

A cross-modal account for synchronic and diachronic patterns of /f/ and /θ/

Molly Babel¹ and Grant McGuire²

¹University of British Columbia and ²University of California at Santa Cruz

Corresponding author:

Grant McGuire
Department of Linguistics
University of California at Santa Cruz
Stevenson Faculty Services
Santa Cruz, CA, USA 95064
Email: gmcguir1@ucsc.edu
Telephone: 831-459-4988

Abstract

While the role of auditory saliency is well accepted as providing insight into the shaping of phonological systems, the influence of visual saliency on such systems has been neglected. This paper provides evidence for the importance of visual information in historical phonological change and synchronic variation through a series of audio-visual experiments with the /f~/θ/ contrast. /θ/ is typologically rare, an atypical target in sound change, acquired comparatively late, and synchronically variable in language inventories. Previous explanations for these patterns have focused on either the articulatory difficulty of an interdental tongue gesture or the perceptual similarity /θ/ shares with labiodental fricatives. We hypothesize that the bias is due to an asymmetry in audio-visual phonetic cues and cross-talker cue variability. Support for this hypothesis comes from a speech perception study that explored the weighting of audio and visual cues for /f/ and /θ/ identification in CV, VC, and VCV syllabic environments in /i/, /a/, or /u/ vowel contexts in Audio, Visual, and Audio-visual experimental conditions using stimuli from ten different talkers. The results indicate that /θ/ identification is more variable than /f/, both in Audio and Visual conditions. We propose that it is this variability which contributes to the unstable nature of /θ/ across time and offers an improved explanation for the observed synchronic and diachronic asymmetries in its patterning.

1. Introduction

While it has been known for some time that visual information plays an important role in speech perception, such information has only rarely been invoked as an explanation for historical changes and synchronic phonological patterns (see Johnson, DiCano, and Mackenzie 2007 for a notable exception). The goal of this paper is to argue for the importance of visual input on the shaping of phonologies, using evidence from the puzzling diachronic and synchronic instability of /θ/ and its frequent and asymmetric substitution with its highly confusable counterpart, /f/. Although the auditory confusability of the two sounds is well known (e.g. Miller and Nicely 1955) and has been argued as the cause of substitutions of the sounds (e.g. Labov et al 1968:93, Jones 2002), none of these accounts has been able to adequately handle the asymmetry in substitution. We argue that a multi-modal model of speech perception, coupled with the heretofore under-analyzed cross-talker articulatory variability of /θ/, can account for observed patterns.

Across the world's languages, and unlike /f/, /θ/ is a rare phoneme, occurring in less than 5% of languages (UPSID, Maddieson and Precoda 1990). In addition to being cross-linguistically rare, it is the frequent target of sound change (Blevins 2004), a synchronically volatile sound in English and Germanic (Kjellmer 1995), and acquired comparatively late by children (Edwards and Beckman 2008). Its synchronic volatility is evident in many varieties of English, where /θ/ often undergoes patterns of stopping to [t] or fronting to [f], particularly in coda position (Dubois & Horvath 1998, Wells 1982).

To explain this lack of stability, scholars have blamed either articulatory difficulty (Wells 1982, Kjellmer 1995) or perceptual weakness (Labov et al. 1968, Jones 2002). Articulatory

difficulty is an unlikely reason; coronals are generally considered easier articulatorily due to the high degree of flexibility and precision inherent in the tongue tip and there seems to be no reason why tongue to teeth contact is any more difficult than lower lip to teeth. Moreover, if it is more complex articulatorily it should be late-acquired, especially relative to /f/. While true for English, this is not true for Greek, where /θ/ is more frequent than /f/ (Edwards and Beckman 2008). However, /f/ and /θ/ are indeed highly confusable perceptually (Miller and Nicely 1955), due primarily to their spectral similarity (Tabain 1998), although listeners are able to identify /f/ and /θ/ at above chance accuracy (Jongman 1989). Consequently, due to this spectral similarity, listeners rely heavily on formant transitions in adjacent vowels to identify these fricatives (Harris 1958).

The high degree of auditory similarity has led researchers to suggest that visual and higher-level disambiguating semantic information are especially important in identifying /f/ and /θ/ (Miller and Nicely 1955). This hypothesis was thoroughly tested and supported by Jongman et al. (2003). They found that semantic context heavily affected labiodental ~ interdental disambiguation; similar information was negligible for the English sibilant distinction. Moreover, they found that visual information significantly improved the perceptibility of the non-sibilants (to a ceiling effect) and that visual information alone was sufficient to disambiguate them.

In addition to simply helping to disambiguate the sounds, the visual information may help account for the noted asymmetry where the sound change /θ/ > /f/ is common, while /f/ > /θ/ is rare, at best. We propose this is due to asymmetries in visual cues and cross-talker cue variability in /θ/ production that disadvantages it relative to /f/. Support for this hypothesis comes from the

study described below, which explores the weighting of audio and visual cues in /f/ and /θ/ identification using stimuli from multiple talkers. The use of multiple talkers is particularly novel. Previous audio-visual work on /f/ and /θ/ has only used a single talker (Jongman et al. 2003, Wang et al. 2008, 2009), assuming that the amount of useful visual-phonetic information provided by each individual is equal, an assumption not supported by our data. Before presenting our study in Section 4, we provide a brief overview of the role of visual saliency (and lack thereof) in speech perception in Section 2. In Section 3 we review the unique patterning of /θ/ across time and space in disparate language families. We discuss the results of the experiment in Section 5 and conclude in Section 6.

2 Visual information in speech perception

It has long been known that visual cues are important contributors to understanding spoken language. Perhaps the most famous example of the role of visual information in speech perception is the “McGurk Effect” (McGurk and McDonald 1976) in which a mismatch in auditory and visual information results in a misperception of the auditory information and sometimes also the visual information. The classic demonstration of this effect is a video of a face producing [gɑ] coupled with an auditory presentation of [ba] which results in a percept of [da] for most subjects. More than just a curiosity, this effect demonstrated the strong coupling of auditory and visual information in speech processing; in fact, such effects are even present early in infancy (Rosenblum et al. 1997). The effect is so robust that it works even when subjects are

aware of the deception (Massaro 1987), when the voice and face are mismatched for gender (Green et al. 1991), and when highly abstracted images are used (Rosenblum and Saldaña 1996).

