessed through Pearson correlation tests (Kendall & Fridland, 2017), summarized in **Table 4**. Pillai score is significantly correlated with birth year ($r = -0.69$, $p = 0.002$), as visualized in **Figure 7**. Older speakers tend to have higher Pillai scores, consistent with their relatively more fronted productions of LOT.

Associations of Pillai score with F1 and F2 of TRAP, LOT, and THOUGHT, as well as F2 of GOOSE, are shown in **Figure 8**. F1 and F2 of LOT show significant positive correlations with age, showing that older speakers produce LOT with a fronter and (acoustically) lower target. By contrast, neither F1 nor F2 of THOUGHT exhibits such a relationship with Pillai score (or with speaker age), confirming its relative lack of variance. TRAP's involvement in both the NCS and LBMS also predicts an association with the position of LOT, which is borne out. A lower and backer TRAP is correlated with a lower Pillai score and a more retracted LOT. It is worth noting, however, that raised productions of TRAP, while most common among the speakers with the highest Pillai scores, are also observed for several speakers with weak LOT-THOUGHT contrasts. CHI019 and CHI013, who have the lowest and third-lowest Pillai scores, produce TRAP with F1 means reasonably close to AE1: 689 Hz (SD 124) and 692 Hz (SD 71), respectively.

2.3. Articulatory results: Tongue position

Consistent with variability in the acoustic strength of the LOT-THOUGHT contrast, speakers also vary as to whether they distinguish the vowels by tongue position. Mid-sagittal tongue contours for LOT and THOUGHT at 35% of the vowel's duration are presented in **Figure 9** for four speakers. Contours represent midsagittal tongue shape with the tongue front at right. The 95% confidence interval for each spline is shaded; overlap suggests that tongue positions for LOT and THOUGHT do not differ significantly at that region. CHI003 and CHI010, whose Pillai scores exceed 0.75, produce LOT and THOUGHT with distinct tongue positions in line with their acoustics. For these speakers, the tongue body for LOT is both higher and less retracted than for THOUGHT, consistent with earlier descriptions of LOT-fronting as also involving raising (Labov et al., 2006, p. 17). Lingual LOT-THOUGHT contrast is smaller for CHI001, whose acoustic contrast is moderate (Pillai = 0.508), although just-significant trends in the expected direction are apparent. For CHI009, whose Pillai score is indicative of complete merger, the two vowels have identical tongue positions.

Measure	Pillai score		Speaker age		TRAP-LOT ED		TRAP F1		TRAP F2		LOT F2	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
TRAP F1	**−0.49**	**0.041**	**−0.64**	**0.004**	**−0.90**	**<0.001**	–	–	–	–	–	–
TRAP F2	**0.61**	**0.007**	**0.69**	**0.002**	**0.86**	**<0.001**	–	–	–	–	–	–
LOT F1	**0.70**	**0.001**	**0.57**	**0.014**	**0.88**	**<0.001**	**−0.88**	**<0.001**	**0.90**	**<0.001**	–	–
LOT F2	**0.71**	**<0.001**	**0.71**	**<0.001**	0.23	0.359	**−0.56**	**0.015**	**0.65**	**0.004**	–	–
THOUGHT F1	−0.38	0.116	−0.33	0.182	−0.14	0.589	0.04	0.884	−0.09	0.721	0.00	0.999
THOUGHT F2	0.11	0.658	0.23	0.365	−0.23	0.368	−0.04	0.870	0.18	0.478	**0.63**	**0.005**
GOOSE F1	0.34	0.163	0.33	0.186	0.31	0.218	−0.28	0.261	0.18	0.485	−0.02	0.945
GOOSE F2	**−0.80**	**<0.001**	**−0.70**	**0.001**	−0.44	0.064	**0.61**	**0.007**	**−0.79**	**<0.001**	**−0.85**	**<0.001**
Speaker Age	**0.69**	**0.002**	–	–	–	–	–	–	–	–	–	–

Table 4: Pearson correlation tests of Pillai score, vowel formants, and speaker age. Significant correlations shown in boldface.

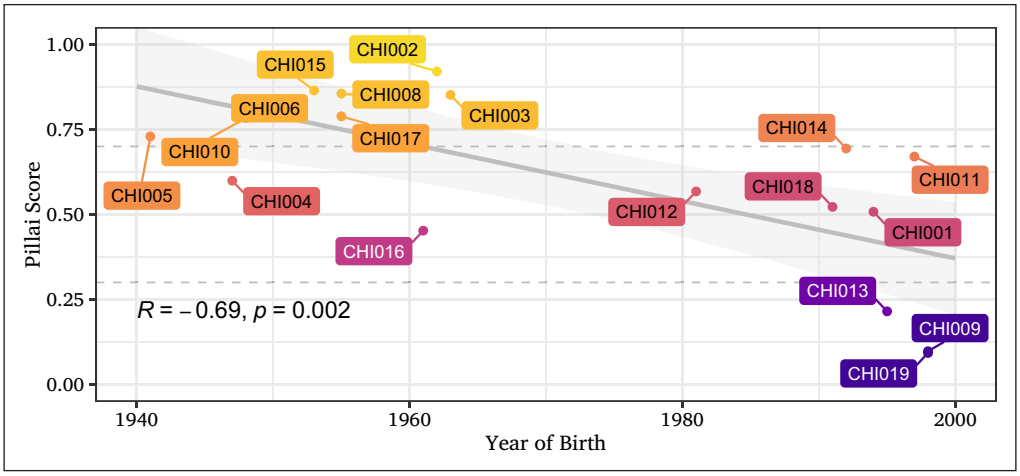


Figure 7: Pillai scores for LOT and THOUGHT by year of birth at 35% of the vowel's duration. Dashed lines indicate boundaries for high vs medium vs low acoustic contrasts.

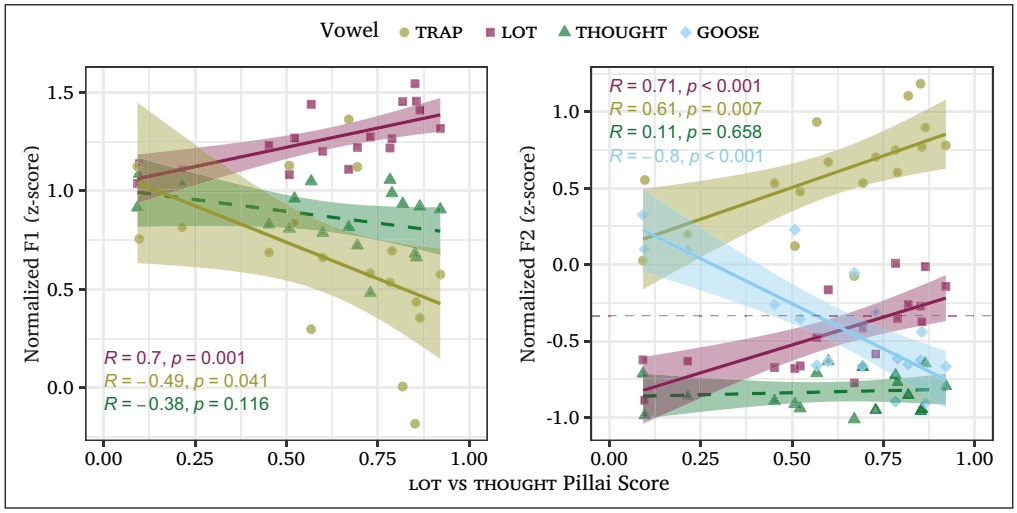


Figure 8: Relationship of Pillai score to mean F1 and F2 measurements for TRAP, LOT, THOUGHT, and GOOSE, with Pearson product-moment correlation coefficients.

Lingual LOT-THOUGHT differences throughout the full vowel interval are shown as three-dimensional difference surfaces in **Figure 10**, obtained with `itsadug::get_difference()`. Like CHI003 and CHI010, the other six speakers with Pillai scores above 0.7 produce LOT and THOUGHT with clear differences in tongue position. The difference is greatest during the initial 50% interval, coinciding with the typical F2 minimum for THOUGHT, at which point acoustic distance between LOT and THOUGHT is most extreme. The lingual LOT-THOUGHT difference is not significant at the offset, consistent with the ingliding trajectory for THOUGHT F2. As most of these speakers (except CHI005 and CHI017) also produce LOT with an F2 exceeding 1450 Hz, LOT-THOUGHT F2 differences correspond well with tongue position. Speakers with Pillai scores between 0.3–0.7

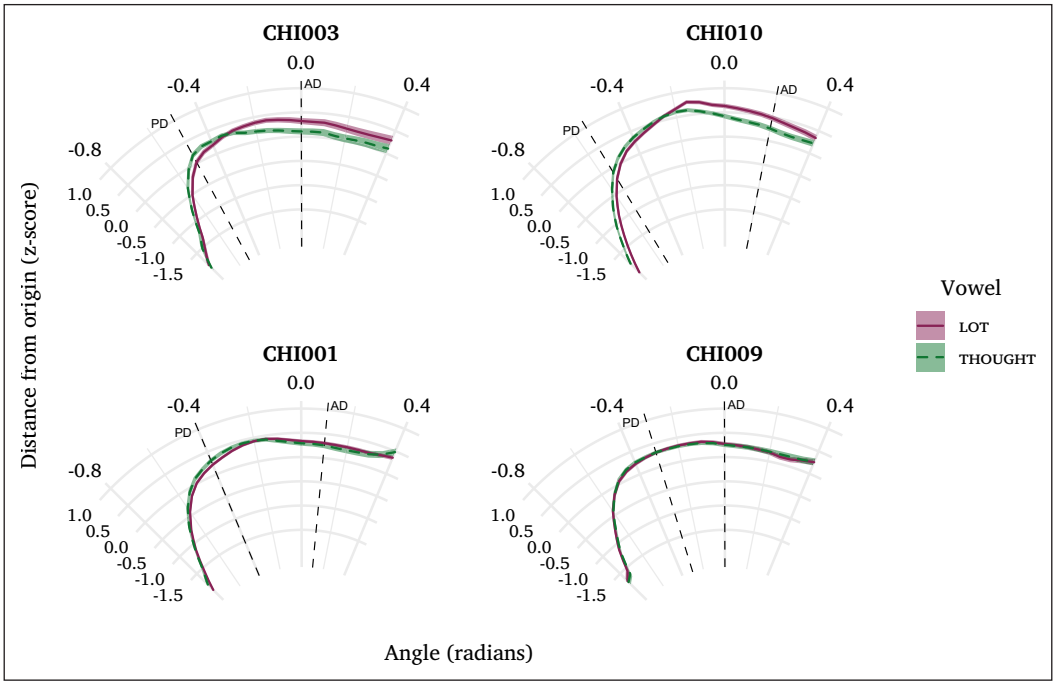


Figure 9: Predicted mid-sagittal GAMM tongue contours at 35% of vowel duration for four representative speakers. Shading indicates 95% confidence interval. Tongue front is to the right. Speakers ordered by Pillai score, highest at upper left.

(CHI004–CHI016) less consistently distinguish LOT and THOUGHT by tongue position. Within this group, CHI004 and CHI018 are similar to the speakers with high Pillai scores, while CHI001, CHI011, and CHI014 produce THOUGHT with greater retraction than LOT, but without a significant difference at the tongue front. A marginal difference for CHI012 runs counter to the expected direction; neither CHI012 or CHI016 use tongue position to achieve their relatively weaker acoustic contrasts. Likewise, CHI009 shows no difference in tongue position at any point in the vowel's duration, consistent with complete merger. On the other hand, CHI013 produces the two vowels with distinct tongue positions throughout their entire duration, despite their acoustic proximity. CHI019 shows some difference in tongue body retraction toward the vowel offset, whereby THOUGHT remains retracted while LOT begins to front, consistent with her F2 trajectory for THOUGHT.

Lingual differences were formally tested with GAMMs fit to two quantitative measures of tongue position. The radial axes of the maximal LOT-THOUGHT difference were identified from the GAMM difference surfaces, along both the anterodorsal (between 0 and 0.3 radians) and posterodorsal (between -0.7 and -0.3 radians) regions of the tongue. These axes are indicated with dashed lines in **Figures 9** and **10**. The distance from the probe origin to the tongue surface along each axis was then determined throughout the entire duration of each vowel. Similar to the approach used by Mielke et al. (2017), anterodorsal distance was subtracted from the posterodorsal distance and the result was divided by two, providing a measure of overall tongue

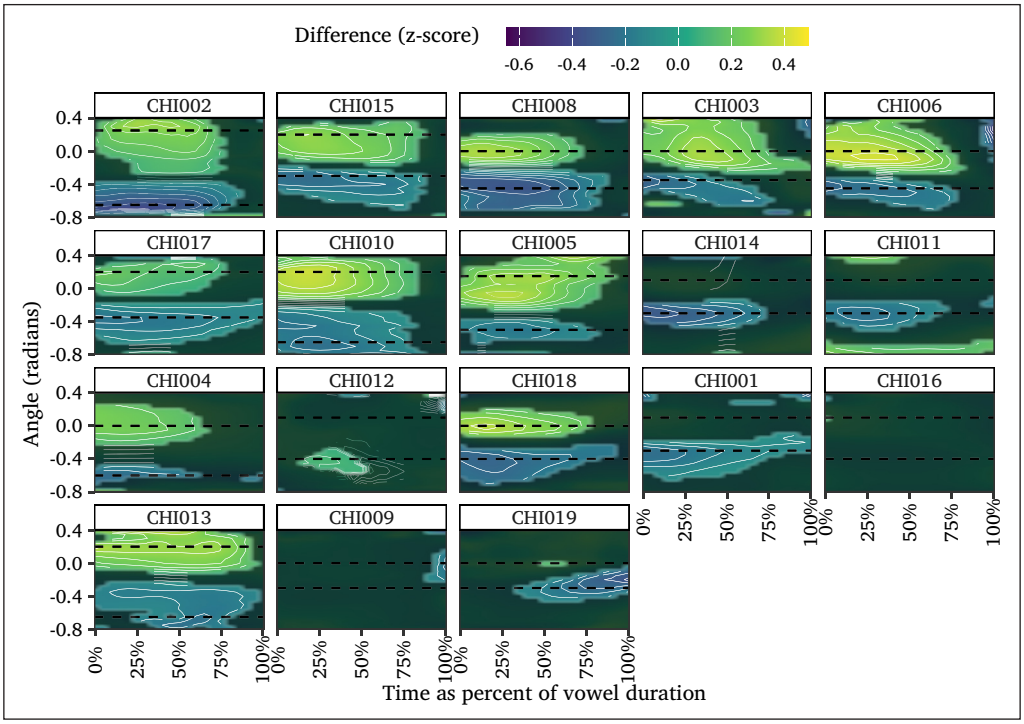


Figure 10: Predicted GAMM difference surface for LOT vs THOUGHT. Slices along the *x* axis indicate the mid-sagittal difference in tongue position at successive moments in time. Radial axes along the tongue's surface are on the *y* axis, tongue front at top. Lighter colors at the top (more anterior angles) indicate a higher tongue front for LOT compared to THOUGHT, and darker colors at the bottom (more posterior angles) indicate less tongue root retraction for LOT compared to THOUGHT. Dark shading indicates times/regions where the difference is not significant (95% CI). Dashed lines indicate speaker-specific locations of maximal antero- and posterodorsal differences, used to quantify tongue body retraction.

body retraction. Larger values for this measure correspond to greater retraction of the tongue body. The y-coordinate for the DeepLabCut-tracked point corresponding to the tongue dorsum (Dorsum1) was used as a measure of overall tongue dorsum height. Higher values indicate a more raised tongue position. These measurements were calculated for all vowels (including /i, æ, u, o/) and z-score normalized. The range of the z-scores therefore also reflect the positions of LOT and THOUGHT within the articulatory vowel space.

Model summaries for both lingual measures are provided in **Table 5**. These models confirm that most speakers with Pillai scores above 0.7 produce LOT with a significantly less retracted tongue body and a significantly higher tongue dorsum. Visualization of the predicted trajectories for each measure in **Figure 11** show that the significant non-linear differences between the two vowels are such that the difference in tongue height and retraction is greatest within the 25–50% interval of the vowel's duration, during which time THOUGHT is the backest and lowest vowel in the entire system. Although visual comparisons of the difference surfaces for speakers

			Parametric coefficient				Difference smooth			
Speaker	**Pillai**	**DV**	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
CHI002 (1962, M)	0.921	TBR	–0.451	0.13	–3.46	0.001	**2.79**	**3.18**	**2.3**	**0.143**
		TDH	0.15	0.0511	2.93	0.007	3.66	4.15	4.15	0.004
CHI015 (1953, F)	0.865	TBR	–0.458	0.0667	–6.86	<0.001	3.87	4.33	5.3	<0.001
		TDH	0.236	0.0616	3.84	<0.001	5.08	5.87	4.86	<0.001
CHI008 (1955, M)	0.855	TBR	–0.416	0.103	–4.05	<0.001	3.21	3.64	3.12	0.040
		TDH	0.125	0.0404	3.11	0.004	**2.73**	**3.03**	**1.17**	**0.620**
CHI003 (1963, F)	0.852	TBR	–0.276	0.0749	–3.68	<0.001	2.42	2.58	4.05	0.019
		TDH	0.219	0.046	4.76	<0.001	5.92	6.74	10.1	<0.001
CHI006 (1952, F)	0.818	TBR	–0.288	0.0857	–3.36	0.002	2.68	3.02	11.1	<0.001
		TDH	0.192	0.0427	4.51	<0.001	6.04	7.08	7.53	<0.001
CHI017 (1955, F)	0.789	TBR	–0.393	0.136	–2.89	0.008	2.73	3.04	4.97	0.004
		TDH	**0.00816**	**0.092**	**0.0887**	**>0.999**	**1.55**	**1.64**	**0.891**	**0.554**
CHI010 (1948, M)	0.783	TBR	–0.447	0.0656	–6.8	<0.001	1.61	1.75	6.82	0.008
		TDH	0.322	0.107	3.02	0.005	**2.31**	**2.47**	**1.74**	**0.393**
CHI005 (1941, F)	0.729	TBR	–0.392	0.133	–2.95	0.006	**1.4**	**1.53**	**3.79**	**0.128**
		TDH	0.34	0.143	2.37	0.036	4.19	5.11	4.99	<0.001
CHI014 (1992, F)	0.694	TBR	**–0.134**	**0.0758**	**–1.77**	**0.153**	2.14	2.29	4.42	0.017
		TDH	**–0.0614**	**0.044**	**–1.4**	**0.326**	**1**	**1**	**0.818**	**0.732**
CHI011 (1997, F)	0.67	TBR	**0.148**	**0.11**	**1.34**	**0.363**	**1**	**1**	**0.452**	**>0.999**
		TDH	**–0.11**	**0.102**	**–1.07**	**0.567**	**1**	**1**	**1.09**	**0.591**

(Contd.)