One aspect where visual information has been shown to be of great importance is under degraded auditory conditions. This is true for both the hearing impaired (Grant et al. 1998) as well as for those with normal hearing where noise has been added to the signal (Sumbly and Pollack 1954, Summerfield 1987, Robert-Ribes et al. 1998). Under such degraded auditory conditions, the addition of visual information boosts subject accuracy in perception.

Additionally, this is apparently a fairly active process. Vatikiotis-Bateson et al. (1998) tracked the gaze location of subjects in noisy audio-visual conditions and found that increased noise led to more gazing directly at the talkers' lips rather than the eyes, though the eyes were still the primary target of looking. It should be noted that difficult speaking conditions are not necessary for these effects: improved accuracy can be seen in clear speech conditions, as in the aforementioned Jongman et al. (2003) work looking at the difficult (even in clear speech) /f~/θ/ contrast examined in the current paper.

Another source of evidence for the importance of visual input in speech perception comes from work with the blind. For example, Mills (1987) reports that blind English-acquiring children show reduced sensitivity to labial contrasts and show more confusions than sighted children. While such differences may not be surprising in acquisition, systematic differences seem to persist into adulthood, as demonstrated by recent work by Lucie Ménard and colleagues. Specifically, Ménard et al. (2009) found that congenitally blind speakers of French both perceived and produced speech differently from sighted speakers – blind speakers produced less distinct vowels, yet showed enhanced auditory acuity. While enhanced auditory acuity has been

shown before (e.g. Hugdahl et al. 2004), the specific contrasts that showed this effect were limited to those in which visual cues are most useful. Similarly, the novel production effect suggests that the lack of visual input degrades production abilities.

Taken together, these various strands of evidence all converge on the fundamental importance of visual input in speech perception. While not absolutely required to be a fluent speaker of a language, visual input nonetheless shapes production and perception in important ways. Therefore, it should not be surprising that visual saliency may play a role in historical change, just as auditory saliency does.

As we argue, an excellent test case for the role visual information in shaping spoken phonological systems is the asymmetric patterning of /f/ and /θ/. Given the high degree of acoustic similarity between these sounds, visual information likely plays an important role in their perception. While much of the work on these sounds has been done with English speaking populations, the $\theta > f$ asymmetry and the general weakness of /θ/ are not confined to English or even Germanic. The following section details the synchronic and diachronic patterns of these sounds across language families.

3 Patterns of /f/ and /θ/

3.1 Synchronic variation

Across dialects of English, /θ/ is highly variable, as has been the case throughout the history of English (Smith 2009). Dubois and Horvath (1998:248) state that it would be

challenging to find a part of the English-speaking world which is not home to a sociolect that exhibits a pattern where /θ/ is replaced by another phoneme. The two most common patterns for first language speakers of English seem to be /θ/-stopping (/θ/ > /t/), where /θ/ loses its fricative quality and becomes a coronal stop, or /θ/-fronting (/θ/ > /f/), where the fricative loses its dental place and merges with the labiodental fricative. Non-native varieties of English often have an additional pattern where /θ/ > /s/.

With respect to the two native patterns, fronting and stopping appear to have relatively different distributions across language varieties. Stopping patterns seem to have developed out of language contact contexts and have been reported in Caribbean English, Irish English, and Indian English (Wells1982), as well as in parts of the rural Midwest of the United States where there were long periods of bilingualism in German and Polish.^a

The /θ/-fronting pattern is one that is frequent in many British dialects – Cockney English is a frequently caricatured example – and in colonial varieties of British English, such as Australian and New Zealand dialects of English. According to Kerswill (2003), /θ/-fronting is a southern British dialect feature that is diffusing to more northerly British dialects through dialect contact. The pattern is being introduced by working class groups who are described as fronting /θ/ more than middle class groups; additionally, for both adult and child populations, males are reported to front /θ/ more than females.

In the US, many varieties of African American English are described as having a mixed system composed of both /θ/-stopping and /θ/-fronting (Labov et al. 1968, Wolfram 1994,

a This is based on our own observations of our native dialects of Midwestern US English.

Rickford 1999). In these dialects, the syllabic environments of the patterns are in complementary distribution: stopping occurs word-initially, and /θ/-fronting is described as taking place intervocally and word-finally (Labov et al. 1968).

The variation in /θ/ production across the English-speaking world has not been lost on linguists. Phoneticians have struggled with how to accurately describe sounds like /θ/ and /ð/. Languages that are claimed to have /θ/ and /ð/ vary between whether these sounds are actually manifested with an interdental (e.g., Castilian Spanish) or dental (e.g., Tamil) tongue gesture (Ladefoged and Maddieson 1996:143). With respect to English, Catford (1982: 151) disagrees with the characterization of [θ ð] as being interdental and instead describes these English fricatives with a dental position in which the tongue tip position is strictly behind the front teeth. He specifically states that the label “inter-dental” does not adequately describe “most normal articulations of these English sounds.” Catford’s assessment of what is a normal articulation of these sounds is certainly context dependent and he is likely writing from the perspective of a speaker of a variety of English spoken in the United Kingdom, as this articulatory feature seems to differ across dialects. Ladefoged and Maddieson (1996:143) provide the results of a study that examined /θ/ productions of 28 Californian university students and 28 British university students. Within these populations, they report that nearly 90% of Californian students had a visible interdental tongue gesture, whereas only 10% of British students had clear interdental gestures. The remaining 90% of the British students produced /θ/ with the tip of the tongue behind the front teeth, as suggested by Catford (1982). Despite this finding, Ladefoged and Maddieson (1996: 173) suggest that talkers are so inconsistent in their productions of /f/ and /θ/ that it is “profitless to try to characterize the acoustic spectra of the fricatives.” This claim is,

perhaps, supported by Harris (1958), who suggests the distinguishing acoustic aspects of these nonsibilant fricatives lie in the adjacent vowel formant transitions. This is considered the generally accepted view (Shadle et al. 1992).

As phoneticians struggle to accurately describe /θ/, infants struggle to acquire its perceptual contrast with /f/. While two month old babies are able to discriminate between modified natural tokens of [fa] and [θa] produced by a male speaker (Levitt et al. 1988), neither six month nor twelve month old babies could discriminate between naturally-produced tokens of [fi]/[θi] or [fa]/[θa] (Eilers et al. 1997). Narayan (to appear) comments that infants' difficulties with the contrast nicely matches the typological distribution. In production, English-acquiring children are slow to acquire an adult-like /θ/, but while this has been previously argued to be due to articulatory difficulty, Beckman and Edwards (2008) demonstrate that age-matched Greek-acquiring children master /θ/ at much earlier ages. This suggests that English-acquiring children are slow to master /θ/ due to its low frequency of occurrence in English and not because of some inherent challenge in producing /θ/.

3.2 Diachronic distribution

Broadly, theories of sound change can be characterized as either talker-driven or listener-driven. Those that are talker-driven focus on mechanical or articulatory changes being the root of systematic sound changes in a language (Osthoff and Brugmann 1878, Paul 1880 [1920], Bloomfield 1933). Listener-driven changes view the origin of sound change as based in the misperception or misinterpretation of a talker's intended production (Ohala 1981, 1993,

Blevins 2004). The $\theta > f$ change has been used as a classic example of a context-free sound change of the listener-driven type (Blevins 2004: 134). This pattern, argues Blevins, is a sweeping sound change, having occurred in Rotuman (Oceanic) diachronically and as a synchronic pattern in many dialects of English and Italian. Specifically, Blevins states the change cannot be explained in terms of “general articulatory variability,” and instead attributes the source of the change to misperception. Blevins (2004: 292) suggests the change may be driven by misperception in the course of first language acquisition. As mentioned above, misperception of these sounds is likely as they are highly confusable perceptually. Notably, the direction of this confusability in laboratory tasks mirrors the direction of the bias in sound change. In Miller and Nicely’s (1955) study, listeners were more likely to identify / θ / as /f/ than /f/ as / θ /.

In discussing the unstable phonetic nature of / θ / across a language system, the following question logically arises: if / θ / is such a less than ideal segment, where does it come from in the first place? While it is true that / θ / is an infrequent sound, it is found in disparate language families, which makes the general focus on English and other Germanic languages in the study of / θ / rather untoward. Fortunately, many of the language families with / θ / have benefited from historical analysis, so the origins of the sound can be reconstructed. A brief review of other language families with / θ / suggests that the voiceless interdental develops out of coronal stops or affricates that were, presumably, dentally articulated. For example, a few Bantu languages are known to have / θ /.

Ndonga, as a case in point, has / θ /, which has further developed into /f/ in Kwanyama. These nonsibilant fricatives appear to have developed out of Proto-Bantu **ki* and **ti* (Larry Hyman, personal communication).

Most Northern Athabaskan languages underwent the Great Northern Series Shift, in which a coronal fricative became interdental in all positions – **ts*, **tsʰ*, **tsʷ*, **s*, **z* > **tθ*, **tθʰ*, **tθʷ*, **θ*, **ð* (Leer 1996). Since then, however, the behavior of the interdental series has varied considerably across languages. For example, Koyukon later merged this interdental series with a voiceless lateral series (Krauss and Leer 1981), in what is considered a perceptually motivated merger by Howe and Fulop (2005). In Dëne Sųłíné, /θ/ merges with /h/ in syllable-initial position (Cook 2004). Gwich'in has a pattern where complex segments with interdentals become labiovelars, and free-standing voiced and voiceless interdental fricatives become labiovelar glides (Howe and Fulop 2005). The θ > f pattern emerges in Tulita-Slavey, where complex segments with interdentals become voiceless bilabial stops, but single interdental fricatives — voiced and voiceless — become labiodental /f/ and /v/. In the Slave dialect continuum, the behaviour of the Proto-Athabaskan coronal **ts* series serves as the major feature on which the continuum is split into its four major groups (Rice 1989). Slave has retained interdental /θ/ segments, while in Mountain, Bearlake, and Hare this series has become an assorted collection which contains labiodentals, labiovelars, and bilabials.

The pattern that emerges is that the θ > f sound change is commonplace, while the reverse, f > θ, appears to be nearly nonexistent. The only attested change that resembles the latter is the case of f > ð from Proto-Chuukic to Pulo-Annan (Bender et al. 2003: 4-5), which may be due to contact with neighboring Palauan (Emerson Odango, personal communication). Additionally, there is a reconstructed, but unattested, case of f > θ > d in Latin (Stuart-Smith 2004). Even considering these possible counterexamples, what becomes clear in our brief survey of /θ/ is that the instability of the segment synchronically across dialects of English is mirrored in

the sound's inability to be maintained in a system. The Great Northern Shift introduced a full series of /θ/s into eight languages but the interdental components of the series remain fully intact for only one language (Dene Tha). Only one other language, Alaska Gwich'in, retains a /θ/ release on a sole complex segment in the series.

4 Experiments

Three audio-visual experiments were designed and conducted to address the following issues. First, these experiments provide a necessary replication of Jongman et al. (2003) to confirm that visual information is indeed highly facilitative to the identification of the /f/ and /θ/ contrast. Moreover, we seek evidence supporting the hypothesis that /f/ is privileged in this modality.

Second, the experiments test the hypothesis that the differences seen by syllabic position in the distribution of the sounds in question find their origin in perceptual differences. We predict that listeners will perform more poorly at identifying the fricatives in intervocalic and coda positions. Not only does this prediction follow from the distribution of the sound patterns in contemporary varieties of English, but, for the CV context, it is in line with previous phonetic research indicating that acoustic-phonetic information is more salient in onset position (Fujimura et al. 1978, Ohala 1990).