			Parametric coefficient				Difference smooth			
Speaker	**Pillai**	**DV**	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
CHI004 (1947, F)	0.599	TBR	–0.466	0.129	–3.61	<0.001	2.18	2.65	6.52	0.001
		TDH	0.352	0.111	3.17	0.003	2.32	2.84	10.7	<0.001
CHI012 (1981, F)	0.568	TBR	**–0.0256**	**0.0283**	**–0.906**	**0.731**	**3.38**	**4.01**	**1.51**	**0.397**
		TDH	**0.107**	**0.0492**	**2.16**	**0.061**	**1.3**	**1.45**	**1.68**	**0.538**
CHI018 (1991, F)	0.522	TBR	–0.299	0.0957	–3.12	0.004	**1.1**	**1.13**	**0.751**	**0.733**
		TDH	**0.131**	**0.102**	**1.28**	**0.401**	**1.79**	**1.93**	**0.473**	**>0.999**
CHI001 (1994, F)	0.508	TBR	**–0.105**	**0.0538**	**–1.95**	**0.103**	**1**	**1**	**3.87**	**0.099**
		TDH	**–0.00851**	**0.0529**	**–0.161**	**>0.999**	**1.27**	**1.32**	**0.728**	**>0.999**
CHI016 (1961, F)	0.452	TBR	**–0.053**	**0.107**	**–0.496**	**>0.999**	**1**	**1**	**1**	**0.635**
		TDH	**0.0341**	**0.0607**	**0.561**	**>0.999**	**1**	**1**	**0.297**	**>0.999**
CHI013 (1995, F)	0.215	TBR	–0.381	0.104	–3.68	<0.001	**1.35**	**1.43**	**0.847**	**>0.999**
		TDH	**0.103**	**0.0612**	**1.68**	**0.188**	**1.49**	**1.62**	**1.11**	**0.428**
CHI009 (1998, M)	0.0975	TBR	**0.002**	**0.0998**	**0.02**	**>0.999**	**1**	**1**	**0.094**	**>0.999**
		TDH	**–0.0139**	**0.0519**	**–0.268**	**>0.999**	**1**	**1**	**5.22e-05**	**>0.999**
CHI019 (1998, F)	0.0925	TBR	**–0.206**	**0.0981**	**–2.1**	**0.072**	**3.47**	**3.89**	**2.41**	**0.118**
		TDH	**–0.0299**	**0.102**	**–0.294**	**>0.999**	**1**	**1**	**0.47**	**0.986**

Table 5: GAMM estimates and binary difference smooths for tongue body retraction (TBR) and tongue dorsum height (TDH) of LOT relative to THOUGHT (z-score).

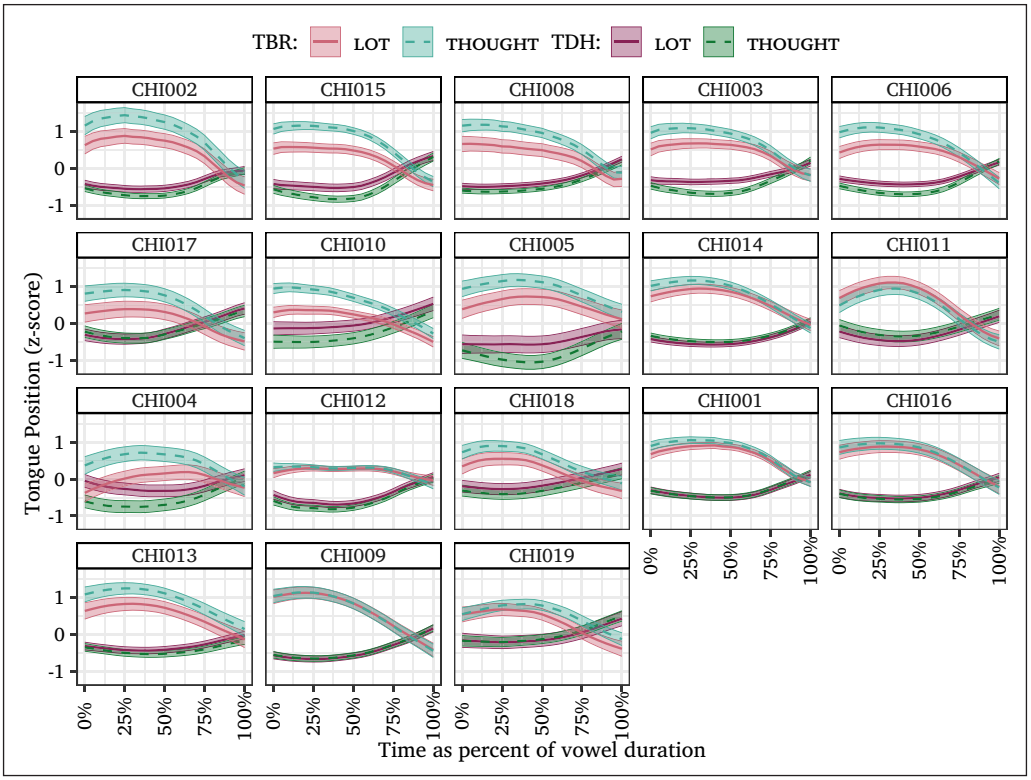


Figure 11: Predicted tongue body retraction (TBR) and tongue dorsum height (TDH) for LOT and THOUGHT, normal speech task. Shading represents 95% confidence interval.

with Pillai scores below 0.7 were suggestive of brief or marginal lingual differences, formal testing of the tongue body retraction and tongue dorsum height measures show that LOT and THOUGHT do not have different tongue positions for most speakers with moderate or low Pillai scores. Notably, however, both CHI013 and CHI019 produce THOUGHT with a significantly more retracted tongue body, which is consistent with significant intercept and non-linear differences observed for F2. For CHI019, the difference in tongue body retraction is largest at the offset, which again coincides with the timing of her greatest acoustic difference between the vowels.

2.4. Articulatory results: Lip rounding

While a lingual distinction between LOT and THOUGHT is consistently observed only for speakers with the highest Pillai scores and most fronted LOT, nearly all speakers distinguish LOT and THOUGHT through differences in lip rounding. Summaries of the GAMM estimates and difference smooths are given in **Table 6**. Whereas differences in lingual retraction, raising, or both were non-significant for most speakers with Pillai scores below 0.7, all speakers (except CHI009) show significant differences between LOT and THOUGHT in terms of lip protrusion. In most cases, LOT and THOUGHT differ in both the overall degree of protrusion and in the non-linear lip

			Parametric coefficient				Difference smooth			
Speaker	**Pillai**	**DV**	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
CHI002 (1962, M)	0.921	LP	1.04	0.0255	40.6	<0.001	5.77	6.55	11.1	<0.001
		LA	0.226	0.0806	2.8	0.011	**3.97**	**4.63**	**1.68**	**0.281**
CHI015 (1953, F)	0.865	LP	0.945	0.0634	14.9	<0.001	7.61	8.9	24.6	<0.001
		LA	0.617	0.0853	7.23	<0.001	6.36	7.29	8.72	<0.001
CHI008 (1955, M)	0.855	LP	0.897	0.0728	12.3	<0.001	6.52	7.59	9.08	<0.001
		LA	–0.408	0.142	–2.87	0.008	**1.73**	**1.98**	**1.31**	**0.571**
CHI003 (1963, F)	0.852	LP	0.996	0.0848	11.7	<0.001	4.66	5.37	11	<0.001
		LA	0.445	0.126	3.53	<0.001	2.85	3.28	3.07	0.042
CHI006 (1952, F)	0.818	LP	0.959	0.0487	19.7	<0.001	8.24	9.72	24.2	<0.001
		LA	–0.221	0.0748	–2.96	0.006	5.65	6.75	9.78	<0.001
CHI017 (1955, F)	0.789	LP	0.738	0.0764	9.65	<0.001	4.8	5.73	10.2	<0.001
		LA	0.404	0.0781	5.17	<0.001	4.33	5.15	6.64	<0.001
CHI010 (1948, M)	0.783	LP	0.551	0.0676	8.15	<0.001	2.84	3.13	7.67	<0.001
		LA	0.311	0.0824	3.78	<0.001	**2.28**	**2.47**	**2.45**	**0.241**
CHI005 (1941, F)	0.729	LP	0.815	0.0495	16.5	<0.001	5.92	7.17	15.5	<0.001
		LA	0.634	0.0873	7.26	<0.001	5.6	6.53	13.4	<0.001
CHI014 (1992, F)	0.694	LP	0.536	0.0692	7.74	<0.001	1	1	12.3	<0.001
		LA	0.307	0.119	2.57	0.020	3.04	3.56	5.18	0.002
CHI011 (1997, F)	0.67	LP	0.44	0.0853	5.16	<0.001	**2.96**	**3.77**	**2.79**	**0.059**
		LA	**0.0703**	**0.0934**	**0.752**	**0.905**	4.08	4.99	3.09	0.019

(Contd.)

			Parametric coefficient				Difference smooth			
Speaker	**Pillai**	**DV**	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
CHI004 (1947, F)	0.599	LP	0.612	0.082	7.46	<0.001	5.19	6.27	15	<0.001
		LA	**0.082**	**0.1**	**0.821**	**0.824**	1.6	1.82	7.52	0.006
CHI012 (1981, F)	0.568	LP	0.475	0.0839	5.67	<0.001	3.79	4.5	3.36	0.015
		LA	0.23	0.0558	4.12	<0.001	4.39	5.1	3.18	0.014
CHI018 (1991, F)	0.522	LP	0.352	0.0847	4.16	<0.001	**2.06**	**2.31**	**0.925**	**0.718**
		LA	0.298	0.0987	3.02	0.005	3.02	3.43	2.99	0.050
CHI001 (1994, F)	0.508	LP	0.922	0.0642	14.3	<0.001	3.62	4.41	4.25	0.003
		LA	**0.125**	**0.0688**	**1.81**	**0.140**	**1**	**1**	**1.24**	**0.532**
CHI016 (1961, F)	0.452	LP	0.374	0.0712	5.24	<0.001	1.2	1.3	5.56	0.019
		LA	**0.105**	**0.102**	**1.03**	**0.611**	2.13	2.63	3.62	0.029
CHI013 (1995, F)	0.215	LP	0.639	0.0833	7.67	<0.001	3.38	4.11	9.06	<0.001
		LA	0.26	0.0744	3.49	0.001	**2.51**	**2.9**	**1.6**	**0.314**
CHI009 (1998, M)	0.0975	LP	**0.149**	**0.118**	**1.26**	**0.417**	**2.05**	**2.38**	**0.955**	**0.616**
		LA	**0.116**	**0.0782**	**1.48**	**0.277**	**1**	**1**	**0.219**	**>0.999**
CHI019 (1998, F)	0.0925	LP	0.242	0.0585	4.14	<0.001	1.28	1.37	12.4	<0.001
		LA	0.133	0.0575	2.31	0.042	1	1	19.9	<0.001

Table 6: GAMM estimates and binary difference smooths for lower lip protrusion (LP) and lip aperture (LA) of LOT relative to THOUGHT (z-score).

trajectories. **Figure 12** provides the predicted GAMM smooths for lower lip protrusion with 95% confidence intervals. For three speakers, the non-linear smooth for LOT does not significantly differ from THOUGHT, while for CHI014 and CHI016, the non-linear smooths have an edf close to 1. For these speakers, LOT and THOUGHT are distinguished by protrusion, but both the degree of protrusion for THOUGHT and the degree of lip spreading for LOT are less extreme. **Figure 13** shows lip protrusion trajectories from GAMMs fit to all vowels for six speakers. Speakers in the top row, all of whom have clear LOT-THOUGHT rounding contrasts, produce THOUGHT with a maximum degree of protrusion comparable to GOOSE and GOAT, but with a distinct trajectory at the offset, while LOT shows comparable spreading to FLEECE and TRAP. For speakers in the bottom row, the trajectory for THOUGHT is variable. For CHI012 and CHI016, its shape is more similar to the unround vowels than to the other round vowels, even though THOUGHT is significantly more protruded throughout its entire duration. For CHI016, LOT is also less spread than TRAP, whereas the low unround vowels are comparable for the other speakers. Thus, while the magnitude of rounding/spreading gestures for LOT and THOUGHT may be reduced, rounding alone is responsible for maintaining a modest acoustic contrast, given that neither speaker distinguishes LOT and THOUGHT through tongue position.

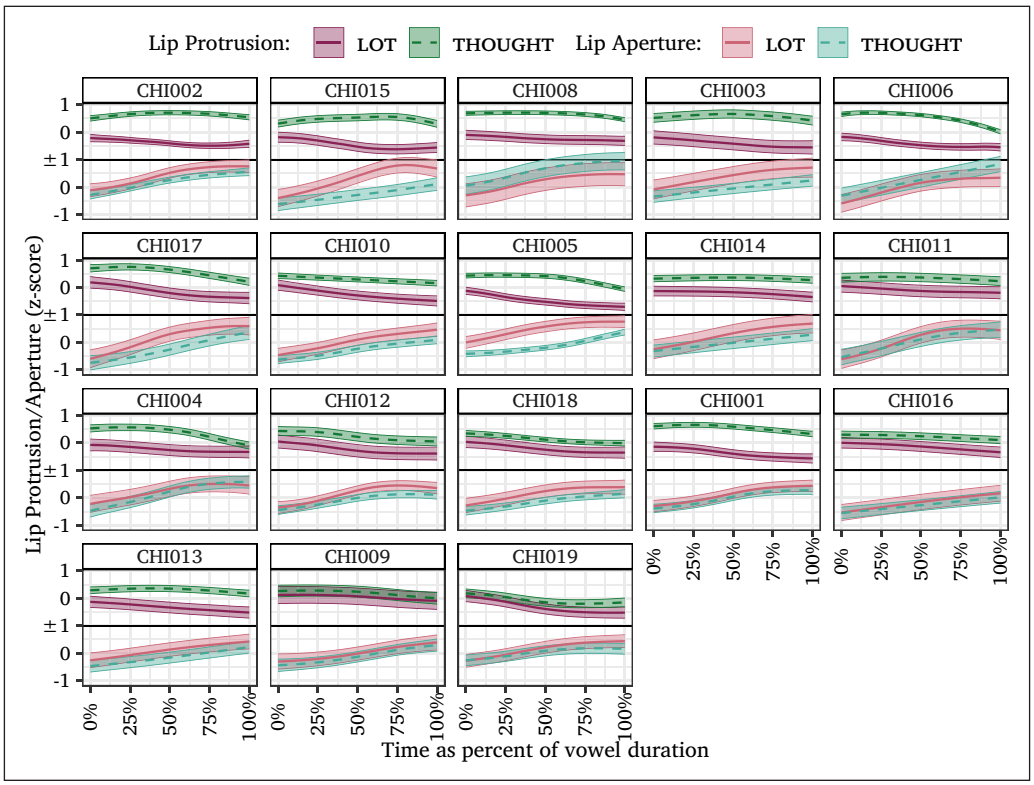


Figure 12: Predicted lower lip protrusion (upper) and lip aperture (lower) for LOT and THOUGHT, normal speech task. Shading represents 95% confidence interval.

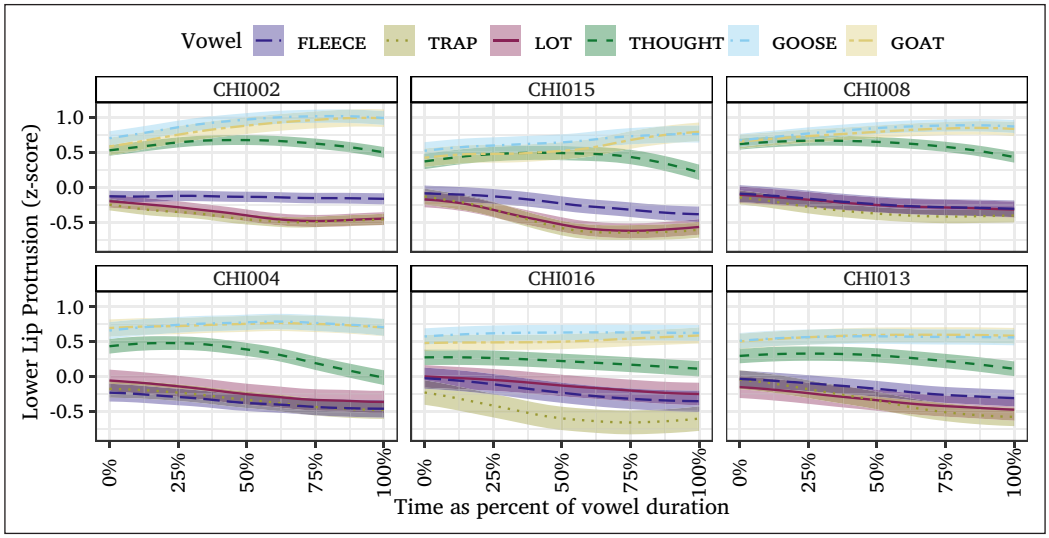


Figure 13: Predicted lower lip protrusion for all vowels. Shading represents 95% confidence interval.

Differences in lip aperture are also variable. For the majority of speakers, LOT has a significantly wider lip aperture than THOUGHT, although the intercept difference is consistently smaller than that for lip protrusion. Lip aperture trajectories also tend to be more similar, and the difference smooths for several speakers are not significant. Both LOT and THOUGHT tend to have a larger aperture toward the vowel offset. For CHI008 and CHI006, the lip aperture for THOUGHT is larger than for LOT, which is attributable to greater outrounding for THOUGHT that increases the vertical distance between the upper and lower lips.

2.5. Summary of Experiment 1

This experiment has shown that Chicagoans vary in the strength of the LOT-THOUGHT contrast, the articulatory gestures used to distinguish them, and in their use of NCS-like vowel targets more generally. Typical NCS vowel systems were observed for most older speakers, who exhibit raising of TRAP, fronting of LOT, and a robust LOT-THOUGHT contrast. A strong acoustic distinction between LOT and THOUGHT is found to be associated with differences in both tongue position and lip rounding, as was also observed in Michigan by Havenhill and Do (2018). For older speakers, LOT has a fronted tongue position and is unround, while THOUGHT has a lower, more retracted tongue position in addition to lip protrusion comparable to GOOSE and GOAT. Younger speakers, by contrast, are less likely to show an acoustically strong LOT-THOUGHT contrast and are more likely to show non-NCS features, including a lower TRAP and more fronted GOOSE. Acoustically-weak contrasts between LOT and THOUGHT are associated with a less-fronted LOT, which is distinguished from THOUGHT predominantly (or entirely) through lip rounding. Only one speaker was found to collapse the LOT-THOUGHT contrast completely, in both acoustics and

articulation. For the remaining speakers, the lingual difference between the vowels is either marginal or not significant. While lip protrusion for THOUGHT may be somewhat reduced in comparison to the other round vowels, the LOT-THOUGHT rounding contrast is maintained to some extent for all but one speaker.

3. Experiment 2: Clear speech enhancement of the LOT-THOUGHT contrast

In Experiment 1, it was found that some speakers produce an acoustically weak LOT-THOUGHT contrast, distinguishing the two vowels by lip rounding but not by tongue position, while others maintain a robust acoustic contrast and distinguish the two vowels by both backness and rounding. As articulatory effort is expected to vary according to speaking style, however, it remains an open question whether speakers will extend these same articulatory configurations to other styles. Theories of phonetic variability, in particular H&H Theory, posit that speech production reflects a compromise between the inherently conflicting goals of maximizing perceptibility and minimizing articulatory effort. If lip gestures alone are sufficient to maintain (audiovisual) perceptibility of the LOT-THOUGHT contrast, speakers may reduce the magnitude of other gestures in less careful speech styles. The finding that some speakers lack a lingual distinction between LOT and THOUGHT is therefore inconclusive—while it may be the case that these speakers have collapsed the lingual LOT-THOUGHT distinction altogether, it is also possible that they will recruit lingual gestures to increase acoustic distance between the two vowels in more careful styles, when doing so is necessary for clarity. At the same time, because speaking style is also sensitive to sociolinguistic factors (Clopper & Pierrehumbert, 2008; Clopper & Tamati, 2014; Clopper et al., 2017, 2019), a third possibility is that these speakers will not enhance a marginal LOT-THOUGHT contrast at all (Clopper et al., 2017; Grama & Kennedy, 2019; Labov, 1991; Nycz, 2013). With an apparent shift in Chicago toward non-NCS vowel systems (D'Onofrio & Benheim, 2020; McCarthy, 2011), speakers may target socially unmarked variants (Clopper & Pierrehumbert, 2008; Clopper et al., 2017), promoting a yet less-fronted LOT and/or weaker LOT-THOUGHT contrast.