Finally, considering the conflicting descriptions as to the exact place of articulation of /θ/ (Catford 1982, Ladefoged and Maddieson 1996) the experiments test the hypothesis that talker

variability, especially with respect to visual cues, disproportionately affects /θ/ to its perceptual detriment.

4.1 General Methods

4.1.1 Stimuli

Five male and six female native speakers of American English with some phonetics training provided the audio-visual stimuli. The extra female subject provided stimuli that were used as practice tokens. Audio and video recordings were made separately in the same session. Videos were recorded using a Casio Exilim Pro EF1 camera attached to a PC running Adobe Premier Pro CS5. The camera was set on a tripod approximately 10 feet from the talker who was seated in a chair in front of a neutral background. The camera was zoomed such that the talker's head filled the frame with a small amount of background visible around the head. The head never left the frame and the lower jaw/upper neck was always visible. Stimuli were displayed visually to the talker in a randomized order and consisted of the fricatives /f/ and /θ/ in CV, VCV, and VC contexts where the vowel was either /a/, /i/, or /u/ for a total of 18 stimuli. The consonants were represented orthographically to the talker using the symbols <f> and <th> while the IPA symbols were used for the vowels. Subjects were coached on the proper vowel quality for the IPA symbols before recording and encouraged not to reduce the vowel of the second syllable in VCV tokens. VCV tokens were consistently produced by the eleven talkers with a H* accent on the initial vowel and L% on the second vowel.

Audio recordings were made in a sound-attenuated room immediately following the video session. Subjects wore a head-mounted AKG C250 microphone positioned about two inches to the side of the mouth. Productions were digitally recorded to the hard drive of a PC at a 44K sampling rate. The orthographic representation of the stimuli noted above was used again and presented in a random order using E-Prime (Schneider et al. 2007). Audio and video recordings were synched using Adobe Premier Pro CS5. Synching was done such that the audio began with the first frame of visible articulation for each token.

4.1.2 Procedures

All experiments used a Yes-No task. Subjects were presented a single token per trial and responded, using a five button response box, whether the stimulus they heard contained an /f/ or /θ/. The left- and right-most buttons were labeled <f> and <th>, respectively. Subjects were encouraged to respond in less than one second and the trial timed out after 1500 milliseconds, as timed from the onset of the stimulus. Subjects were given feedback on their reaction time and were offered breaks between each block.

Stimuli were presented and data were recorded using the E-prime software suite. The experiment began with 18 trials – each possible token given the two fricatives, three vowel environments, and three syllable contexts – using the practice talker. Following that, subjects classified all tokens for each of the 10 talkers, three repetitions each in a randomized order for a total of 540 total trials. Stimulus presentation was blocked by talker such that subjects classified tokens for one talker before moving on to the next. Block order was randomized across subjects.

Additionally, subjects filled out a background questionnaire soliciting basic information about languages spoken and areas lived in. Following the experiment subjects filled out a brief questionnaire about the experiment and the stimuli. Finally, subjects were asked during the experiment, after each block, how “pleasant” the particular talker was. These data are not analyzed further in this paper.

4.1.3 Conditions and Participants

There were three experimental conditions: Audio, Visual, and Audio-Visual. In the Audio experiment only the audio stimuli were used. For the Visual experiment the synched audio-visual stimuli were used, but the audio was muted. In the Audio-Visual experiment the synched audio-visual stimuli were used. In all experiments the same female talker was reserved for practice stimuli and not used in the main experiment. Under all conditions instructions were kept identical except for minor variations in wording about whether subjects would hear, see, or hear and see talkers producing the stimuli.

Each experiment made use of a different group of participants. All were native speakers of West Coast English with no reported speech or hearing problems or bilingualism, although many had some training in second languages. The specific breakdown is as follows:

AUDIO EXPERIMENT: Twenty-seven (18 female) undergraduates from the University of California at Santa Cruz participated. They were compensated with course credit for participation.

VISUAL EXPERIMENT: Sixteen undergraduates (10 female) from the University of British Columbia participated. They were compensated with \$10 CAD for participation.

AUDIO-VISUAL EXPERIMENT: Sixteen undergraduates (9 female) from the University of California at Santa Cruz participated. They were compensated with course credit.

4.2 Condition and Segmental Effects

All accuracy scores were converted to the sensitivity measure d' (d-prime) according to MacMillan and Creelman (2005). This measure controls for any response bias the subject may have and allows a more accurate comparison across conditions and subjects. To do this the correct /f/ responses were arbitrarily assigned as 'hits' and over-application of /f/ responses to /θ/ trials were called 'false alarms'. These two measures were then combined to form a single measure of sensitivity, d' , which served as the dependent measure in the statistical analyses. A $d' = 0$ indicates no sensitivity to the contrast, meaning the subject is responding randomly. The upper limit for d' for these results is approximately 3.65, and this can be considered near perfect perception. Table 1 reports summary statistics for each condition.

	Sensitivity	Bias	Proportion Correct
Audio Only	1.38 (0.59)	-0.06 (0.43)	0.8 (0.1)
Visual Only	1.35 (0.65)	-0.1 (0.47)	0.79 (0.13)
Audio-Visual	1.83 (0.6)	0 (0.31)	0.87 (0.1)

Table 1. Summary statistics for each condition. The number in each cell indicates the mean value for each measure: sensitivity in d' , bias in c (criterion point; 0 = no bias, negative indicates bias to respond /f/), and proportion correct. Numbers in parentheses are standard deviations.

The overall design was a 2 (Fricative: /f θ/) x 3 (Vowel: /i a u/) x 3 (Syllable: CV, VC, or CVC) x 3 (Condition: Audio, Visual, or Audio-Visual) factorial design. The first analysis was conducted to establish the effects of syllable position and vowel quality on classifying the two consonants. To this end, the ten individual talkers were treated as variation for this purpose and are analyzed as a factor in the analysis reported in §4.4. The specific analysis consisted of a repeated-measures ANOVA with Syllable and Vowel as within-subjects factors and Condition as a between-subjects factor. Syllable ($F[2,102]=43.8$, $p < 0.001$), Vowel ($F[2,102]=40.6$, $p < 0.001$), and Condition ($F[2,51]=7.63$, $p < 0.01$) all returned as significant main effects. The ANOVA also revealed significant two-way Syllable x Condition ($F[4,102]=8.17$, $p < 0.001$), Vowel x Condition ($F[4,102]=90.2$, $p < 0.001$), and Syllable x Vowel ($F[4,204]=6.09$, $p < 0.001$) interactions, as well as a three-way Syllable x Vowel x Condition interaction ($F[8,204]=3.11$, $p < 0.01$).

The main effect of Condition was explored through Bonferroni corrected t -tests, which showed that the Audio-visual condition was significantly more salient than the Audio and Visual conditions ($p < 0.001$), but that the Audio and Visual conditions were not significantly different

from each other. This finding replicates previous research suggesting that the presence of the visual channel amplifies the gain on the auditory system (e.g. Summerfield, 1979). Post-hoc analyses of the main effect of Syllable were similarly explored and found that CV and VCV environments were significantly more salient than VC ($p < 0.001$), but that CV and VCV were not significantly different from each other. Post-hoc testing of the Vowel main effect revealed /u/ to be significantly more salient than either of the other two vowel contexts ($p < 0.01$). All main effects are depicted in the three panels in Figure 1.

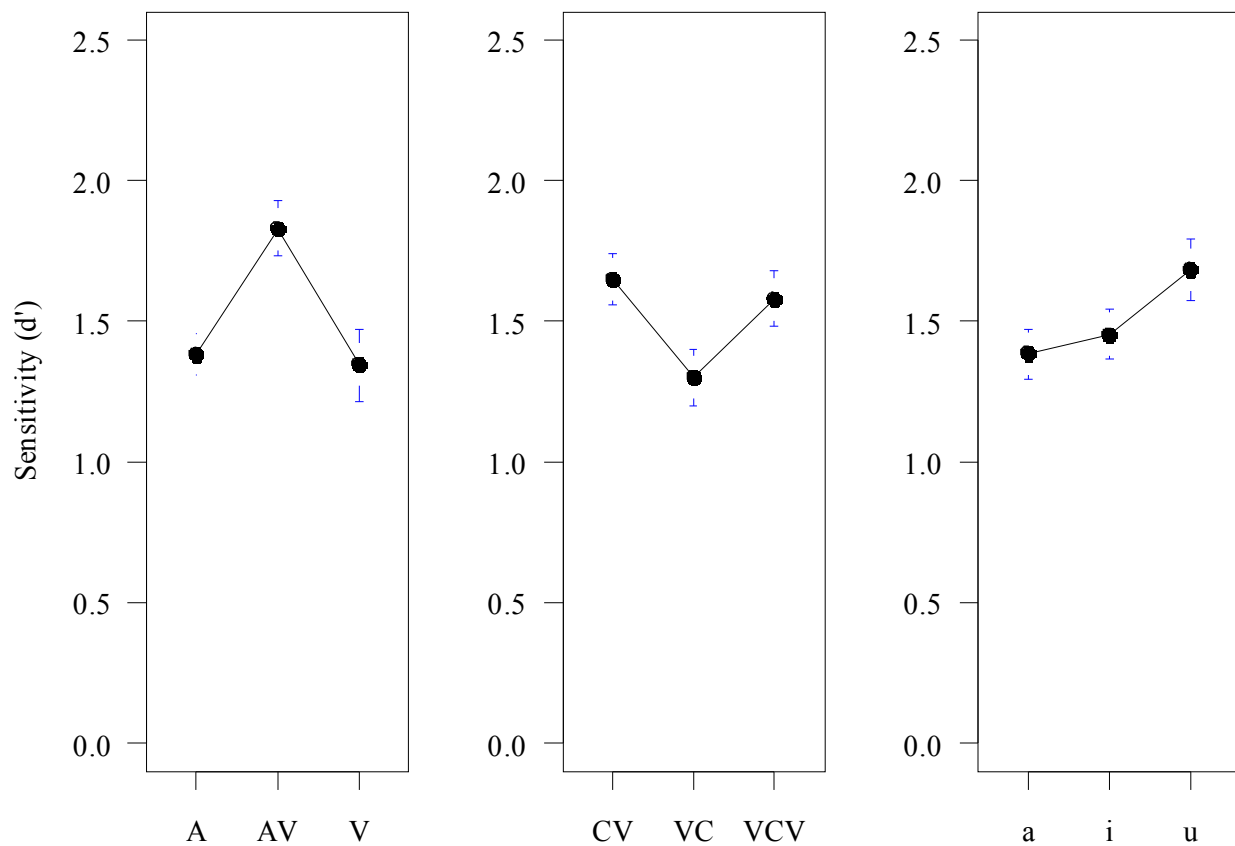


Figure 1. Main effects of Condition (left panel), Syllable (middle panel), and Vowel (right panel). The plots present listeners' d' scores for each group within each factor. A higher d' value indicates that listeners were more sensitive to the f/θ contrast. Error bars represent 95% confidence intervals.

These main effects demonstrate two main points. First, the Condition results replicate the intuitive finding of Jongman et al. (2003) that presenting the audio and visual channels simultaneously improves recognition of the stimuli. Second, our findings for the Syllable environments follows the pattern expected from the typological facts, where /θ/ alternation patterns are most frequent in coda position. Our results offer credence to the perceptual explanation for these patterns, that the coda position is simply a less salient position for the presentation of /θ/ cues. The vowel context effects suggest that /u/ is the most salient context for this fricative contrast. To our knowledge, this is a new finding; the details of this result are discussed further below, as we explore the interactions of the factors.

The interaction between Condition and Syllable is shown in Figure 2 for the Audio, Visual, and Audio-visual conditions. Selected Bonferroni corrected *t*-tests showed that for all conditions CV syllables were more salient than VC syllables ($p < 0.05$). The VCV environment was more salient than the VC environment only for the Audio condition ($p < 0.05$) and trends in that direction for the Audio-visual condition. Overall, the pattern of response to the three syllable environments is very similar across conditions, with the CV environment being the most salient, VC the least, and the saliency of the VCV context being the most variable across presentation modes.