How speakers with a stronger LOT-THOUGHT contrast produce the two vowels in hyperarticulated speech is also of interest, as the relative contributions of synergistic articulators (e.g., the tongue and lips) to vowel contrast enhancement has not generally been a focus of clear speech research. Acoustic dispersion between vowels differing in both backness and rounding could, in principle, be maximized by hyperarticulating both lingual and labial gestures. On the other hand, competing pressure to limit articulatory effort might restrict hyperarticulation only to those gestures that maximally contribute to perceptibility. That is, if hyperarticulation is listener focused, then speakers may prefer to hyperarticulate gestures for which a smaller articulatory change will have a greater influence on the acoustic output (e.g., due to quantal relations; Stevens 1989; Stevens and Keyser 2010) or which have other means of being perceived

(e.g., visually). Some gestures otherwise used in normal speech may accordingly be reduced, either to counterbalance effort expended to hyperarticulate more perceptible gestures, or to modulate the acoustic effects of hyperarticulation (e.g., avoiding socially marked variants or confusion with other sounds). Thus, for speakers with a strong LOT-THOUGHT contrast, will hyperarticulation of the contrast involve enhancement of both lingual and labial gestures, or will speakers preferentially enhance one over the other? These questions are addressed in Experiment 2, in which speakers were asked to explicitly contrast LOT-THOUGHT minimal pairs in clear speech. The experiment aims to examine a) whether speakers increase acoustic distance between LOT and THOUGHT (and whether this varies by strength of the contrast in normal speech) and b) whether enhancement (or reduction) of acoustic LOT-THOUGHT distance are the result of modifications to lip rounding, tongue position, or both.

3.1. Methods

Experiment 2 was conducted in the same sitting as Experiment 1, allowing articulatory data for both experiments to be directly compared. All participants in Experiment 1 also participated in Experiment 2, except CHI009, who was unable to participate due to time constraints. The wordlist was a subset of that used in Experiment 1, only including the target vowels LOT and THOUGHT. Participants were instructed to "speak clearly and with as much emphasis as possible, as though you are correcting someone who misheard you" (cf. Bradlow, 2002; Picheny et al., 1986). Two words containing each target vowel and two containing the contrasting vowel were embedded in the carrier phrase "I said $\underline{target_x}$ and $\underline{target_y}$, not $\underline{contrast_a}$ and $\underline{contrast_b}$." The items $target_x$ and $contrast_a$ were a minimal or near-minimal pair, as were $target_y$ and $contrast_b$. For example, with the word pairs *nod* and *sod* vs. *gnawed* and *sawed*, the participant would say: "I said *nod* and *sod*, not *gnawed* and *sawed*." The full list of phrases is provided in supplementary materials. To avoid prosodic differences, only $target_x$ and $contrast_a$ are analyzed. Items alternated such that each word appeared in all four positions, for a total of 36 phrases.

Data processing and analysis followed the same procedures as those described for Experiment 1. However, because Pillai scores are sensitive to sample size (Stanley & Sneller, 2023), scores for the normal speech context were re-calculated after dropping the third repetition of each word. Two tokens of each word for both tasks yielded 36 repetitions per vowel per task (144 tokens in total). As a result, the normal speech Pillai scores differ somewhat between experiments 1 and 2. The reduced token count yields slightly higher scores for most participants, although interspeaker differences in score are generally preserved. Because the aim of Experiment 2 is to examine acoustic and articulatory contrast in comparison with Experiment 1, this decision was made in order to more accurately estimate the difference in acoustic contrast between the two speaking styles.

Dynamic articulatory and acoustic trajectories were analyzed with an additional set of GAMMs. Models from Experiment 1 were expanded with the inclusion of two terms for clear

speech in addition to the main term for vowel class, described in more detail below. Rather than by-speaker models, models were fit for three groups of speakers categorized according to their Pillai score in Experiment 1. Speakers with a score above 0.7 were included in the high score group, speakers with a score below 0.3 were included in the low score group, and other speakers were included in the medium score group. As such, models also included random reference-difference smooths for each term by speaker. A set of by-speaker models were also fit in order to visualize formant trajectories and articulatory changes for individual speakers. For visualization of tongue contours, by-speaker models were fit with the same structure as in Experiment 1, where independent tensor product smooths were fit for each level of a combined factor for vowel × task. In addition to the same measures of tongue position and lip rounding used in Experiment 1, Experiment 2 also includes analysis of jaw height in clear vs. normal speech, given that it has potential to serve as an additional visual cue to vowel quality and also to take on socioindexical meaning (Pratt & D'Onofrio, 2017). As previously noted, the position of the short tendon, which attaches the genioglossus to the mental spine of the mandible, was tracked by the DeepLabCut model. Following orientation of the ultrasound images with respect to the occlusal plane, the vertical coordinate of this point was z-score normalized across all vowel tokens and used as a measure of jaw height. Visualization of the short tendon's position in all vowel contexts is provided for reference in supplementary materials.

3.2. Results

3.2.1. Acoustic enhancement

Pillai scores for all speakers in normal and clear speech tasks are given in **Figure 14**. While the majority (all but four) make a larger acoustic LOT-THOUGHT distinction in clear speech than in normal speech, interspeaker differences are apparent. For most speakers with normal speech Pillai scores greater than 0.7, between-task differences were modest, within the range -0.10 to $+0.05$. A paired two-sample t-test shows that Pillai scores in the clear speech task are only marginally higher overall ($t(16) = -1.89$, $p = 0.077$). Relatively large increases are observed for both speakers with the lowest Pillai scores, including CHI019 and CHI013, both of whom were suggested in Experiment 1 to exhibit an incomplete merger. CHI019 had a normal speech Pillai score of 0.148 when sampled at 35% of the vowel's duration; in clear speech, her Pillai score increases to 0.265 ($+0.117$). The Pillai score increase was even greater for CHI013, whose score increased to 0.420 ($+0.198$).

Two speakers with extreme changes also stand out. CHI011, who had a mid-range Pillai score of 0.717 for normal speech, showed a decrease in LOT-THOUGHT acoustic distance in clear speech, with a score of only 0.448 (-0.269). On the other hand, CHI014 had the greatest increase in score between the two tasks, from 0.680 in normal speech to 0.899 ($+0.219$) in clear speech. With these two speakers excluded, the overall between-task difference in score is significant ($t(14) = -3.56$, $p = 0.0031$).

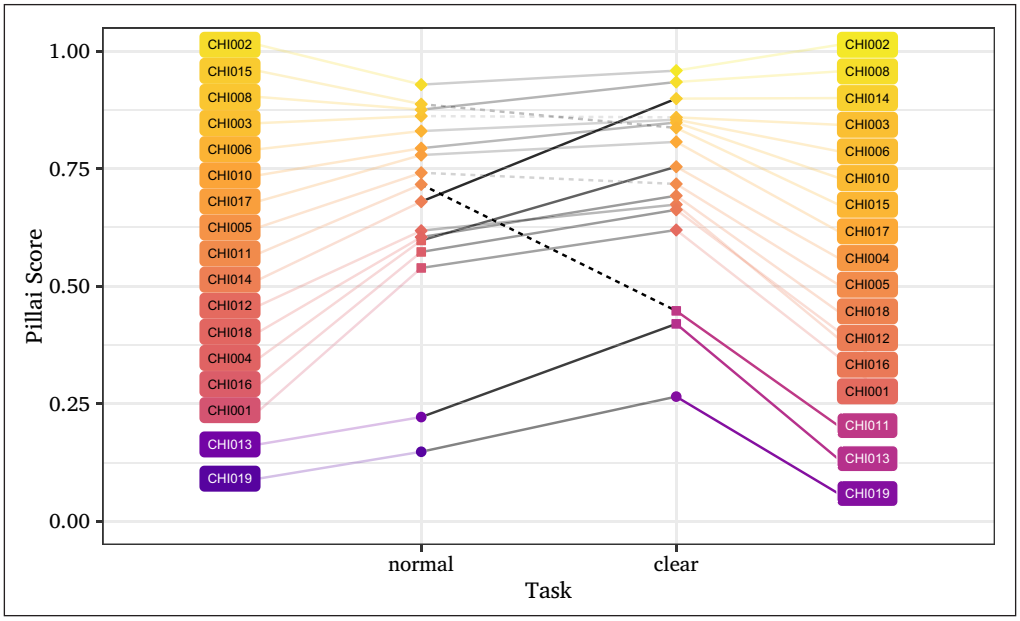


Figure 14: Pillai score by task. Solid line indicates increase in Pillai score for clear speech task, dashed line indicates decrease.

Summaries for group-level F1 and F2 GAMMs are given in **Table 7**. As CHI011's substantial decrease in Pillai score is clearly distinct from all other speakers, she was excluded from the Medium group prior to model fitting and is presented separately. Here, terms listed for the normal speech task indicate the constant and non-linear LOT-THOUGHT differences in F1 and F2 for each group. As in the ordered factor models in Experiment 1, the intercept/reference smooth corresponds to the predicted formant values for THOUGHT, while the term for LOT indicates the constant/non-linear difference of LOT relative to THOUGHT. Normal speech estimates for LOT and THOUGHT confirm that speakers in the High and Medium groups (Pillai scores above 0.3) are similar in terms of the F1 × F2 position of THOUGHT, as well as the constant F1 difference between THOUGHT and LOT. For speakers in the Medium group, the F2 difference between the two vowels is smaller than for speakers in the High group, consistent with the F2 variability for LOT observed in Experiment 1. For speakers in the Low contrast group, LOT and THOUGHT do not significantly differ from one another in either F1 or F2, although the non-linear trajectories for the two vowels show significant differences.

Of present interest are the differences between clear and normal speech, modeled with difference smooths. The first term, clear vs normal, is equal to 1 for all clear speech tokens (without specification for vowel category), which generally corresponds to the change from normal to clear speech for THOUGHT. The second term, LOT v THOUGHT, is equal to 1 for clear speech tokens of LOT alone and therefore corresponds to the difference in the effect of clear speech for LOT relative to THOUGHT. Two versions of each model were fit. The first included binary difference smooths

						Parametric coefficient				Difference smooth			
Group	**Task**	**Avg. Pillai**	**DV**	**Term**	**Bin. smooth *p***	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
High	normal	0.736	F1	THOUGHT (Int.)		0.788	0.0362	21.8	<0.001	23.3	24.6	70.9	<0.001
				LOT		0.442	0.0604	7.32	<0.001	8.47	10.9	3.77	<0.001
			F2	THOUGHT (Int.)		–0.648	0.0557	–11.6	<0.001	13.9	17	14.4	<0.001
				LOT		0.477	0.04	11.9	<0.001	7.08	9.02	9.06	<0.001
	clear	0.794	F1	clear vs. normal	<0.001	0.0887	0.0335	2.65	0.016	8.51	11	3.23	<0.001
				LOT vs. THOUGHT	**0.254**	0.183	0.0345	5.31	<0.001	**4.21**	**5.2**	**1.62**	**0.295**
			F2	clear vs. normal	<0.001	–0.0723	0.0296	–2.44	0.029	3.05	3.69	10.1	<0.001
				LOT vs. THOUGHT	**0.116**	**0.0677**	**0.0443**	**1.53**	**0.254**	**1.52**	**1.7**	**2.5**	**0.096**
Medium	normal	0.526	F1	THOUGHT (Int.)		0.776	0.072	10.8	<0.001	20.2	22.7	43.1	<0.001
				LOT		0.332	0.0611	5.44	<0.001	**1.01**	**1.02**	**0.035**	**>0.999**
			F2	THOUGHT (Int.)		–0.637	0.0541	–11.8	<0.001	11.6	14.9	13.2	<0.001
				LOT		0.234	0.0622	3.76	<0.001	4.36	5.41	4.02	0.002
	clear	0.637	F1	clear vs. normal	**0.131**	**0.0637**	**0.0888**	**0.717**	**0.946**	**4.12**	**5.09**	**2.11**	**0.126**
				LOT vs. THOUGHT	**0.505**	0.198	0.0602	3.29	0.002	**1.6**	**1.8**	**1.09**	**0.483**
			F2	clear vs. normal	<0.001	–0.117	0.0474	–2.47	0.027	3.25	3.96	5.96	<0.001
				LOT vs. THOUGHT	0.013	**0.109**	**0.0703**	**1.55**	**0.242**	2.91	3.58	3.4	0.017
Low	normal	0.185	F1	THOUGHT (Int.)		0.819	0.0994	8.24	<0.001	12.6	15.7	13.4	<0.001
				LOT		**0.0844**	**0.0956**	**0.883**	**0.754**	7.35	9.33	5.09	<0.001
			F2	THOUGHT (Int.)		–0.647	0.0566	–11.4	<0.001	10.2	13.1	12.6	<0.001
				LOT		**0.102**	**0.0491**	**2.08**	**0.076**	7	9.12	9.07	<0.001

(Contd.)

						Parametric coefficient				Difference smooth			
Group	**Task**	**Avg. Pillai**	**DV**	**Term**	**Bin. smooth *p***	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
	clear	0.343	F1	clear vs. normal	<0.001	**0.0881**	**0.0719**	**1.23**	**0.441**	7.15	9.11	5.7	<0.001
				LOT vs. THOUGHT	**0.161**	**0.153**	**0.115**	**1.33**	**0.370**	4.35	5.51	5.15	<0.001
			F2	clear vs. normal	0.033	–0.227	0.0637	–3.57	<0.001	**2.64**	**3.18**	**1.55**	**0.393**
				LOT vs. THOUGHT	<0.001	**0.154**	**0.0817**	**1.88**	**0.120**	5.72	7.38	5.86	<0.001
CHI011	normal	0.717	F1	THOUGHT (Int.)		0.624	0.0436	14.3	<0.001	19.5	22.6	43.6	<0.001
				LOT		0.273	0.0387	7.04	<0.001	**2.74**	**3.31**	**1.52**	**0.445**
			F2	THOUGHT (Int.)		–0.751	0.0606	–12.4	<0.001	15.6	18.9	10.7	<0.001
				LOT		0.156	0.0313	4.99	<0.001	5.01	6.37	4.16	<0.001
	clear	0.448	F1	clear vs. normal		**0.0515**	**0.0383**	**1.34**	**0.358**	1	1	6.91	0.019
				LOT vs. THOUGHT		**–0.0599**	**0.0538**	**–1.11**	**0.532**	**1**	**1**	**1.1**	**0.575**
			F2	clear vs. normal		–0.112	0.0351	–3.2	0.003	9.71	12.7	9.84	<0.001
				LOT vs. THOUGHT		**–0.00559**	**0.0393**	**–0.142**	**>0.999**	**2.56**	**3.34**	**0.783**	**0.971**

Table 7: GAMM summaries for F1 and F2, clear vs. normal speech. "Bin. smooth *p*" indicates significance of binary smooth fit without a separate parametric term.

without the corresponding parametric term, capturing both linear and non-linear differences. The *p*-value for this term is given under "Bin. smooth *p*" as a holistic test of clear vs. normal speech differences, i.e., whether inclusion of the difference smooth in the model is justified. A second set of models were then fit with parametric terms that correspond to the difference smooths, in order to assess whether differences (if any) correspond to the overall position of either vowel or to their non-linear trajectories. The *p*-values for these terms were Bonferroni-corrected for multiple comparisons. Absolute changes to the positions of both vowels are visualized in **Figure 15**, which includes kernel density estimates for LOT and THOUGHT in clear speech, along with the predicted formant trajectories for clear and normal speech from the by-speaker GAMMs.

For speakers with high Pillai scores, above 0.7, binary smooths for clear vs normal speech are significant for both F1 and F2, indicating significant changes to both measures. Estimates for the intercept differences show that changes to the overall vowel positions in clear speech is relatively small, only 0.09 z for F1 (i.e., lowering by approx. 13 Hz) and only −0.07 z for F2 (i.e., retraction by approx. −31 Hz). The non-linear differences are also significant, however, which are elucidated by the predicted trajectories in **Figure 15**. Most speakers with high Pillai scores produce THOUGHT with a longer overall trajectory, with greater change in F2 throughout

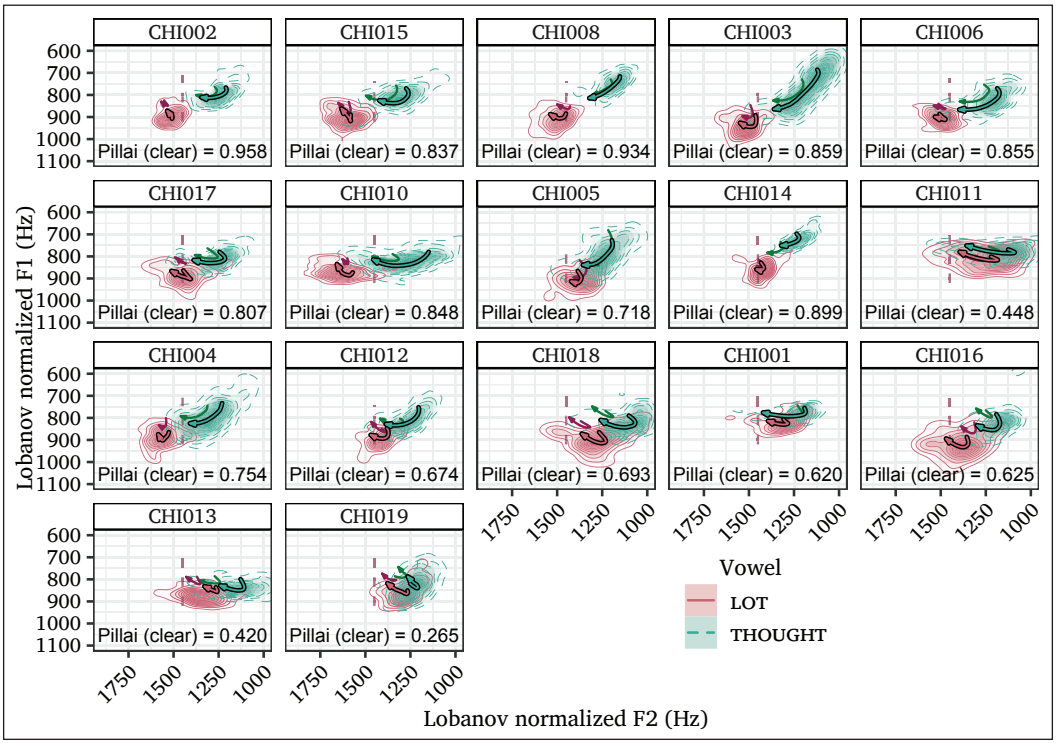


Figure 15: Kernel density estimates for LOT and THOUGHT in clear speech, sampled at five equidistant time points within 20–80% of the vowel's duration. Arrows indicate GAMM-predicted formant trajectories for clear (outlined) and normal speech.

its duration. For these speakers, the F2 minimum for THOUGHT is lower in clear speech than in normal speech during the early part of the vowel interval. In most cases, the maximum F2 at the vowel offset is similar for clear and normal speech. Thus, while the absolute position of THOUGHT is relatively stable, its F2 displacement is more extreme in clear speech. At the same time, binary smooths capturing the LOT-THOUGHT clear speech difference are not significant, which indicates that acoustic changes to both vowels are similar. That is, the overall distance between them is no larger or smaller than in normal speech, consistent with the relatively small increases in Pillai score. Inspection of the formant trajectories for LOT show that it also has a more dynamic formant trajectory for several speakers.