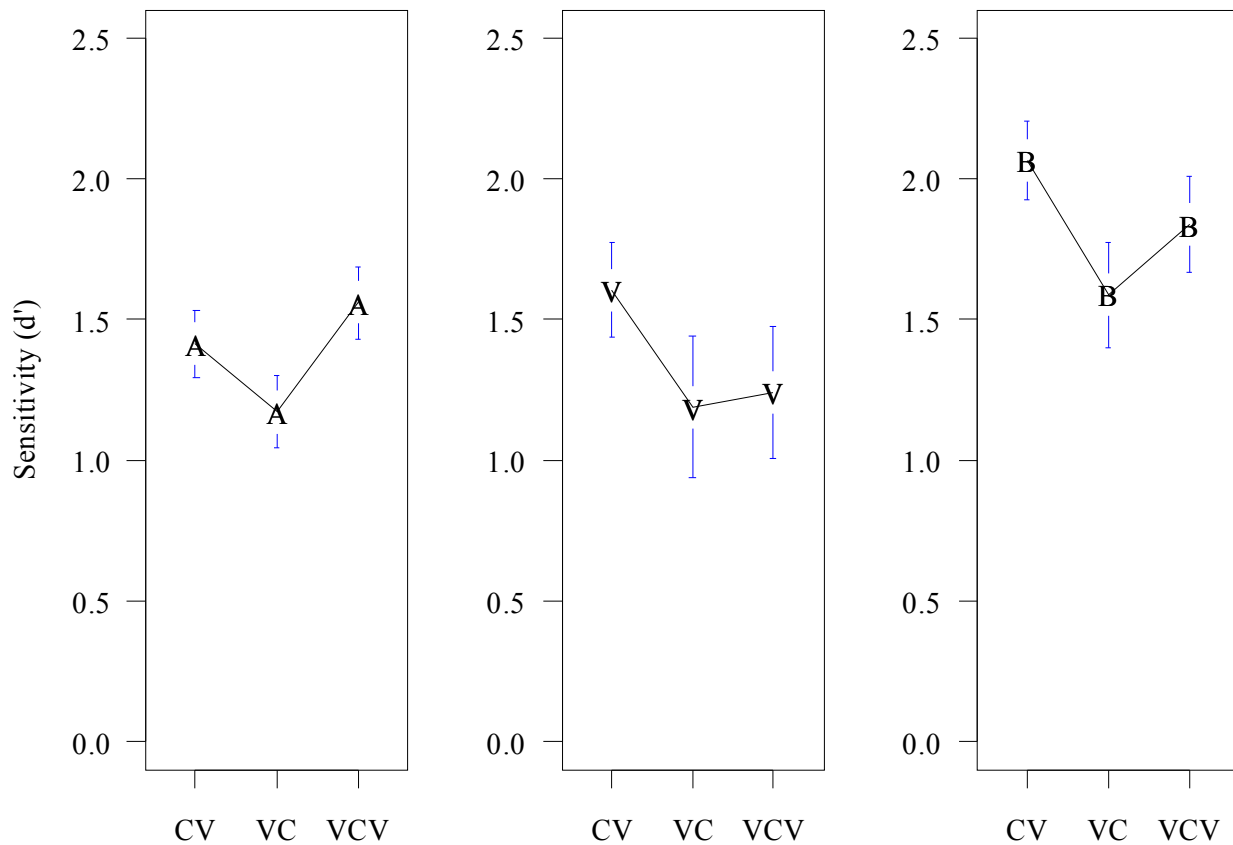


Figure 2. Sensitivity by syllable position for each condition. The left panel is Audio only, the middle panel is Visual only, and the right panel is Audio-visual (B = both audio and visual). Error bars represent 95% confidence intervals.

The Condition and Vowel interaction is shown in Figure 3 and demonstrates the highly divergent patterns across modalities. Post-hoc tests within each condition confirm that only the /u/ context was significantly different from the other vowel contexts for both Audio and Visual conditions ($p < 0.001$), but in very different directions. The /u/ context for the Audio condition garnered the highest sensitivity scores, but for the Visual condition, /u/ provided the environment in which listeners were the least sensitive. There were no differences across the vowel contexts for the Audio-visual condition.

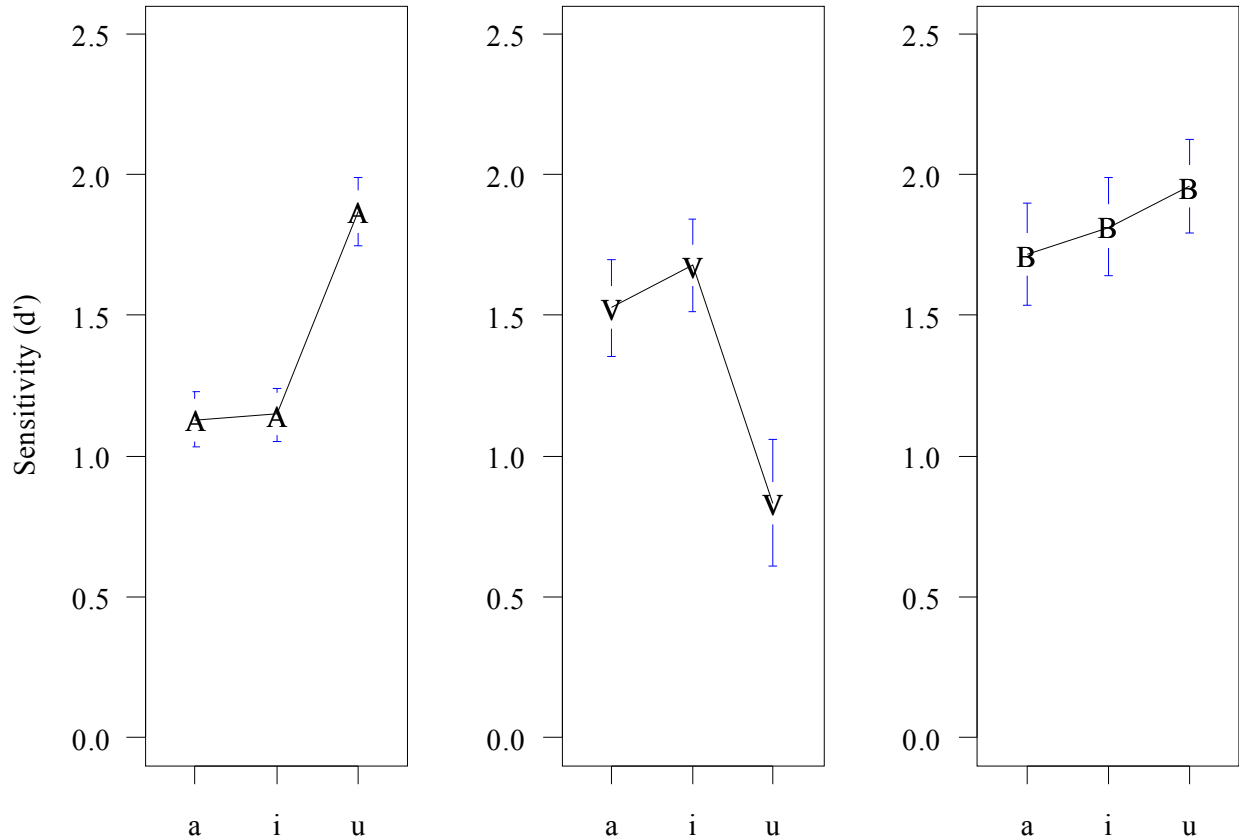


Figure 3. Sensitivity by vowel environment for each condition. The left panel is Audio only, the middle panel is Visual only, and the right panel is Audio-visual. Error bars represent 95% confidence intervals.

4.4 Condition and Indexical Effects

In order to assess the effects of individual talkers on subjects' sensitivity we conducted an analysis similar to the previous, but instead we pooled across syllable and vowel contexts in the calculation of d' . With this set of data, the first analysis was a repeated-measures ANOVA with Talker as a within-subject factor and Condition as an across-subject factor. Both Talker ($F[9,441]=10.3$, $p < 0.001$) and Condition ($F[2,49]=10.6$, $p < 0.001$) returned as significant. The

interaction between Talker x Condition was significant as well ($F[18,441]=7.74, p < 0.001$).

Figure 4 shows listener sensitivities for each of the ten talkers in each of the three conditions. Talker numbers are presented on the x -axis; f1-f5 are the five female talkers and m1-m5 are the five male talkers. Of special interest here is that the interaction reinforces the different patterns seen in listener sensitivity to each talker. Note that while listener sensitivity increases from the Audio to the Audio-visual conditions, the basic patterns in response to each talker are the same. This is not the case with the Visual condition, where we find a divergent pattern of talker-specific sensitivity that is not reducible to a gender difference. While the previous literature indicates that female talkers are on the whole more intelligible than male talkers (Bradlow et al. 1996), the five female talkers in this data set do not stand out as having higher listener sensitivities than the male talkers.

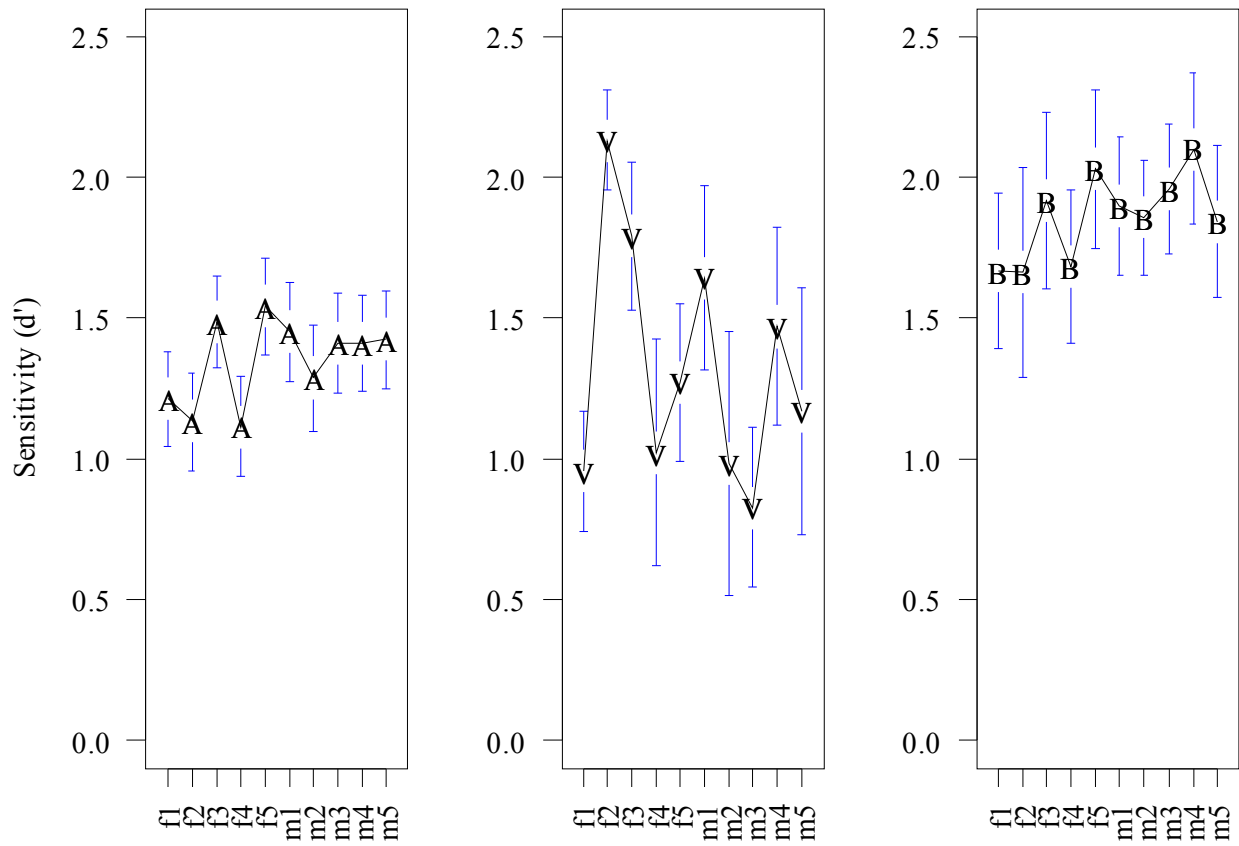


Figure 4. Listener sensitivity to the f/θ contrast for each talker in each condition. The left panel shows the Audio (A) condition, the middle panel is the Visual (V) condition, and the right panel shows the Audio-visual (B) condition. Error bars represent 95% confidence intervals.