Speakers with normal speech Pillai scores between 0.3 and 0.7 (excluding CHI011) show significant differences in F2, but not in F1. Whereas speakers with high Pillai scores consistently showed an increase in the F1 of LOT, speakers with mid-range Pillai scores are less consistent in this regard. For CHI018 and CHI016, both vowels have a higher F1 in clear than in normal speech, but for CHI014 and CHI001, there is no F1 change. For F2, both the clear vs. normal and LOT-THOUGHT differences are significant. The parametric estimate for F2 reflects a larger overall decrease of −0.12 z (approx. −51 Hz) for THOUGHT. As the LOT-THOUGHT difference is also significant and has a positive (albeit non-significant) intercept estimate, this suggests that the decrease in F2 predominantly affects THOUGHT. Because LOT does not change to the same degree, Pillai scores increase more for speakers with moderate scores than for speakers with high scores.

CHI013 and CHI019, the only low-score speakers in Experiment 2, are qualitatively similar to one another in their acoustic clear speech differences. Both showed relatively large increases in Pillai score, reflected in significant binary difference smooths for clear vs. normal speech in both F1 and F2. The intercept difference is significant only for F2, which reflects an overall F2 decrease of −0.23 z (approx. −99 Hz). The non-significant intercept difference for LOT vs THOUGHT shows that this change predominantly affects THOUGHT; the non-linear F2 trajectories for LOT also differ in clear vs. normal speech, but the overall position of LOT does not. Again, this is consistent with both speakers' higher Pillai scores in clear than in normal speech. For CHI011, the only speaker to show a meaningful decrease in Pillai score, neither THOUGHT nor LOT significantly differs in F1 between the two tasks. There is a significant effect of clear speech on F2, however: both vowels have a lower F2 in clear speech than in normal speech, and both vowels show significantly larger F2 increases throughout the vowel intervals. This pattern is qualitatively distinct from most other speakers, for whom THOUGHT, but not LOT, showed a significant change in its F2 trajectory.

3.2.2. Lip rounding

Table 8 presents GAMM summaries for lip protrusion and aperture. Binary difference smooths for speakers with medium and high Pillai scores reveal significant changes in both rounding

						Parametric coefficient				Difference smooth			
Group	**Task**	**Avg. Pillai**	**DV**	**Term**	**Bin. smooth *p***	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
High	normal	0.736	LP	THOUGHT (Int.)		–0.337	0.0406	–8.3	<0.001	15.5	18.5	12.1	<0.001
				LOT		0.654	0.0407	16.1	<0.001	13.8	16.7	13	<0.001
			LA	THOUGHT (Int.)		**0.00722**	**0.0958**	**0.0754**	**>0.999**	12.2	15.1	5.68	<0.001
				LOT		**0.148**	**0.125**	**1.18**	**0.475**	6.85	8.32	3.12	0.003
	clear	0.794	LP	clear vs. normal	0.017	**–0.116**	**0.0556**	**–2.08**	**0.075**	4.59	5.46	2.82	0.024
				LOT vs. THOUGHT	0.005	0.223	0.0648	3.44	0.001	4.5	5.45	3.69	0.004
			LA	clear vs. normal	<0.001	**0.153**	**0.0927**	**1.65**	**0.198**	6.16	7.46	5.27	<0.001
				LOT vs. THOUGHT	**>0.999**	**–0.0993**	**0.0811**	**–1.22**	**0.442**	**1**	**1**	**0.103**	**>0.999**
Medium	normal	0.526	LP	THOUGHT (Int.)		–0.207	0.0445	–4.65	<0.001	9.69	12	8.4	<0.001
				LOT		0.404	0.0696	5.81	<0.001	6.9	8.45	10.3	<0.001
			LA	THOUGHT (Int.)		**–0.108**	**0.0731**	**–1.48**	**0.278**	10	12.4	7.2	<0.001
				LOT		0.128	0.0526	2.44	0.029	4.35	5.27	3.46	0.007
	clear	0.637	LP	clear vs. normal	0.024	**–0.179**	**0.0831**	**–2.15**	**0.063**	5.26	6.44	2.79	0.018
				LOT vs. THOUGHT	<0.001	0.357	0.0831	4.3	<0.001	5.5	6.86	4.93	<0.001
			LA	clear vs. normal	<0.001	0.21	0.0449	4.67	<0.001	4.9	5.94	4.04	0.001
				LOT vs. THOUGHT	**0.131**	**0.0792**	**0.0553**	**1.43**	**0.303**	**3.59**	**4.35**	**2.11**	**0.165**
Low	normal	0.185	LP	THOUGHT (Int.)		**–0.0842**	**0.0886**	**–0.951**	**0.684**	**4.39**	**5.29**	**1.58**	**0.308**
				LOT		0.336	0.121	2.77	0.011	5.76	7.15	7.25	<0.001
			LA	THOUGHT (Int.)		**–0.0775**	**0.0866**	**–0.895**	**0.741**	8.27	10.2	4.95	<0.001
				LOT		0.144	0.049	2.94	0.007	4.14	5.05	3.8	0.004

(Contd.)

						Parametric coefficient				Difference smooth			
Group	**Task**	**Avg. Pillai**	**DV**	**Term**	**Bin. smooth *p***	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
	clear	0.343	LP	clear vs. normal	0.040	–0.257	0.0977	–2.63	0.017	1	1	5.42	0.040
				LOT vs. THOUGHT	**0.550**	**0.301**	**0.227**	**1.33**	**0.370**	**2.03**	**2.31**	**1.35**	**0.525**
			LA	clear vs. normal	<0.001	0.224	0.0256	8.76	<0.001	5.05	6.11	5.52	<0.001
				LOT vs. THOUGHT	**0.206**	**–0.0171**	**0.0654**	**–0.262**	**>0.999**	**3.83**	**4.8**	**2.1**	**0.133**
CHI011	normal	0.717	LP	THOUGHT (Int.)		–0.252	0.0731	–3.45	0.001	7.92	9.91	16.8	<0.001
				LOT		0.369	0.0712	5.18	<0.001	6.48	8.3	9.11	<0.001
			LA	THOUGHT (Int.)		**–0.0279**	**0.111**	**–0.251**	**>0.999**	9.86	12.2	8.35	<0.001
				LOT		**–0.00101**	**0.0643**	**–0.0157**	**>0.999**	7.26	9.19	4.55	<0.001
	clear	0.448	LP	clear vs. normal		**–0.0998**	**0.0851**	**–1.17**	**0.482**	**2.28**	**2.95**	**2.39**	**0.137**
				LOT vs. THOUGHT		**–0.124**	**0.0825**	**–1.5**	**0.268**	1	1	8.06	0.009
			LA	clear vs. normal		**–0.103**	**0.0785**	**–1.31**	**0.381**	**1**	**1**	**3.69**	**0.109**
				LOT vs. THOUGHT		**0.089**	**0.0817**	**1.09**	**0.552**	**3.92**	**5.1**	**1.36**	**0.486**

Table 8: GAMM summaries for lower lip protrusion and lip aperture, clear vs. normal speech.

measures in clear speech. For both groups, the non-linear protrusion trajectories differ by task, although the parametric estimates fail to reach significance after Bonferroni correction, suggesting that, on the whole, lip protrusion for THOUGHT is not significantly greater in clear speech, but has a different trajectory. Trajectories for lip rounding and aperture in clear speech are visualized in **Figure 16**, while **Figure 17** provides difference smooths comparing clear to normal speech. Predictions are shown for both high and medium groups; the low group (CHI013 and CHI019) are visualized individually, as is CHI011. For high and medium groups, and especially CHI013, maximum lip protrusion for THOUGHT is significantly greater than maximum protrusion in normal speech. As indicated in **Figure 16**, peak protrusion is achieved closer to the vowel onset, which partially explains the non-significant intercept differences between conditions. As seen in **Figure 17**, the 95% confidence interval for a pairwise comparison of THOUGHT in clear vs. normal smooth excludes 0 for the first ~50% of the vowel interval, but the two conditions do not significantly differ during the later part of the vowel. Enhancement of lip protrusion therefore involves changes to both the magnitude and timing of the lip rounding gesture; peak lip protrusion is more extreme and a greater proportion of the gesture precedes the vowel onset as anticipatory coarticulation (Zellou & Chitoran, 2023).

The reverse pattern is observed for LOT, which shows more extreme lip spreading toward the vowel offset. Confidence intervals for pairwise clear vs. normal comparisons for LOT in **Figure 17** show that LOT exhibits significantly greater lip spreading in clear speech, at least for CHI013 and

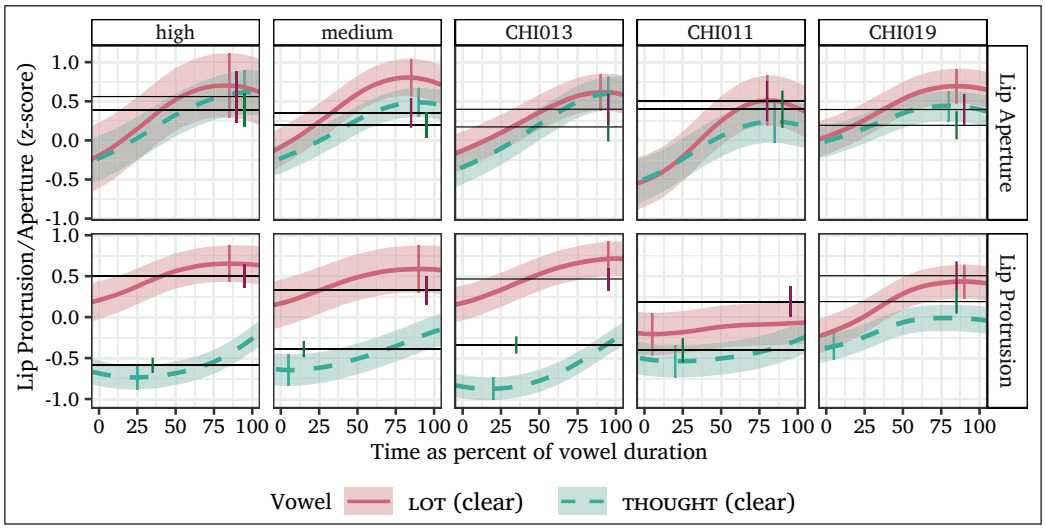


Figure 16: Lip protrusion and lip aperture for LOT and THOUGHT, clear speech task. Horizontal lines indicate degree of maximum protrusion/spreading/aperture in normal speech task; vertical segments indicate time of gestural maximum in each task.

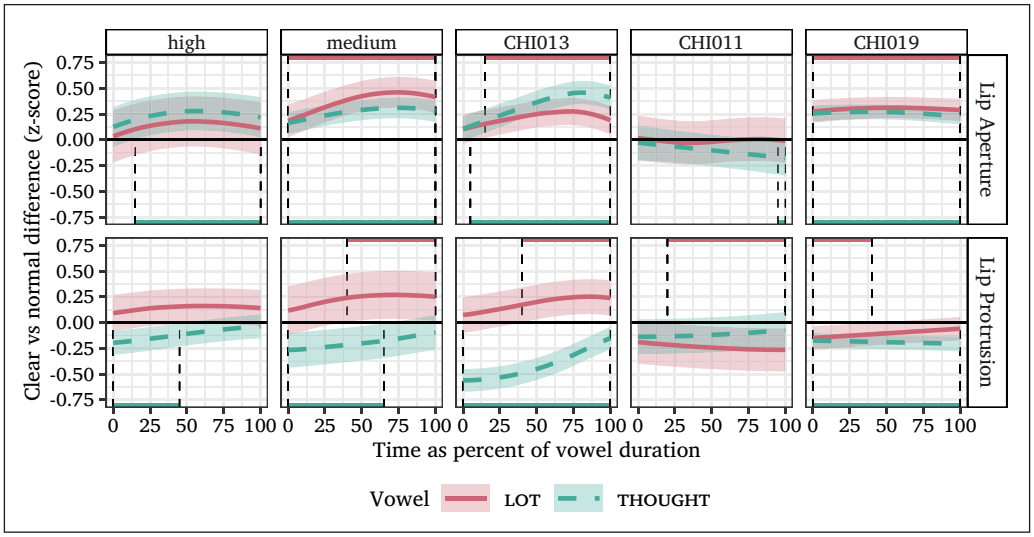


Figure 17: Difference in lip protrusion and lip aperture for LOT and THOUGHT, clear vs normal speech. Difference is significant when 95% confidence interval excludes zero, indicated at top and bottom of each panel.

medium-score speakers. As a result, the parametric and smoothing terms for LOT vs. THOUGHT protrusion are significant for both high-score and medium-score groups (**Table 8**), indicating that the overall between-vowel difference in lip protrusion is greater in clear than in normal speech. For CHI011 and CHI019, by contrast, the degree of lip spreading for LOT is significantly less in clear speech than in normal speech. For CHI019, this is partially offset by a global increase in the degree of protrusion for THOUGHT. As a result, she produces the two vowels more distinctly in clear speech, although the shape of the articulatory trajectories clearly differs from other speakers.

With respect to lip aperture, all three groups (but not CHI011) show task-related changes; significant binary smooths indicate that lip aperture is greater in clear than in normal speech, reflected by the significant difference smooths in **Figure 17**. For all speakers, maximum aperture occurs at vowel offset and this maximum is significantly higher, with the exceptions of CHI011 and high-score speakers' LOT. Just as aperture did not distinguish LOT and THOUGHT in normal speech, however, there is no contrast-specific enhancement to lip aperture for any group. Binary smooths for LOT vs THOUGHT are not significant (as also visualized in **Figure 16**), indicating that small between-vowel differences are fully captured by the main vowel term.

3.2.3. Tongue position

Figure 18 provides predicted mid-sagittal tongue contours for LOT and THOUGHT in both tasks for four speakers. Of these speakers, CHI003 and CHI008 were previously found to distinguish LOT and THOUGHT by tongue position in addition to lip rounding, which was associated with greater

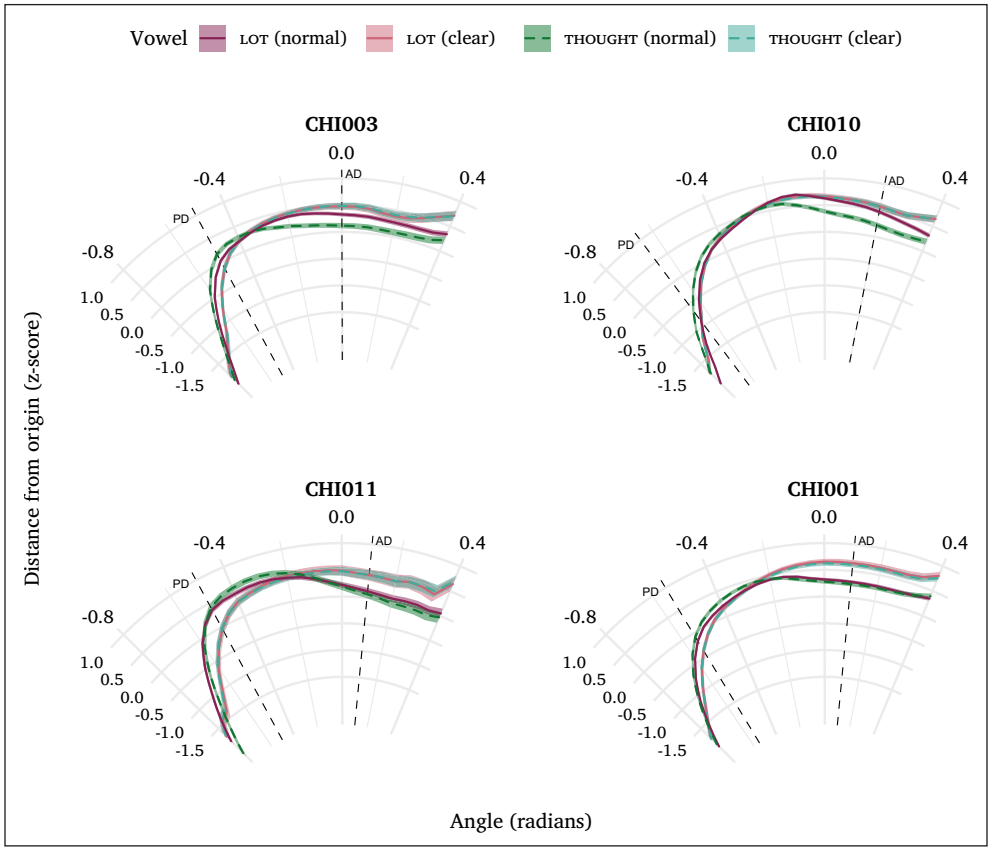


Figure 18: Predicted mid-sagittal GAMM tongue contours at 35% of vowel duration for four representative speakers, clear vs normal speech tasks. Shading represents 95% confidence interval. Tongue front is to the right.

acoustic distance between the two vowels. For both speakers, however, the lingual LOT-THOUGHT contrast is fully neutralized in clear speech. Tongue position for LOT is significantly higher and less retracted; given these speakers' NCS-shifted LOT, that change on its own may be expected to increase its acoustic distance from THOUGHT, especially as LOT-fronting has been argued to aid preservation of the low back contrast (Labov, 2019). That change is offset, however, by similar (yet more extreme) changes to tongue position for THOUGHT, which is also higher and fronter in clear speech, rather than backer and/or lower. As a result, the lingual distinction between LOT and THOUGHT collapses; the two vowels have identical clear-speech tongue positions for all speakers. CHI001 and CHI011 did not distinguish the vowels with significantly different tongue positions in normal speech, and do not recruit distinct tongue positions to enhance the contrast in clear speech. Rather, both vowels undergo parallel fronting and raising in clear speech.

Table 9 shows that this pattern holds for all speaker groups. For speakers with Pillai scores above 0.7, the main effect of vowel confirms a significant LOT-THOUGHT difference in both dorsum height ($\hat{\beta}$ 0.174, $p < 0.001$) and tongue body retraction ($\hat{\beta}$ −0.357, $p < 0.001$),

						Parametric coefficient				Difference smooth			
Group	**Task**	**Avg. Pillai**	**DV**	**Term**	**Bin. smooth *p***	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
High	normal	0.736	TDH	THOUGHT (Int.)		–0.392	0.0517	–7.57	<0.001	13.6	16.6	65.4	<0.001
				LOT		0.174	0.0419	4.15	<0.001	6.78	8.49	7.11	<0.001
			TBR	THOUGHT (Int.)		0.647	0.0753	8.59	<0.001	15	18.1	30.2	<0.001
				LOT		–0.357	0.0924	–3.86	<0.001	4.91	5.82	5.46	<0.001
	clear	0.794	TDH	clear vs. normal	<0.001	0.175	0.0584	2.99	0.006	14.3	17.7	68.3	<0.001
				LOT vs. THOUGHT	<0.001	–0.151	0.0264	–5.74	<0.001	6.09	7.81	9.64	<0.001
			TBR	clear vs. normal	<0.001	–0.318	0.0585	–5.42	<0.001	6.26	7.91	15.2	<0.001
				LOT vs. THOUGHT	0.004	0.23	0.0629	3.65	<0.001	3.98	5.05	6.67	<0.001
Medium	normal	0.526	TDH	THOUGHT (Int.)		–0.344	0.0426	–8.08	<0.001	11.5	14.5	72	<0.001
				LOT		**0.0762**	**0.0421**	**1.81**	**0.141**	**1.37**	**1.53**	**1.1**	**0.463**
			TBR	THOUGHT (Int.)		0.51	0.0769	6.63	<0.001	10.3	12.6	15.1	<0.001
				LOT		–0.168	0.0739	–2.27	0.046	**1.55**	**1.69**	**1.77**	**0.230**
	clear	0.637	TDH	clear vs. normal	<0.001	0.172	0.0355	4.85	<0.001	10.1	12.9	58	<0.001
				LOT vs. THOUGHT	**0.411**	–0.0691	0.0262	–2.64	0.017	**1**	**1**	**4.46**	**0.069**
			TBR	clear vs. normal	**0.211**	–0.231	0.0624	–3.7	<0.001	**1.76**	**2.06**	**1.87**	**0.290**
				LOT vs. THOUGHT	**0.101**	0.196	0.0732	2.68	0.015	**1**	**1**	**2.1**	**0.294**
Low	normal	0.185	TDH	THOUGHT (Int.)		**–0.199**	**0.137**	**–1.45**	**0.295**	8.75	11	25.7	<0.001
				LOT		**0.0138**	**0.0611**	**0.226**	**>0.999**	**1**	**1**	**0.834**	**0.722**
			TBR	THOUGHT (Int.)		0.657	0.15	4.38	<0.001	9.02	11	18.2	<0.001
				LOT		–0.277	0.115	–2.41	0.032	**2.33**	**2.64**	**0.987**	**>0.999**

(Contd.)

						Parametric coefficient				Difference smooth			
Group	**Task**	**Avg. Pillai**	**DV**	**Term**	**Bin. smooth *p***	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
	clear	0.343	TDH	clear vs. normal	<0.001	**–0.0178**	**0.0854**	**–0.209**	**>0.999**	6.81	8.82	22	<0.001
				LOT vs. THOUGHT	**0.297**	**0.012**	**0.054**	**0.221**	**>0.999**	**1**	**1**	**0.0528**	**>0.999**
			TBR	clear vs. normal	**0.926**	**–0.252**	**0.169**	**–1.49**	**0.270**	**3.02**	**3.73**	**1.73**	**0.278**
				LOT vs. THOUGHT	**>0.999**	**0.193**	**0.114**	**1.69**	**0.182**	**1.28**	**1.44**	**0.08**	**>0.999**
CHI011	normal	0.717	TDH	THOUGHT (Int.)		**–0.0861**	**0.0746**	**–1.15**	**0.497**	9.53	11.7	27.4	<0.001
				LOT		**–0.124**	**0.105**	**–1.18**	**0.478**	**1**	**1**	**0.321**	**>0.999**
			TBR	THOUGHT (Int.)		0.327	0.0985	3.32	0.002	12.1	14.9	36.1	<0.001
				LOT		**0.164**	**0.139**	**1.18**	**0.474**	**1**	**1**	**0.119**	**>0.999**
	clear	0.448	TDH	clear vs. normal		**–0.0737**	**0.0585**	**–1.26**	**0.416**	6.55	8.3	9.63	<0.001
				LOT vs. THOUGHT		**0.0813**	**0.0783**	**1.04**	**0.599**	**1.3**	**1.42**	**3.13**	**0.191**
			TBR	clear vs. normal		–0.248	0.107	–2.31	0.042	5.5	7.08	5.01	<0.001
				LOT vs. THOUGHT		**0.0234**	**0.151**	**0.155**	**>0.999**	**1**	**1**	**0.0884**	**>0.999**

Table 9: GAMM summaries for tongue dorsum height (TDH) and tongue body retraction (TBR), clear vs. normal speech.

such that LOT is fronter and higher than THOUGHT in articulatory space. In clear speech, both the (general) clear speech difference smooth and the LOT vs THOUGHT smooths are significant (Clear: $p < 0.001$, LOT-THOUGHT: $p < 0.001$), indicating that tongue positions differ for clear vs. normal speech and also that the effect of clear speech differs for LOT and THOUGHT. The parametric clear speech term for dorsum height indicates a significant increase in height in clear speech ($\hat{\beta}$ 0.175, $p = 0.003$), with effectively the same coefficient as the LOT-THOUGHT difference in normal speech. That is, the tongue position for THOUGHT raises in clear speech to the same height that is observed for LOT in normal speech. The parametric LOT-THOUGHT term for dorsum height shows a significant difference in the opposite direction ($\hat{\beta}$ −0.151, $p < 0.001$), i.e., tongue height for LOT is similar in both tasks, so that the height difference is neutralized.

The pattern for tongue body retraction is similar. In normal speech, the intercept LOT-THOUGHT retraction difference was significant ($\hat{\beta}$ −0.357, $p < 0.01$), but a significant clear speech change to THOUGHT ($\hat{\beta}$ −0.318, $p < 0.01$) collapses this distinction. Trajectories in **Figure 20** show that both vowels have a significantly higher dorsum in clear speech and that THOUGHT (but not LOT) has a significantly less retracted tongue body. The resulting lingual trajectories for LOT and THOUGHT are identical, as shown in **Figure 19**. For speakers with mid-range Pillai scores (and CHI013), the patterns are generally similar, although binary difference smooth terms are not significant for tongue body raising or for the clear speech LOT-THOUGHT difference in height. That is, these speakers do not distinguish LOT and THOUGHT by tongue body retraction in either task, and while they exhibit a significant clear speech change in tongue dorsum height, that change applies equally to both vowels.

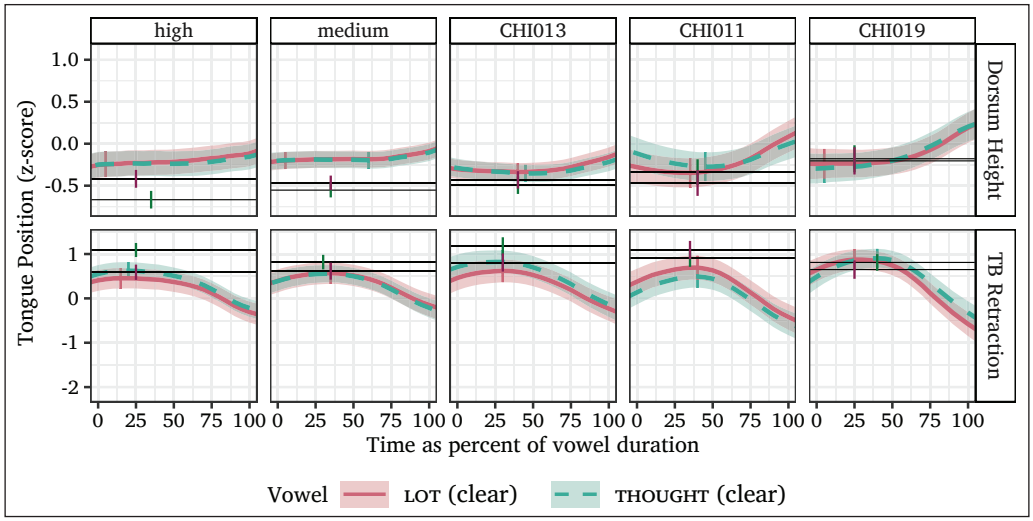


Figure 19: Tongue body retraction and tongue dorsum height for LOT and THOUGHT, clear speech task. Horizontal lines indicate degree of maximum height/retraction in normal speech task; vertical segments indicate time of gestural peak in each task.

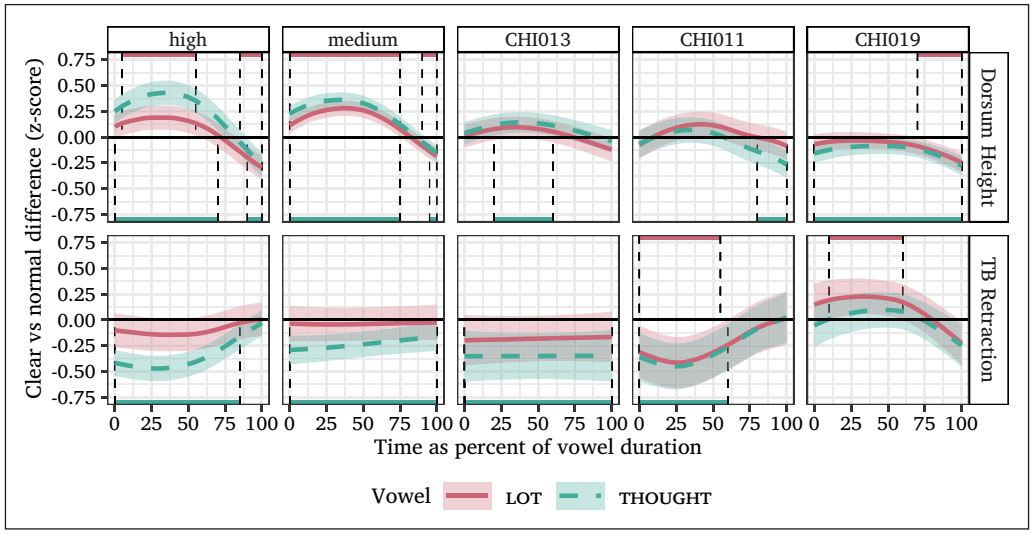


Figure 20: Difference in tongue body retraction and tongue dorsum height for LOT and THOUGHT, clear vs normal speech. Difference is significant when 95% confidence interval excludes zero, indicated at top and bottom of each panel.

3.2.4. Jaw height

Consistent with the raised tongue positions for both vowels in clear speech, jaw height also varies between the two tasks but does not distinguish LOT and THOUGHT. GAMM summaries for the height of the mandible and the short tendon are shown in **Table 10**, while the trajectories for both clear and normal speech are visualized in **Figure 21**. In normal speech, raising of the jaw for both vowels begins after 40–50% of the vowel's overall duration. In clear speech, jaw trajectories are less dynamic—the lowest position reached by the jaw is sustained throughout the vowel's entire duration. However, the lowest position achieved by the jaw tends to be higher, rather than lower, in clear vs normal speech. While jaw height has been argued to be a visually perceptible prosodic cue (Scarborough et al., 2009) and has potential to take on socioindexical meaning (Pratt & D'Onofrio, 2017), it does not distinguish LOT and THOUGHT in either task.

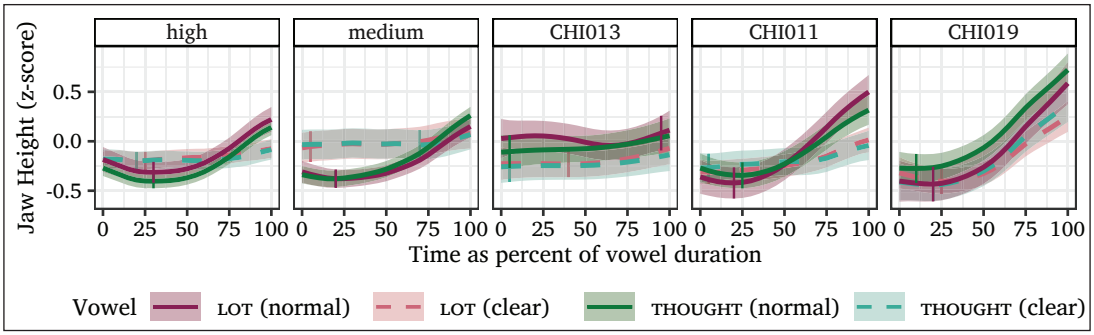


Figure 21: Predicted vertical position of the short tendon in clear and normal speech.

				Parametric coefficient				Difference smooth				
Group	**Task**	**Avg. Pillai**	**Term**	**Bin. smooth *p***	**Est.**	**SE**	***t***	**adj. *p***	**edf**	**Ref df**	**F**	**adj. *p***
High	normal	0.736	THOUGHT (Int.)		–0.212	0.0371	–5.71	<0.001	8.96	11	13.8	<0.001
			LOT		**0.0917**	**0.0425**	**2.16**	**0.062**	**1.72**	**2**	**0.58**	**>0.999**
	clear	0.794	clear vs. normal	<0.001	**0.0477**	**0.0491**	**0.971**	**0.663**	8.15	10.2	12.7	<0.001
			LOT vs. THOUGHT	**0.881**	–0.0784	0.0288	–2.72	0.013	**1**	**1**	**0.428**	**>0.999**
Medium	normal	0.526	THOUGHT (Int.)		–0.156	0.0367	–4.26	<0.001	8.62	10.7	26.3	<0.001
			LOT		**–0.0351**	**0.0378**	**–0.93**	**0.705**	4.05	5.06	3.18	0.014
	clear	0.637	clear vs. normal	<0.001	**0.151**	**0.0747**	**2.02**	**0.086**	8.03	10.2	21	<0.001
			LOT vs. THOUGHT	0.004	**0.0366**	**0.0296**	**1.23**	**0.435**	1	1	9.48	0.004
Low	normal	0.185	THOUGHT (Int.)		**0.0311**	**0.0575**	**0.54**	**>0.999**	2.99	3.38	6.91	<0.001
			LOT		**–0.0599**	**0.078**	**–0.768**	**0.885**	**1.5**	**1.78**	**0.6**	**>0.999**
	clear	0.343	clear vs. normal	**0.177**	–0.225	0.0583	–3.86	<0.001	2.3	2.71	3.9	0.028
			LOT vs. THOUGHT	**>0.999**	**0.0984**	**0.0608**	**1.62**	**0.211**	**1**	**1**	**0.0164**	**>0.999**
CHI011	normal	0.717	THOUGHT (Int.)		**–0.0943**	**0.0624**	**–1.51**	**0.261**	8.7	11.3	13.3	<0.001
			LOT		**0.0184**	**0.0873**	**0.211**	**>0.999**	1	1	6.02	0.028
	clear	0.448	clear vs. normal		**–0.0437**	**0.0515**	**–0.85**	**0.791**	6.94	8.97	11.8	<0.001
			LOT vs. THOUGHT		**–0.0312**	**0.0633**	**–0.493**	**>0.999**	**2**	**2.47**	**1.39**	**0.384**

Table 10: GAMM summaries for vertical position of the short tendon, clear vs. normal speech.

3.3. Summary of Experiment 2

Acoustic and articulatory differences for LOT vs THOUGHT in clear and normal speech are summarized in **Figure 22**. These figures show each speaker's predicted maximum acoustic difference (Euclidean distance) between LOT and THOUGHT in each task, which are compared to articulatory distances calculated in the same fashion, e.g., $\sqrt{\text{protrusion}^2 + \text{aperture}^2}$. For all speakers other than CHI011, LOT and THOUGHT show increased differentiation by lip rounding in clear speech, although the difference is negligible for CHI019. In general, this increase in rounding corresponds to a comparable increase in acoustic distance between LOT and THOUGHT. While CHI011 and CHI019 show some increase in the lingual LOT-THOUGHT distinction in clear speech, most other speakers show a decrease in lingual distance.

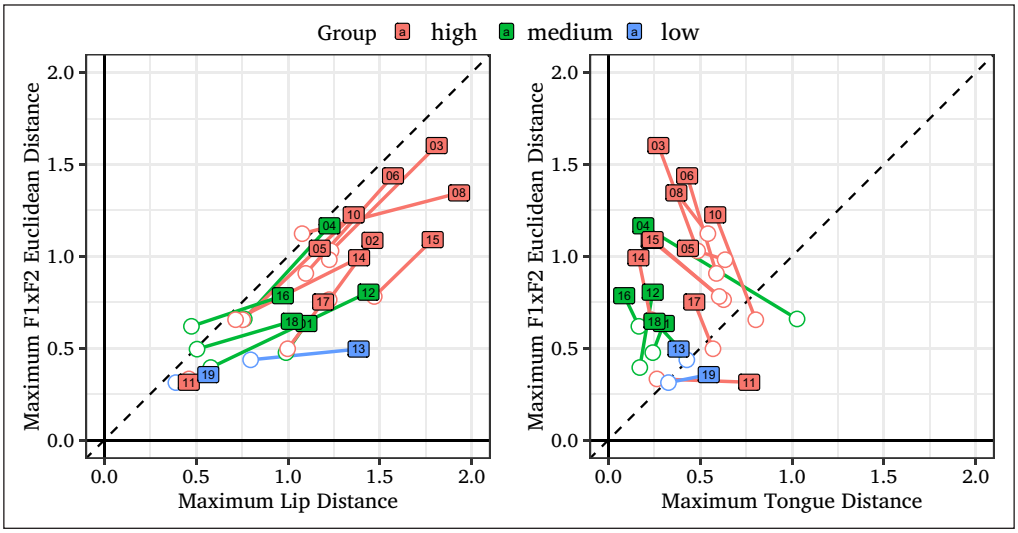


Figure 22: Maximum difference between LOT and THOUGHT in tongue position, lip rounding, and F1 × F2 Euclidean distance in clear vs. normal speech. Empty points indicate normal speech, filled speaker labels correspond to clear speech.

4. Experiment 3: Audiovisual perception of lip rounding

The combined results of experiments 1 and 2 suggest that Chicagoans are more likely to realize the LOT-THOUGHT contrast through differences in lip protrusion than through differences in tongue backness or height. Although speakers with the strongest LOT-THOUGHT contrast distinguish the vowels by tongue position in normal speech, this difference was found to be neutralized in clear speech. In normal speech, all speakers who retain the contrast produce the two vowels with significantly different degrees of lip protrusion, but not all speakers realize the vowels with distinct tongue positions. A third pattern, in which speakers maintain the contrast solely with a difference in tongue position, is hypothetically possible but was not observed. Experiment 3 was

conducted in order to test whether visual cues associated with lip rounding improve perceptibility of the LOT-THOUGHT contrast, as this may contribute to a bias for speakers to preserve and/or enhance labial contrasts.

4.1. Methods

4.1.1. Participants

All participants who took part in Experiment 1 also completed Experiment 3, including CHI009 (who did not complete Experiment 2). One participant (CHI003) started but was unable to complete the experiment due to an error that caused the software to irrecoverably crash.

4.1.2. Materials

The stimulus list for the perception experiment contains 120 items and is provided in the supplementary materials. Video recordings of 60 monosyllabic nonce words were created, including 10 words for each of the vowels LOT, THOUGHT, FLEECE, GOOSE, FACE, GOAT. As in the production experiment, the target vowels were LOT and THOUGHT, while the others served as fillers and controls. Nonce words were generated by finding combinations of phonotactically legal onsets and codas which did not form a real English word when any of the target or filler vowels were inserted and that rhymed with at least one real English word. Stimuli were produced by talkers raised in Metro Detroit who exhibit the Northern Cities Vowel Shift, who maintain an acoustic LOT-THOUGHT contrast, and who produce THOUGHT with lip rounding. Talkers were chosen from among the participants of the production experiment conducted by Havenhill and Do (2018) and therefore known to produce THOUGHT with visibly round lips. One of the talkers also served as the talker for the perception experiment presented in that study, but a new set of stimuli were recorded. Four talkers (2 men, 2 women) were included to control for intertalker differences that might have an influence on visual integration (Gagné et al., 1994; Kricos, 1996; McGuire & Babel, 2012; Traunmüller & Öhrström, 2007), yielding 480 stimuli in total.

Video was recorded at a resolution of 1920 × 1080 pixels at 120 fps using a Sony RX10-III digital camera. Audio was simultaneously recorded with an AKG C-417L lavaliere microphone and a Focusrite Scarlett 2i2 USB audio interface. Audio for each nonce word was extracted using Praat. Pink noise was added at a +15 dB signal-to-noise ratio and the mean amplitude of each stimulus was scaled to 70 dB. Each audio recording was then paired with one of two video recordings: the original, congruous video, and video that was incongruous in lip rounding (for target items) or in height (for control items). For target items, recordings of words containing LOT and THOUGHT were mismatched to produce round and unround variants of each vowel. For control items, recordings of the vowel pairs FLEECE-FACE and GOOSE-GOAT were mismatched to produce visually high and mid variants of each vowel. Video editing was performed using

the command line tool `ffmpeg` (FFmpeg Developers, 2018). When necessary, duration of the video was scaled (on a segment-by-segment basis) to match that of the incongruent audio. The `minterpolate` filter was used to smoothly interpolate between frames when increasing the duration of the video. Video for each talker was cropped such that the apparent size of the talker's head was consistent across talkers, and the position of the talker's mouth was centered at the lower third of the frame. After editing was complete, the stimuli were downsampled to a resolution of 1280 × 720 pixels and a frame rate of 60 fps for presentation.

Prior to running the experiment, stimuli were verified for naturalness by two independent raters who were naïve to the purpose of the experiment. Stimuli that were flagged by one or both raters were checked for issues and, when necessary, manually re-aligned or replaced with video from another take. The new set of stimuli was then re-checked in the same manner by a single rater, after which none of the stimuli were identified as problematic.

4.1.3. Procedure

Data for Experiment 3 were collected in the same session as experiments 1 and 2, with a short break between tasks. Experiment 3 was completed after experiments 1 and 2 in order to avoid influence from the talkers' speech on participants' production patterns.

Participants were seated in a sound-attenuated booth approximately one meter away from a 27 inch computer monitor, with video presented approximately at eye level. Audio was presented to participants through AKG K701 headphones. Stimuli were presented in pseudo-random order with PsychoPy (Peirce, 2007). The randomized stimulus list was generated such that no two stimuli containing the same vowel were presented in successive order, nor were two stimuli containing both members of a vowel pair (i.e., FACE stimuli were not followed by FLEECE, GOAT stimuli were not followed by GOOSE, LOT stimuli were not followed by THOUGHT, and vice versa). The stimuli for each talker were presented in separate blocks, with block order randomized by participant. Participants were given the opportunity to take a break of up to one minute between blocks.

The experimental design is presented schematically in **Figure 23**. Following Havenhill and Do (2018), participants identified the perceived vowel by selecting a rhyming word of English from one of two choices. Both words were shown simultaneously, immediately following the end of the stimulus, with the on-screen order (left vs. right) randomized for each trial. A 2000 millisecond time limit was imposed on responses, after which the experiment automatically advanced. Participants selected their response by pressing a colored button on a Cedrus RB-30 response pad, which also recorded their response time relative to the appearance of the choices on screen. Participants completed five practice trials (using real words rather than nonce words) at the beginning of the experiment.

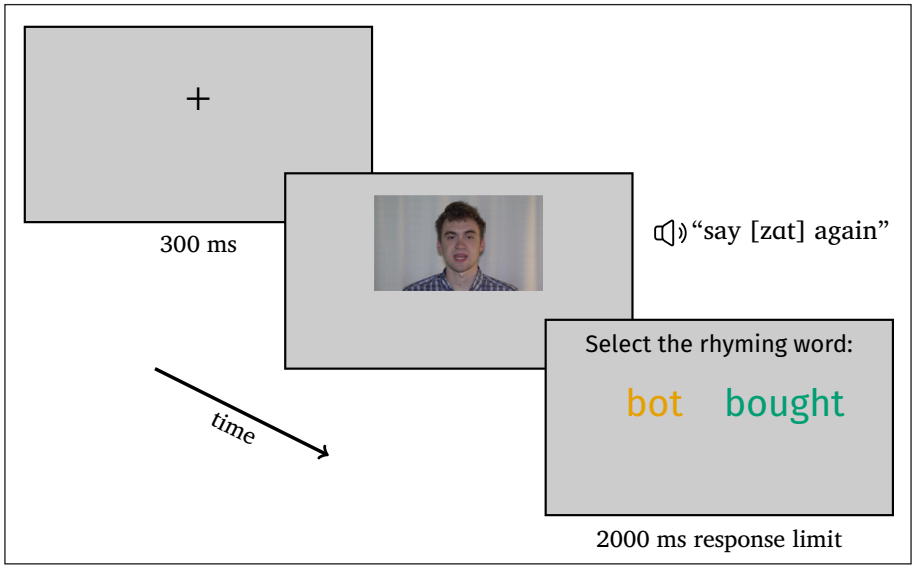


Figure 23: Perception experiment design.

4.2. Results

Figure 24a shows results for the control stimuli, which were visually (in)congruous in terms of vowel height. That is, auditory FLEECE was presented either with the original (visually congruous) video or with visually incongruous video from its mid counterpart FACE, while auditory GOOSE was shown with either GOOSE or GOAT video. Auditory FACE and GOAT were likewise shown with either visually congruous mid or visually incongruous high (FLEECE, GOOSE) video. For the high vowels, FLEECE and GOOSE, participants identified the stimulus as high in 97.1% of trials. A two-sample t-test for each vowel shows that there is no significant difference between visually high and visually mid stimuli for either FLEECE ($t(23.961) = 0.398$, $p = 0.694$) or GOOSE ($t(30.494) = 0.147$, $p = 0.884$). FACE was identified as high in 5.8% of trials, which is higher than expected. These responses come mostly from a single participant, CHI015, who identified auditorily mid/visually high tokens of FACE as FLEECE in 37% of trials. However, as with the high vowels, the effect of visual incongruity was not significant ($t(23.299) = 0.77$, $p = 0.449$). Finally, GOAT was identified as high in 2.8% of trials, with no significant difference between congruous and incongruous stimuli ($t(29.97) = –0.599$, $p = 0.554$). In sum, audiovisual incongruity did not have a significant effect on participants' perception of vowel height.

A significant effect of visual rounding cues was observed, however, for identification of the target LOT-THOUGHT contrast. **Figure 24b** presents results for participants who produced a medium or high acoustic LOT-THOUGHT contrast in Experiment 1. For these participants, auditory LOT was correctly identified as LOT in 74.2% of trials when presented with visually unround lips. Perception of auditory THOUGHT was substantially less accurate; participants in

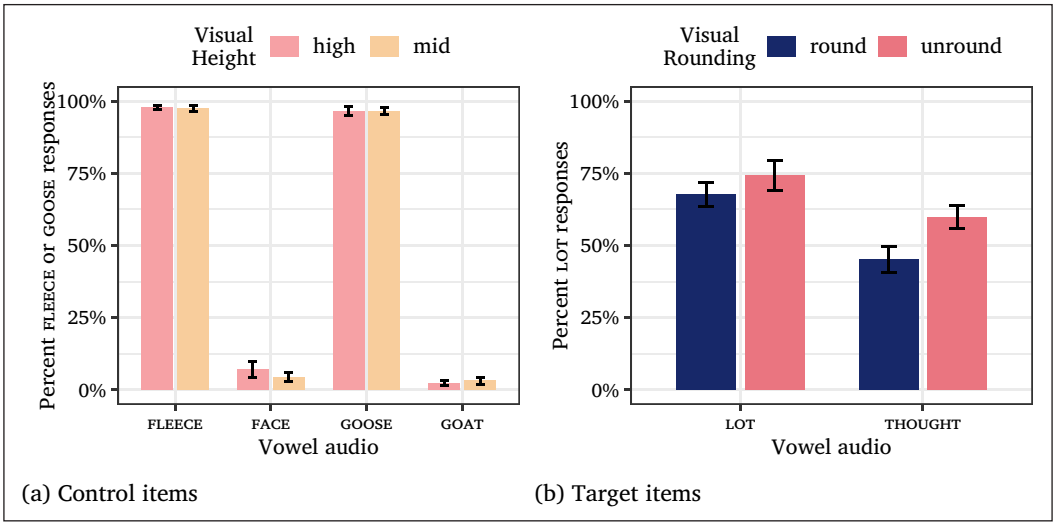


Figure 24: Perception results for control and target items. For target items, responses are shown for participants (N = 14) who produce medium or high acoustic contrast between LOT and THOUGHT. Error bars indicate standard error.

this group identified auditory THOUGHT as LOT in 45.2% of trials when presented with congruous lip rounding, such that accuracy was just better than chance. When paired with incongruous video of unround lips, auditory THOUGHT was much more likely to be perceived as LOT (59.8% of trials). The effect of incongruity for LOT is comparatively weak: 67.6% of auditory LOT stimuli were perceived as LOT even when presented with visibly round lips.

For participants who produced an acoustically weak contrast between LOT and THOUGHT in Experiment 1, perception results are presented in **Figure 25**. For CHI009, who exhibits complete merger of LOT and THOUGHT, perception of LOT is at chance: 52.5% when presented with congruous unround lips and 50.0% when presented with incongruous lip rounding. CHI009's perception of THOUGHT was also close to chance, but showed a small effect of visual incongruity in the direction opposite to what was predicted. CHI009 heard auditory THOUGHT stimuli as LOT in 47.5% in trials with congruous rounding, but in only 40.0% of trials with incongruous unrounding. This result suggests that this participant is unable to reliably distinguish between the two vowels in perception, and that his ability to perceive the LOT-THOUGHT contrast is not aided by visible lip gestures. While CHI019 was also close to chance in her perception of both vowels, she showed the predicted effect of visual cues. When paired with video of unround LOT, she responded with LOT in 65.8% of auditory LOT trials and 57.9% of auditory THOUGHT trials. This result is consistent with this speaker's production of a significant rounding distinction between LOT and THOUGHT in Experiment 1, even though her acoustic contrast was weak and she only marginally increased the lip rounding contrast in Experiment 2.

A mixed effects logistic regression model (**Table 11**) was run for all participants who contrast THOUGHT from LOT with lip rounding in their own speech production, i.e., all participants except

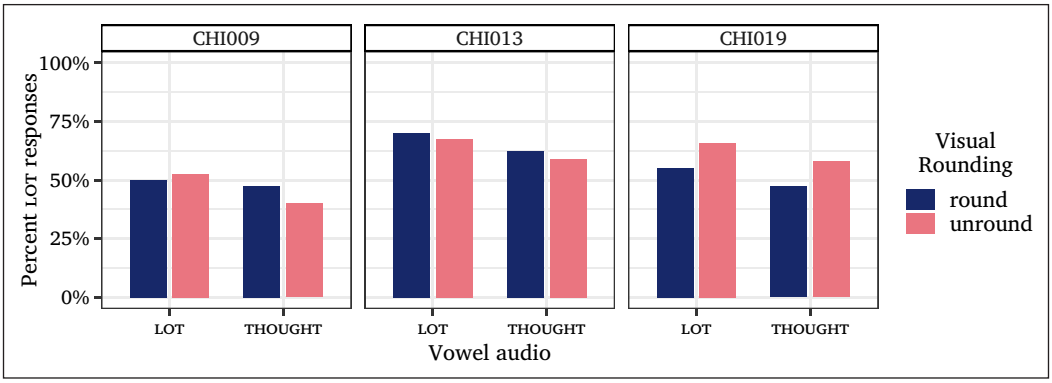


Figure 25: Perception results for participants ($N = 3$) who produce weak acoustic contrast between LOT and THOUGHT.

Predictor	Estimate	SE	z value	Pr(>|z|)	
(Intercept)	0.550	0.231	2.379	p = 0.017	*
Vowel Audio					
/ɔ/	–0.416	0.099	–4.207	p < 0.001	***
Visual Congruity					
Incongruous	0.071	0.099	0.717	p = 0.473	
Audio * Congruity					
/ɔ/ * Incongruous	0.227	0.099	2.297	p = 0.022	*

Table 11: Mixed effects logistic regression model for participants ($N = 16$) who produce significant /ɑ/-/ɔ/ contrast in tongue position, lip rounding, or both.

CHI009. Fixed effects included auditory vowel class, visual congruity, and their interaction. By-participant random slopes for vowel class and visual congruity were included, as were random intercepts for talker and item. A significant negative coefficient for auditory vowel class indicates that stimuli containing auditory THOUGHT are less likely to be perceived as LOT than stimuli containing auditory LOT. The non-significant main effect of incongruity confirms that LOT is not perceived as THOUGHT when produced with visually round lips. A significant interaction of vowel class and visual congruity shows, however, that THOUGHT is more likely to be perceived as LOT when produced with unround lips.

Figure 26 shows response times by vowel class and visual congruity. Response times for LOT and THOUGHT were significantly longer than for any other vowel. A linear mixed effects model (**Table 12**) fit to log response time, with random intercepts for participant, talker, and item, shows that vowel class significantly predicts response time ($\chi^2(5)$, $p < 0.001$). Pairwise post hoc comparisons show that response times for each of LOT and THOUGHT is significantly longer than for all non-low vowels ($p < 0.001$), but that LOT and THOUGHT do not significantly

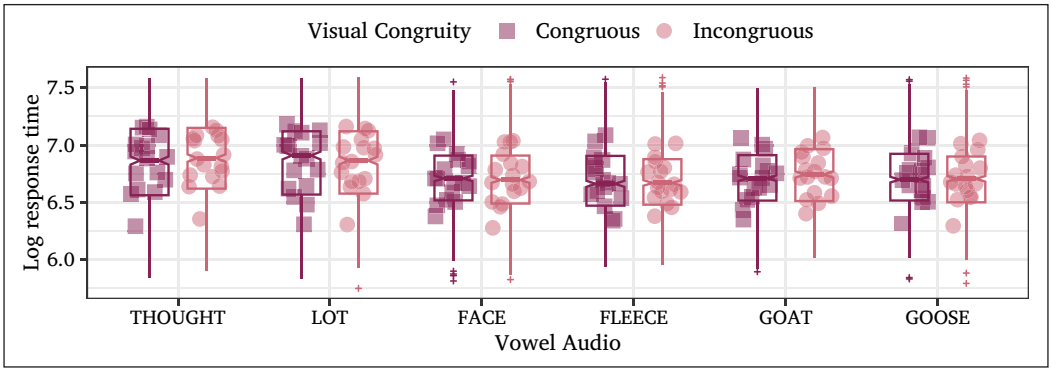


Figure 26: Response time by vowel and visual congruity. Points represent individual means.

Predictor	Estimate	SE	t value	Pr(>|t|)	
(Intercept)	6.865	0.052	130.801	$p < 0.001$	***
Vowel Audio					
COT	–0.016	0.022	–0.756	$p = 0.451$	
FACE	–0.148	0.022	–6.897	$p < 0.001$	***
FLEECE	–0.173	0.022	–8.038	$p < 0.001$	***
GOAT	–0.133	0.022	–6.204	$p < 0.001$	***
GOOSE	–0.147	0.022	–6.837	$p < 0.001$	***

Table 12: Linear mixed effects regression model for response time (all participants).

differ from each other ($p = 0.9747$), nor do any of the non-low vowels differ from one another. Inclusion of visual congruity as a predictor does not improve model fit ($\chi^2(1)$, $p = 0.628$), indicating that response time does not differ for stimuli presented with video that is incongruous in terms of rounding (LOT, THOUGHT) or height (all other vowels).

4.3. Summary of Experiment 3

Experiment 3 confirms that visual rounding cues influence listener identification of LOT and THOUGHT. Stimuli containing auditory THOUGHT were significantly more likely to be perceived as LOT when presented with unround lips. As THOUGHT in many cases exhibits significant acoustic overlap with LOT, visual perception of lip rounding may be one of the primary means by which listeners distinguish LOT and THOUGHT. If so, perceptual benefits conferred by lip rounding may contribute to a bias toward visually distinct articulatory strategies and allow for greater spectral overlap between LOT and THOUGHT without complete collapse of the contrast. In general, listeners were found to exhibit a bias toward perceiving LOT when presented with auditory THOUGHT stimuli, regardless of the visual stimulus. A nearly identical bias toward LOT was also seen by

Havenhill and Do (2018), who tested participants on different recordings of the same nonce words used here. In that study, participants identified auditory LOT stimuli as LOT in 65–75% of trials regardless of visual lip rounding cues. Even when THOUGHT was presented with congruous rounding, participants identified the vowel as LOT in over 35% of trials, increasing to 55% of trials when presented with incongruous unrounding. A contributing factor for this bias is likely frequency: Mines et al. (1978) show that the token frequency of LOT in conversational American English is approximately twice that of THOUGHT. As the words presented to participants were nonce words, a frequency-driven bias toward LOT is likely stronger than what would be observed for identification of real THOUGHT words. It may also be the case that for auditory LOT tokens with a relatively high F2, visible rounding is insufficient to override the auditorily perceived frontness of the vowel. For THOUGHT, which is acoustically more back, the auditory cues are inherently more ambiguous. The results of Experiment 1 show that THOUGHT rarely, if ever, exhibits high F2 values, while LOT ranges from a low back to low central vowel. A low vowel with an F2 around 1200–1300 Hz may thus correspond either to THOUGHT or to a back token of LOT; in that case, visual cues (and in real speech, information obtained from discourse and syntactic context) may determine what listeners perceive. As this experiment tested only audiovisual perception and did not include audio-only or visual-only tasks, it may also be that case that listeners will show greater reliance on visual cues under noisier conditions, including for LOT. Future work may further explore the trading relationship between auditory and visual perception by varying the amount of auditory noise, visibility of the lips, and acoustic quality of the vowel.

5. General Discussion

This study has examined production and perception of the LOT-THOUGHT contrast among speakers from Chicago, where LOT and THOUGHT are in flux—first due to the spread of the Northern Cities Shift, and then by its apparent ongoing reversal. Contrary to most other varieties of North American English, phonemic contrast between LOT and THOUGHT has historically been retained in Chicago. As younger generations reverse the trends of the NCS and show increasing orientation toward exogenous speech norms, however, the ongoing status of the LOT-THOUGHT contrast remains to be seen. Retraction of LOT results in its merger with THOUGHT in varieties where the Low-Back-Merger Shift has taken hold, but it is not certain this will occur in regions where the initial configurations of LOT and THOUGHT are different (D'Onofrio & Benheim, 2020; Nesbitt & Stanford, 2021; Nesbitt et al., 2019). Labov (2019) identifies the LOT-THOUGHT rounding contrast as a key differentiator of unmerged from merged varieties, although the articulatory characteristics of LOT and THOUGHT are not typically examined (cf. Havenhill & Do, 2018; Majors & Gordon, 2008).

Experiment 3 found that listeners are sensitive to the presence of visible rounding cues in the identification of THOUGHT. Participants were generally slower and less accurate in identifying

LOT-THOUGHT nonce words, consistent with the somewhat marginal status of the contrast. Nevertheless, auditory THOUGHT was significantly more likely to be identified as THOUGHT when presented to listeners with visible rounding. Visually unround variants of THOUGHT were more likely to be (mis)identified as LOT, which was also reported for Michigan listeners by Havenhill and Do (2018). Previous work has argued that as spectral overlap between vowels increases, speakers may rely on cues other than F1 and F2 to preserve phonemic contrasts. Labov and Baranowski (2006) proposed duration to be a feature by which Northern Cities-shifted speakers distinguish LOT from DRESS (another vowel implicated in the shift), while dynamic formant trajectories have received similar attention elsewhere (Farrington et al., 2018; R. A. Fox & Jacewicz, 2009). For LOT and THOUGHT specifically, Fridland et al. (2014) argue that greater spectral overlap corresponds to an increased difference in duration. While duration was not found to distinguish LOT and THOUGHT here (nor did Fridland et al. observe duration differences among their Northern speakers, of whom only one had LOT-THOUGHT merger), visual perception of lip rounding may serve a similar role. The relevance of visual perception to the preservation of contrast has previously been proposed for English /θ/-/f/ and /ɹ/-/w/, which may be misperceived due to their acoustic similarities but have visually distinct articulations (King & Chitoran, 2022; McGuire & Babel, 2012). If the same is true for LOT-THOUGHT, weakening of the acoustic distinction between them does not entail that they will merge, as long as listeners are able to categorize the vowel classes on the basis of visual cues (Labov, 1991).

Under usage-based models of perception and production (Bybee, 2000; Johnson, 1997; Pierrehumbert, 2001), a reliance on visual rounding cues for the identification of LOT and THOUGHT would in turn be predicted to influence speakers' production patterns. One approach to contrastive hyperarticulation proposes that a bias toward hyperarticulated or more contrastive productions emerges due to "perceptual restructuring." That is, perceptually salient tokens are more likely than perceptually ambiguous tokens to be accurately perceived and categorized, so production targets drawn from this pool will resemble the more easily perceived hyperarticulated variants. In Experiment 1, the strength of the LOT-THOUGHT contrast was found to vary according to how NCS-like the speaker's vowel system was. Speakers who produce LOT and THOUGHT with greater acoustic distance also have more NCS-fronted LOT, which is distinguished from THOUGHT by both tongue position and lip rounding. Among younger speakers, LOT is retracted with an F2 in proximity to (or overlapping with) THOUGHT. The articulatory strategy used to produce retracted LOT cannot be predicted on a purely acoustic basis, as lowering the F2 of a fronted, unround vowel may involve introducing rounding, retracting the tongue, or both. If the categorization of LOT and THOUGHT is mediated by the visibility of lip rounding, as suggested by Experiment 3, then listeners may correctly categorize acoustically-ambiguous tokens of THOUGHT more often when they are visually perceived to be round (and LOT when perceived to be unround). LOT-THOUGHT tokens that are distinguished by rounding are then more likely to serve as the basis for

the listener-turned-speaker's own production, motivating preservation of the rounding contrast. Consistent with this prediction, Experiment 1 shows that speakers with an acoustically weak LOT-THOUGHT contrast still distinguish the vowels by lip rounding; clear lingual contrasts are observed only among speakers with fronted LOT. Hyperarticulation of rounding for all but one speaker in Experiment 2 also suggests that the LOT-THOUGHT rounding contrast is not fully collapsed. The rounding distinction may, nevertheless, be somewhat diminished for speakers with weak contrasts; this appears to be the result of decreased rounding on THOUGHT, rather than rounding of LOT, but merits further investigation. As rounding of LOT has been proposed to facilitate its merger with THOUGHT in other regions (Labov, 2019), comparison of LOT-THOUGHT articulation in merged and unmerged regions may elucidate whether Chicagoans with weak contrasts resemble speakers who lack the contrast altogether.

Visual perceptibility may play a similar role in the ongoing fronting of the back vowels /u, ʊ, o/, which is observed in many varieties of English (e.g., Ferragne & Pellegrino, 2010; Harrington et al., 2008; Labov, 2008). In the case of /u/, fronting is caused not by pressure for this vowel to become more distinct from any other, but the opposite: the coarticulatory influence of a neighboring coronal consonant interferes with the acoustic realization of /u/, causing it to become *less* distinct from /i/ (Harrington et al., 2008; Ohala, 1993). For vowel systems in which /u/ undergoes acoustic fronting, realizing the auditorily optimal [u] is not viable; speakers must instead choose one of [y], [ɨ], [ʉ], or [ɯ] as their production target. In this situation, articulatory dispersion of the sort rejected by Diehl and Kluender (1989) may come into play. Whereas [ɨ] and [ɯ] may yield acoustically equivalent output, only variants like [ʉ] and [y] maintain visible articulatory dispersion from /i/, while [ɯ] and [ɨ] do not. Although some impressionistic observations describe fronted /u/ as unrounded (Hagiwara, 1997; Hinton et al., 1987), articulatory studies of this change have consistently found that speakers tend to retain rounding on /u/ and that fronted /u/ is contrasted from /i/ largely through lip rounding (Gorman & Kirkham, 2020; Harrington et al., 2011; Havenhill, 2024; Lawson et al., 2019; Scobbie et al., 2012). A confounding factor for /u/ is the fact that tongue fronting is initially driven by coarticulation, so preservation of the /i/-/u/ lingual contrast may be precluded by articulatory pressures. Thus while visual perceptibility of lip rounding may contribute to the tendency to retain lip rounding for /u/, these competing effects are difficult to disentangle. In the cases of both the Northern Cities Shift and the Low-Back-Merger Shift, advancement and retraction of LOT and THOUGHT are unconditioned and not motivated by coarticulation. Reconfigurations of the tongue or lips are equally viable strategies to achieve changes in F2, but visual perceptibility of lip rounding may contribute to a bias toward its preservation.

At the same time, the presence of merger/near-merger of LOT and THOUGHT in Chicago raises the obvious question of whether visual cues are in fact sufficient for inhibiting merger.

In Experiment 1, at least one speaker was found to use identical tongue positions and lip configurations for LOT and THOUGHT, which were also merged in acoustics. Acoustic distance between LOT and THOUGHT was low for CHI013 and CHI019, although both were found to distinguish the vowels to some extent and also to increase the acoustic and articulatory distance between them in clear speech. A number of forces are known to influence the direction of sound change and phonological evolution, so even if visual cues do play a role, they are at best only one of several factors and cannot be expected to be decisive. For one, the functional load between LOT and THOUGHT is relatively low (Wedel et al., 2013), so pressure to maintain this contrast is not particularly high from the perspective of lexical contrast. With regard to social factors, D'Onofrio and Benheim (2020) have argued that NCS reversal in some Chicago neighborhoods is driven by orientation away from stigmatized ideologies linked to socially salient NCS features. If the LOT-THOUGHT contrast itself carries social meaning in this way, then social pressures to merge LOT and THOUGHT may outweigh system-internal pressures toward retaining the contrast. In many areas, the reversal of local sound patterns has been interpreted as an orientation toward supralocal features (Labov et al., 2016; Prichard & Tamminga, 2012; Wagner et al., 2016). Such importation of non-local variants can be understood through the mechanisms proposed by Labov (2007), who distinguishes between two distinct types of sound change: transmission, in which dialect features are passed down generationally within a speech community, and diffusion, in which sound patterns spread piecemeal from one variety to another. If reversal of the NCS is the result of diffusion, rather than transmission, the visibility of lip rounding may be rendered irrelevant. While it may enable listeners to more readily perceive a LOT-THOUGHT distinction, social and other pressures may motivate them not to produce one.

Finally, further consideration of audiovisual perceptibility may inform theories of clear speech and hyperarticulation. Clear speech strategies have been shown to vary considerably on a speaker-by-speaker basis (e.g., Ferguson, 2004; Gagné et al., 1994; Perkell et al., 2002), and evidence is mixed for whether clear speech targets specific contrasts (Bradlow, 2002; Schertz, 2013; Wedel et al., 2018; Wright, 2004) and to what extent hyperarticulation is listener-oriented (Baese-Berk & Goldrick, 2009; Lee & Baese-Berk, 2020). As speakers possess phonetic knowledge of both the auditory-acoustic and visual-articulatory properties of their language, it is not necessarily the case that their primary goal in clear speech will be to increase the acoustic distance between sounds. Indeed, in real communicative situations, clear speech is frequently used when speaking under noisy conditions or when communicating with interlocutors with hearing loss. In situations where attempts to enhance auditory perceptibility are potentially futile, visual enhancement may more effectively fulfill the speaker's communicative goals. In this study, the degree to which participants enhanced their speech specifically for visual perceptibility may have been underestimated by the design of the clear speech task. Listener-oriented accounts predict that speakers will adapt their speech production efforts depending on their estimation

of the listener's perceptual requirements in a given communicative context. Here, speakers were told to speak as though they were correcting someone who misheard them, but the participants were not actually communicating with another person, nor were they speaking under noisy conditions. While previous work indicates that hyperarticulated speech can be observed with both real and imagined interlocutors, real listener-directed speech has been found to differ from imagined listener-directed speech (Hazan & Baker, 2011; Scarborough & Zellou, 2013, 2022). By examining articulatory modifications in a range of speaking contexts, future work may elucidate the extent to which speakers explicitly enhance their speech for visual as opposed to auditory perceptibility, and how visual vs. auditory goals vary according to speaking context.

6. Conclusion

This study contributes to a small but growing body of literature which suggests that the organization of phonological systems is influenced not only by the auditory-acoustic quality of speech sounds, but also by their visual perceptibility. When competing articulatory strategies yield acoustically similar output, speakers (and listeners) may be biased toward those that are perceptibly more distinct in both the auditory and visual domains. This preference is proposed to be reflected not only in their phonological inventory, but also in the articulatory strategies that speakers recruit when enhancing their speech for maximum perceptibility. Consideration of both articulatory and audiovisual-perceptual factors is crucial to understanding the mechanisms of sound change and speech production, uncovering patterns that cannot be explained by auditory-acoustic factors alone.

Supplementary file

Supplementary materials can be found here: https://doi.org/10.16995/labphon.11002.s1

Ethics and consent

This study was carried out in accordance with the recommendations of the Georgetown University Social and Behavioral Sciences Institutional Review Board (IRB-C) with written informed consent from all subjects.

Acknowledgements

Thank you to Jennifer Cole, Matt Goldrick, Annette D'Onofrio, and Chun Liang Chan for generously providing lab space and for their support in participant recruitment. Special thanks go to May Pik Yu Chan and Arthur Thompson, who assisted with data processing and model training, and to members of the Georgetown University PhonLab and the HKU Language Development Lab.

Competing interests

The author has no competing interests to declare.

References

Anderson, A. H., Bard, E. G., Sotillo, C., Newlands, A., & Doherty-Sneddon, G. (1997). Limited visual control of the intelligibility of speech in face-to-face dialogue. *Perception and Psychophysics*, *59*(4), 580–592. https://doi.org/10.3758/bf03211866

Articulate Instruments Ltd. (2008). *Ultrasound stabilisation headset users manual: Revision 1.4*.

Articulate Instruments Ltd. (2012). *Articulate Assistant advanced user guide: Version 2.14*.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527–554. https://doi.org/10.1080/01690960802299378

Barreda, S. (2015). phonTools: Functions for phonetics in R [R package version 0.2-2.1].

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

Bateman, N. (2011). On the typology of palatalization. *Language and Linguistics Compass*, *5*(8), 588–602. https://doi.org/10.1111/j.1749-818X.2011.00294.x

Becker, K. (Ed.) (2019). *The low-back-merger shift: Uniting the Canadian vowel shift, the California vowel shift, and short front vowel shifts across North America*. Duke University Press. https://doi.org/10.1215/00031283-8032913

Beddor, P. S., Krakow, R. A., & Goldstein, L. M. (1986). Perceptual constraints and phonological change: A study of nasal vowel height. *Phonology Yearbook*, *3*, 197–217. https://doi.org/10.1017/S0952675700000646

Bladon, R., & Nolan, F. (1977). A video-fluorographic investigation of tip and blade alveolars in English. *Journal of Phonetics*, *5*(2), 185–193. https://doi.org/10.1016/s0095-4470(19)31128-3

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics*, *32*(2), 117–166. https://doi.org/10.1515/TL.2006.009

Boersma, P., & Weenink, D. (2023). Praat: Doing phonetics by computer (version 6.3.11).

Bradlow, A. R. (1996). A perceptual comparison of the /i/–/e/ and /u/–/o/ contrasts in English and in Spanish: Universal and language-specific aspects. *Phonetica*, *53*(1–2), 55–85. https://doi.org/10.1159/000262189

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

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities. *Journal of Speech, Language, and Hearing Research*, *46*(1), 80–97. https://doi.org/10.1044/1092-4388(2003/007)

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, *89*, 68–86. https://doi.org/10.1016/j.jml.2015.12.009

Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-Based Models of Language* (pp. 65–85). CSLI.

Cahill, M. (1999). Aspects of the phonology of labial-velar stops. *Studies in African Linguistics*, *28*(2), 155–184. https://doi.org/10.32473/sal.v28i2.107374

Carignan, C. (2014). An acoustic and articulatory examination of the "oral" in "nasal": The oral articulations of French nasal vowels are not arbitrary. *Journal of Phonetics*, *46*, 23–33. https://doi.org/10.1016/j.wocn.2014.05.001

Carignan, C. (2017). Covariation of nasalization, tongue height, and breathiness in the realization of F1 of southern French nasal vowels. *Journal of Phonetics*, *63*, 87–105. https://doi.org/10.1016/j.wocn.2017.04.005

Carignan, C. (2018a). Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels. *The Journal of the Acoustical Society of America*, *143*(5), 2588–2601. https://doi.org/10.1121/1.5034760

Carignan, C. (2018b). Using naïve listener imitations of native speaker productions to investigate mechanisms of listener-based sound change. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 18. https://doi.org/10.5334/labphon.136

Carignan, C., Hoole, P., Kunay, E., Pouplier, M., Joseph, A., Voit, D., Frahm, J., & Harrington, J. (2020). Analyzing speech in both time and space: Generalized additive mixed models can uncover systematic patterns of variation in vocal tract shape in real-time MRI. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *11*(1). https://doi.org/10.5334/labphon.214

Carignan, C., Shosted, R. K., Fu, M., Liang, Z.-P., & Sutton, B. P. (2015). A real-time MRI investigation of the role of lingual and pharyngeal articulation in the production of the nasal vowel system of French. *Journal of Phonetics*, *50*, 34–51. https://doi.org/10.1016/j.wocn.2015.01.001

Chen, C. (2022). Voice quality of the nasal vowels in Chaoshan Chinese. In R. Billington (Ed.), *Proceedings of the 18th Australasian International Conference on Speech Science and Technology* (pp. 76–80). Australasian Speech Science and Technology Association. https://sst2022.com/wp-content/uploads/2022/12/sst2022_proceedings.pdf

Chen, C., & Havenhill, J. (2023). Articulation of the nasal vowels in Shanghai Chinese. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 1012–1016). Guarant International.

Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, *39*(3), 344–361. https://doi.org/10.1016/j.wocn.2011.02.005

Clarke, S., Elms, F., & Youssef, A. (1995). The third dialect of English: Some Canadian evidence. *Language Variation and Change*, *7*(2), 209–228. https://doi.org/10.1017/s0954394500000995

Clopper, C. G., Burdin, R. S., & Turnbull, R. (2019). Variation in /u/ fronting in the American Midwest. *The Journal of the Acoustical Society of America*, *146*(1), 233–244. https://doi.org/10.1121/1.5116131

Clopper, C. G., Mitsch, J. F., & Tamati, T. N. (2017). Effects of phonetic reduction and regional dialect on vowel production. *Journal of Phonetics*, *60*, 38–59. https://doi.org/10.1016/j.wocn.2016.11.002

Clopper, C. G., & Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *The Journal of the Acoustical Society of America*, *124*(3), 1682–1688. https://doi.org/10.1121/1.2953322

Clopper, C. G., & Tamati, T. N. (2014). Effects of local lexical competition and regional dialect on vowel production. *The Journal of the Acoustical Society of America*, *136*(1), 1–4. https://doi.org/10.1121/1.4883478

Dart, S. N. (1998). Comparing French and English coronal consonant articulation. *Journal of Phonetics*, *26*(1), 71–94. https://doi.org/10.1006/jpho.1997.0060

de Boer, B. (2001). *The origins of vowel systems*. Oxford University Press.

Delattre, P., & Freeman, D. C. (1968). A dialect study of American R's by X-ray motion picture. *Linguistics*, *6*(44), 29–68. https://doi.org/10.1515/ling.1968.6.44.29

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*(2), 121–144.

Dinkin, A. J. (2022). Adjusting to the new normal(ization): Adapting Atlas of North American English benchmarks to Lobanov-normalized data. *University of Pennsylvania Working Papers in Linguistics: Selected Papers from NWAV 49*, *28*(2). https://doi.org/20.500.14332/45376

D'Onofrio, A., & Benheim, J. (2020). Contextualizing reversal: Local dynamics of the Northern Cities Shift in a Chicago community. *Journal of Sociolinguistics*, *24*(4), 469–491. https://doi.org/10.1111/josl.12398

D'Onofrio, A., Eckert, P., Podesva, R. J., Pratt, T., & Van Hofwegen, J. (2016). The low vowels in California's Central Valley. *Publication of the American Dialect Society*, *101*(1), 11–32. https://doi.org/10.1215/00031283-3772879

Eckert, P. (2008). Where do ethnolects stop? *International Journal of Bilingualism*, *12*(1–2), 25–42. https://doi.org/10.1177/13670069080120010301

Espy-Wilson, C. Y. (2004). Articulatory strategies, speech acoustics and variability. In J. Slifka, S. Manuel, & M. Matthies (Eds.), *Proceedings of Sound to Sense: Fifty+ Years of Discoveries in Speech Communication* (pp. B62–B76). MIT Research Laboratory of Electronics.

Farrington, C., Kendall, T., & Fridland, V. (2018). Vowel dynamics in the Southern Vowel Shift. *American Speech*, *93*(2), 186–222. https://doi.org/10.1215/00031283-6926157

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *The Journal of the Acoustical Society of America*, *116*(4), 2365–2373. https://doi.org/10.1121/1.1788730

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

Ferragne, E., & Pellegrino, F. (2010). Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association*, *40*(1), 1–34.

FFmpeg Developers. (2018). FFmpeg v3.4.2 [computer software].

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically-Based Phonology*. Cambridge University Press.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics*, *14*(1), 3–28. https://doi.org/10.1016/s0095-4470(19)30607-2

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, *99*(3), 1730–1741. https://doi.org/10.1121/1.415237

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(3), 816–828. https://doi.org/10.1037/0096-1523.17.3.816

Fox, N. P., Reilly, M., & Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language*, *83*, 97–117. https://doi.org/10.1016/j.jml.2015.04.002

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America*, *126*(5), 2603–2618. https://doi.org/10.1121/1.3212921

Fricke, M., Baese-Berk, M. M., & Goldrick, M. (2016). Dimensions of similarity in the mental lexicon. *Language, Cognition and Neuroscience*, *31*(5), 639–645. https://doi.org/10.1080/23273798.2015.1130234

Fridland, V., & Kendall, T. (2019). On the uniformity of the Low-Back-Merger Shift in the U.S. West and beyond. In K. Becker (Ed.), *The Low-Back-Merger Shift: Uniting the Canadian Vowel Shift,*

the California Vowel Shift, and short front vowel shifts across North America (pp. 100–119). Duke University Press. https://doi.org/10.1215/00031283-8032957

Fridland, V., Kendall, T., & Farrington, C. (2014). Durational and spectral differences in American English vowels: Dialect variation within and across regions. *The Journal of the Acoustical Society of America*, *136*(1), 341–349. https://doi.org/10.1121/1.4883599

Fung, R. S. Y., & Lee, C. K. C. (2019). Tone mergers in Hong Kong Cantonese: An asymmetry of production and perception. *The Journal of the Acoustical Society of America*, *146*(5), EL424–EL430. https://doi.org/10.1121/1.5133661

Gagné, J.-P., Masterson, V., Munhall, K. G., Bilida, N., & Querengesser, C. (1994). Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech. *The Journal of the Academy of Rehabilitative Audiology*, *27*, 135–158.

Gagné, J.-P., Querengesser, C., Folkeard, P., Munhall, K. G., & Masterson, V. M. (1995). Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: An investigation of conversational and clear speech. *The Volta Review*, *97*(1), 33–51.

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

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789–806. https://doi.org/10.1016/j.jml.2011.11.006

Gardner, M. H., & Roeder, R. V. (2022). Phonological mergers have systemic phonetic consequences: PALM, trees, and the Low Back Merger Shift. *Language Variation and Change*, *34*(1), 29–52. https://doi.org/10.1017/s0954394522000059

Garellek, M., Ritchart, A., & Kuang, J. (2016). Breathy voice during nasality: A cross-linguistic study. *Journal of Phonetics*, *59*, 110–121. https://doi.org/10.1016/j.wocn.2016.09.001

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, *462*(7272), 502–504. https://doi.org/10.1038/nature08572

Gorman, E., & Kirkham, S. (2020). Dynamic acoustic-articulatory relations in back vowel fronting: Examining the effects of coda consonants in two dialects of British English. *The Journal of the Acoustical Society of America*, *148*(2), 724–733. https://doi.org/10.1121/10.0001721

Grama, J., & Kennedy, R. (2019). Dimensions of variance and contrast in the Low Back Merger and the Low-Back-Merger Shift. In K. Becker (Ed.), *The Low-Back-Merger Shift: Uniting the Canadian Vowel Shift, the California Vowel Shift, and Short Front Vowel Shifts Across North America* (pp. 31–55). Duke University Press. https://doi.org/10.1215/00031283-8032924

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *The Journal of the Acoustical Society of America*, *102*(1), 655–658. https://doi.org/10.1121/1.419712

Hall-Lew, L. (2013). 'Flip-flop' and mergers-in-progress. *English Language and Linguistics*, *17*(2), 359–390. https://doi.org/10.1017/s1360674313000063

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/- fronting, and sound change in Standard Southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America*, *123*(5), 2825–2835. https://doi.org/10.1121/1.2897042

Harrington, J., Kleber, F., & Reubold, U. (2011). The contributions of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern British English. *Journal of the International Phonetic Association*, *41*(2), 137–156. https://doi.org/10.1017/S0025100310000265

Havenhill, J. (2024). Articulatory and acoustic dynamics of fronted back vowels in American English. *The Journal of the Acoustical Society of America*, *155*(4), 2285–2301. https://doi.org/10.1121/10.0025461

Havenhill, J., & Do, Y. (2018). Visual speech perception cues constrain patterns of articulatory variation and sound change. *Frontiers in Psychology*, *9*, 728. https://doi.org/10.3389/fpsyg.2018.00728

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*(3). https://doi.org/10.1515/tlr.2006.014

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484. https://doi.org/10.1016/j.wocn.2005.10.001

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America*, *130*(4), 2139–2152. https://doi.org/10.1121/1.3623753

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Hinton, L., Moonwomon, B., Bremner, S., Luthin, H., Van Clay, M., Lerner, J., & Corcoran, H. (1987). It's not just the Valley Girls: A study of California English. *Annual Meeting of the Berkeley Linguistics Society*, *13*, 117–128. https://doi.org/10.3765/bls.v13i0.1811

Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., & Schiele, B. (2016). DeeperCut: A deeper, stronger, and faster multi-person pose estimation model. *European Conference on Computer Vision*, 34–50.

Jacewicz, E., & Fox, R. A. (2020). Perception of local and non-local vowels by adults and children in the South. *The Journal of the Acoustical Society of America*, *147*(1), 627–642. https://doi.org/10.1121/10.0000542

Jibson, J. (2021). Assessing merged status with Pillai scores based on dynamic formant contours. *Proceedings of the Linguistic Society of America*, *6*(1), 203. https://doi.org/10.3765/plsa.v6i1.4961

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variation in Speech Processing* (pp. 145–165). Academic Press.

Johnson, K. (2015). Audio-visual factors in stop debuccalization in consonant sequences. In *UC Berkely Phonology Lab Annual Report* (pp. 227–242). University of California, Berkeley.

Johnson, K., DiCanio, C. T., & MacKenzie, L. (2007). The acoustic and visual phonetic basis of place of articulation in excrescent nasals. In *UC Berkeley Phonology Lab Annual Report* (pp. 529–561). University of California, Berkeley.

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, *94*(2), 701–714. https://doi.org/10.1121/1.406887

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384. https://doi.org/10.1006/jpho.1999.0100

Kendall, T., & Fridland, V. (2017). Regional relationships among the low vowels of U.S. English: Evidence from production and perception. *Language Variation and Change*, *29*(2), 245–271. https://doi.org/10.1017/s0954394517000084

Kennedy, R., & Grama, J. (2012). Chain shifting and centralization in California vowels: An acoustic analysis. *American Speech*, *87*(1), 39–56.

King, H., & Chitoran, I. (2022). Difficult to hear but easy to see: Audio-visual perception of the /r/-/w/ contrast in Anglo-English. *The Journal of the Acoustical Society of America*, *152*(1), 368–379. https://doi.org/10.1121/10.0012660

King, H., & Ferragne, E. (2020). Labiodentals /r/ here to stay: Deep learning shows us why. *Anglophonia* 30. https://doi.org/10.4000/anglophonia.3424

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*(3), 419–454. https://doi.org/10.1353/lan.1994.0023

Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, *36*(1), 28–54. https://doi.org/10.1016/j.wocn.2007.02.001

Kochetov, A. (2011, April). Palatalization. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1–25). Wiley. https://doi.org/10.1002/9781444335262.wbctp0071

Krakow, R. A., Beddor, P. S., Goldstein, L. M., & Fowler, C. A. (1988). Coarticulatory influences on the perceived height of nasal vowels. *The Journal of the Acoustical Society of America*, *83*(3), 1146–1158. https://doi.org/10.1121/1.396059

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, *115*(1), 362–378. https://doi.org/10.1121/1.1635842

Kricos, P. B. (1996). Differences in visual intelligibility across talkers. In D. G. Stork & M. E. Hennecke (Eds.), *Speechreading By Humans and Machines: Models, Systems, and Applications* (pp. 43–53). Springer. https://doi.org/10.1007/978-3-662-13015-5_4

Labov, W. (1966). *The social stratification of English in New York City*. Center for Applied Linguistics.

Labov, W. (1991). Three dialects of English. In P. Eckert (Ed.), *New Ways of Analyzing Variation in English* (pp. 1–4). Academic Press.

Labov, W. (1994). *Principles of linguistic change*. Wiley-Blackwell.

Labov, W. (2007). Transmission and diffusion. *Language*, *83*(2), 344–387.

Labov, W. (2008). Triggering events. *Studies in the history of the English language IV: Empirical and analytical advances in the study of English language change*, *61*, 11–54.

Labov, W. (2019). Predicting the path of sound change. In K. Becker (Ed.), *The Low-Back-Merger Shift: Uniting the Canadian Vowel Shift, the California Vowel Shift, and Short Front Vowel Shifts Across North America* (pp. 166–179). Duke University Press. https://doi.org/10.1215/00031283-8032990

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English*. Walter de Gruyter. https://doi.org/10.1515/9783110206838

Labov, W., & Baranowski, M. (2006). 50 msec. *Language Variation and Change*, *18*(03). https://doi.org/10.1017/s095439450606011x

Labov, W., Fisher, S., Gylfadottír, D., Henderson, A., & Sneller, B. (2016). Competing systems in Philadelphia phonology. *Language Variation and Change*, *28*(3), 273–305. https://doi.org/10.1017/s0954394516000132

Lawson, E., Stuart-Smith, J., & Rodger, L. (2019). A comparison of acoustic and articulatory parameters for the GOOSE vowel across British Isles Englishes. *The Journal of the Acoustical Society of America*, *146*(6), 4363–4381. https://doi.org/10.1121/1.5139215

Lee, D.-Y., & Baese-Berk, M. M. (2020). The maintenance of clear speech in naturalistic conversations. *The Journal of the Acoustical Society of America*, *147*(5), 3702–3711. https://doi.org/10.1121/10.0001315

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. https://doi.org/10.1016/0010-0277(85)90021-6

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*(4), 839. https://doi.org/10.2307/411991

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental Phonology* (pp. 13–44). Academic Press.

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. https://doi.org/10.1007/978-94-009-2037-8_16

Lindblom, B., & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, *50*(4B), 1166–1179. https://doi.org/10.1121/1.1912750

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*, 606–608. https://doi.org/10.1121/1.1912396

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, *60*(4), 602–619. https://doi.org/10.3758/bf03206049

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*(2), 131–141. https://doi.org/10.3109/03005368709077786

Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press.

Majors, T., & Gordon, M. J. (2008). The [+spread] of the Northern Cities Shift. *University of Pennsylvania Working Papers in Linguistics*, *14*(2), 111–120. https://doi.org/20.500.14332/44693

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, *28*(3), 213–228. https://doi.org/10.3758/bf03204377

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America*, *69*(2), 548–558. https://doi.org/10.1121/1.385483

Martinet, A. (1955). *Économie des changements phonétiques*. Francke.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281–1289.

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)

Mayer, C., Gick, B., Tamra, W., & Whalen, D. H. (2013). Perceptual integration of visual evidence of the airstream from aspirated stops. *Canadian Acoustics*, *41*(3), 23–27.

McCarthy, C. (2010). The Northern Cities Shift in real time: Evidence from Chicago. *University of Pennsylvania Working Papers in Linguistics*, *15*(2). https://doi.org/20.500.14332/44736

McCarthy, C. (2011). The Northern Cities Shift in Chicago. *Journal of English Linguistics*, *39*(2), 166–187. https://doi.org/10.1177/0075424210384226

McGuire, G., & Babel, M. (2012). A cross-modal account for synchronic and diachronic patterns of /f/ and /θ/ in English. *Laboratory Phonology*, *3*(2), 1–41. https://doi.org/10.1515/lp-2012-0014

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748. https://doi.org/10.1038/264746a0

Ménard, L., Trudeau-Fisette, P., Côté, D., & Turgeon, C. (2016). Speaking clearly for the blind: Acoustic and articulatory correlates of speaking conditions in sighted and congenitally blind speakers. *PLOS ONE*, *11*(9), e0160088. https://doi.org/10.1371/journal.pone.0160088

Mielke, J. (2015). An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *The Journal of the Acoustical Society of America*, *137*(5), 2858–2869. https://doi.org/10.1121/1.4919346

Mielke, J., Baker, A., & Archangeli, D. (2016). Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language*, *92*(1), 101–140.

Mielke, J., Carignan, C., & Thomas, E. R. (2017). The articulatory dynamics of pre-velar and pre-nasal /æ/-raising in English: An ultrasound study. *The Journal of the Acoustical Society of America*, *142*(1), 332–349. https://doi.org/10.1121/1.4991348

Mines, M. A., Hanson, B. F., & Shoup, J. E. (1978). Frequency of occurrence of phonemes in conversational English. *Language and Speech*, *21*(3), 221–241. https://doi.org/10.1177/002383097802100302

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

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1048–1058. https://doi.org/10.1044/1092-4388(2004/078)

Nasir, S. M., & Ostry, D. J. (2006). Somatosensory precision in speech production. *Current Biology*, *16*(19), 1918–1923. https://doi.org/10.1016/j.cub.2006.07.069

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, *14*(7), 2152–2176.

Nelson, N. R., & Wedel, A. (2017). The phonetic specificity of competition: Contrastive hyperarticulation of voice onset time in conversational English. *Journal of Phonetics*, *64*, 51–70. https://doi.org/10.1016/j.wocn.2017.01.008

Nesbitt, M. (2023). Phonological emergence and social reorganization: Developing a nasal /æ/ system in Lansing, Michigan. *Language Variation and Change*, *35*, 273–297. https://doi.org/10.1017/S0954394523000182

Nesbitt, M., & Stanford, J. N. (2021). Structure, chronology, and local social meaning of a supra-local vowel shift: Emergence of the Low-Back-Merger Shift in New England. *Language Variation and Change*, *33*(3), 269–295. https://doi.org/10.1017/s0954394521000168

Nesbitt, M., Wagner, S. E., & Mason, A. (2019). A tale of two shifts: Movement toward the Low-Back-Merger Shift in Lansing, Michigan. In K. Becker (Ed.), *The Low-Back-Merger Shift: Uniting the Canadian Vowel Shift, the California Vowel Shift, and Short Front Vowel Shifts Across North America* (pp. 144–165). Duke University Press. https://doi.org/10.1215/00031283-8032979

Nycz, J. (2013). New contrast acquisition: Methodological issues and theoretical implications. *English Language and Linguistics*, *17*(2), 325–357. https://doi.org/10.1017/s1360674313000051

Nycz, J. (2018). Stylistic variation among mobile speakers: Using old and new regional variables to construct complex place identity. *Language Variation and Change*, *30*(2), 175–202. https://doi.org/10.1017/s0954394518000108

Nycz, J., & Hall-Lew, L. (2014). Best practices in measuring vowel merger. *Proceedings of Meetings on Acoustics*. https://doi.org/10.1121/1.4894063

O'Brien, J. P. (2012). *An experimental approach to debuccalization and supplementary gestures* [Doctoral dissertation, University of California].

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the Parasession on Language and Behavior* (pp. 178–203). Chicago Linguistic Society.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical Linguistics: Problems and Perspectives* (pp. 237–278). Longman.

Ohala, J. J. (1994). Acoustic study of clear speech: A test of the contrastive hypothesis. In *International Symposium On Prosody* (pp. 75–89).

Ohala, J. J., & Lorentz, J. (1977). The story of [w]: An exercise in the phonetic explanation for sound patterns. *Proceedings of the Annual Meeting of the Berkeley Linguistic Society*, *3*, 577–599. https://doi.org/10.3765/bls.v3i0.2264

Peirce, J. W. (2007). PsychoPy: Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Perkell, J. S., Zandipour, M., Matthies, M. L., & Lane, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *The Journal of the Acoustical Society of America*, *112*(4), 1627–1641. https://doi.org/10.1121/1.1506369

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184. https://doi.org/10.1121/1.1906875

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

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Typological Studies in Language* (pp. 137–157). John Benjamins. https://doi.org/10.1075/tsl.45.08pie

Pratt, T., & D'Onofrio, A. (2017). Jaw setting and the California Vowel Shift in parodic performance. *Language in Society*, *46*(3), 283–312. https://doi.org/10.1017/S0047404517000227

Prichard, H., & Tamminga, M. (2012). The impact of higher education on Philadelphia vowels. *University of Pennsylvania Working Papers in Linguistics*, *18*(2). https://doi.org/20.500.14332/44867

R Core Team. (2024). R: A language and environment for statistical computing R Foundation for Statistical Computing. Vienna, Austria.

Riordan, C. J. (1977). Control of vocal-tract length in speech. *The Journal of the Acoustical Society of America*, *62*(4), 998–1002. https://doi.org/10.1121/1.381595

Roeder, R. V., & Gardner, M. H. (2013). The phonology of the Canadian Shift revisited: Thunder Bay & Cape Breton. *University of Pennsylvania Working Papers in Linguistics*, *19*(2), 161–170. https://doi.org/20.500.14332/44923

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., & Yuan, J. (2015). FAVE (forced alignment and vowel extraction) program suite v1.2.2 [computer software]. https://doi.org/10.5281/zenodo.22281

Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology 10* (pp. 557–586). De Gruyter Mouton. https://doi.org/10.1515/9783110224917.5.557

Scarborough, R., Keating, P., Mattys, S. L., Cho, T., & Alwan, A. (2009). Optical phonetics and visual perception of lexical and phrasal stress in English. *Language and Speech*, *52*(2–3), 135–175. https://doi.org/10.1177/0023830909103165

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

Scarborough, R., & Zellou, G. (2022). Out of sight, out of mind: The influence of communicative load and phonological neighborhood density on phonetic variation in real listener-directed speech. *The Journal of the Acoustical Society of America*, *151*(1), 577–586. https://doi.org/10.1121/10.0009233

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in english. *Journal of Phonetics*, *41*(3–4), 249–263. https://doi.org/10.1016/j.wocn.2013.03.007

Scobbie, J. M., Lawson, E., & Stuart-Smith, J. (2012). Back to front: A socially-stratified ultrasound tongue imaging study of Scottish English /u/. *Rivista di Linguistica*, *24*(1), 103–148.

Shosted, R., Carignan, C., & Rong, P. (2012). Managing the distinctiveness of phonemic nasal vowels: Articulatory evidence from Hindi. *The Journal of the Acoustical Society of America*, *131*(1), 455–465. https://doi.org/10.1121/1.3665998

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

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass*, *3*(1), 236–264. https://doi.org/10.1111/j.1749-818x.2008.00112.x

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, 101017. https://doi.org/10.1016/j.wocn.2020.101017

Stanley, J. A. (2020). *Vowel dynamics of the elsewhere shift: A sociophonetic analysis of English in Cowlitz County, Washington* [Doctoral dissertation, University of Georgia].

Stanley, J. A., Renwick, M. E. L., Kuiper, K. I., & Olsen, R. M. (2021). Back vowel dynamics and distinctions in Southern American English. *Journal of English Linguistics*, *49*(4), 389–418. https://doi.org/10.1177/00754242211043163

Stanley, J. A., & Sneller, B. (2023). Sample size matters in calculating Pillai scores. *The Journal of the Acoustical Society of America*, *153*(1), 54–67. https://doi.org/10.1121/10.0016757

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*(1–2), 3–45. https://doi.org/10.1016/s0095-4470(19)31520-7

Stevens, K. N., & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, *38*(1), 10–19. https://doi.org/10.1016/j.wocn.2008.10.004

Stevens, K. N., Keyser, S. J., & Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 426–449). Psychology Press.

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, *19*(6–7), 455–501. https://doi.org/10.1080/02699200500113558

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Swan, J. T. (2019). The Low-Back-Merger Shift in Seattle, Washington, and Vancouver, British Columbia. In K. Becker (Ed.), *The Low-Back-Merger Shift: Uniting the Canadian Vowel Shift, the California Vowel Shift, and short front vowel shifts across North America* (pp. 74–99). Duke University Press. https://doi.org/10.1215/00031283-8032946

Thomas, E. R. (2001). *An acoustic analysis of vowel variation in New World English* (Vol. 85). Duke University Press.

Thomas, E. R. (2019). A retrospective on the Low-Back-Merger Shift. In K. Becker (Ed.), *The Low-Back-Merger Shift: Uniting the Canadian Vowel Shift, the California Vowel Shift, and short front vowel shifts across North America* (pp. 180–204). Duke University Press. https://doi.org/10.1215/00031283-8033001

Traunmüller, H., & Öhrström, N. (2007). Audiovisual perception of openness and lip rounding in front vowels. *Journal of Phonetics*, *35*(2), 244–258. https://doi.org/10.1016/j.wocn.2006.03.002

Tupper, P., Leung, K. W., Wang, Y., Jongman, A., & Sereno, J. A. (2021). The contrast between clear and plain speaking style for Mandarin tones. *The Journal of the Acoustical Society of America*, *150*(6), 4464–4473. https://doi.org/10.1121/10.0009142

Uchanski, R. M. (2005). Clear speech. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 207–235). Blackwell Publishing. https://doi.org/10.1002/9780470757024.ch9

Wade, L. (2017). The role of duration in the perception of vowel merger. *Journal of the Association for Laboratory Phonology*, *8*(1), 30. https://doi.org/10.5334/labphon.54

Wagner, S. E., Mason, A., Nesbitt, M., Pevan, E., & Savage, M. (2016). Reversal and reorganization of the Northern Cities Shift in Michigan. *University of Pennsylvania Working Papers in Linguistics*, *22*(2), 19. https://doi.org/20.500.14332/45120

Wedel, A., & Fatkullin, I. (2017). Category competition as a driver of category contrast. *Journal of Language Evolution*, *2*(1), 77–93. https://doi.org/10.1093/jole/lzx009

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, *128*(2), 179–186. https://doi.org/10.1016/j.cognition.2013.03.002

Wedel, A., Nelson, N., & Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language*, *100*, 61–88. https://doi.org/10.1016/j.jml.2018.01.001

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

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116.

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3–36.

Wood, S. (2017). *Generalized additive models: An introduction with R* (Second Edition). Chapman, Hall.

Wrench, A., & Balch-Tomes, J. (2022). Beyond the edge: Markerless pose estimation of speech articulators from ultrasound and camera images using DeepLabCut. *Sensors*, *22*. https://doi.org/https://doi.org/10.3390/s22031133

Wright, R. (2004). Factors of lexical competition in vowel articulation. In *Phonetic Interpretation: Papers in Laboratory Phonology 6* (pp. 75–87). Cambridge University Press. https://doi.org/10.1017/cbo9780511486425.005

Zellou, G., & Chitoran, I. (2023). Lexical competition influences coarticulatory variation in French: Comparing competition from nasal and oral vowel minimal pairs. *Glossa: A Journal of General Linguistics*, *8*(1). https://doi.org/10.16995/glossa.9801