As is observable from Figure 4, within the Audio and Visual conditions, several pairs of talkers have significantly different saliencies, as determined by Bonferroni-corrected t -tests ($p < 0.05$). In the Audio condition talkers f3 and f5 are significantly more salient than f2 and f4. However, in the Visual condition talker f2 is highly salient and significantly more so than f3 and m1, in addition to all other talkers. Note that in the Audio and Audio-visual conditions, f2 is actually less salient compared to those talkers. This was significant in the Audio only condition ($p < 0.05$). There were no differences between any pair of talkers in the Audio-visual condition, demonstrating that with both audio and visual channels available to listeners, all talkers produce an equally distinguishable contrast as measured by d' , statistically speaking.

Since the true interdental nature of / θ / is debatable (Catford 1982, Ladefoged and Maddieson 1996) and in an effort to understand the talker-specific findings, a trained phonetician (GM) coded the presence or absence of a visible tongue gesture for each / θ / stimulus for each talker frame-by-frame with the visual stimuli. To test the hypothesis that this visible tongue gesture was responsible for listeners' responses in the visual condition, the binary tongue present or absent scoring was correlated with sensitivity to each talker for each condition. This analysis reveals that a visible tongue gesture and talkers' sensitivities are highly correlated for the Visual condition (Spearman = 29.2, $p < 0.01$, $\rho = 0.82$), but not for the Audio (Spearman = 109, $p = 0.34$, $\rho = 0.33$) or Audio-visual (Spearman = 164, $p = 0.97$, $\rho = 0.01$) conditions. Figure 5 presents these correlations for each condition. Note that f2, the talker to whom listeners were most sensitive in the Visual Condition, has the highest tongue gesture score. In each production of / θ /, f2 exhibited a visible tongue gesture. Participants were able to make use of this information in the Visual condition, but, given her comparably low sensitivity scores in the

Audio and Audio-visual condition, listeners did not capitalize on this visual cue in the presence of the audio channel.

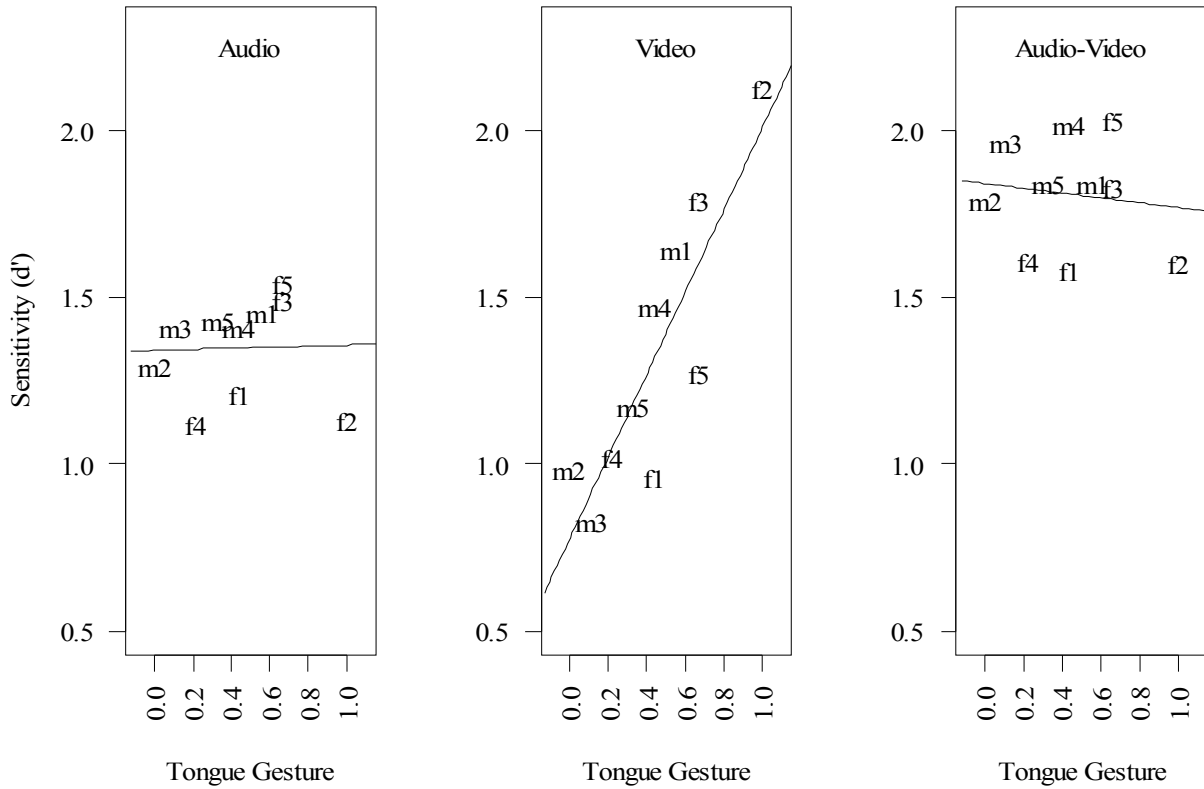


Figure 5. The average tongue gesture score for each talker plotted against average d' sensitivity values for each talker by condition. A Tongue Gesture score of 0 indicates that the talker's tongue was never visible during / θ / production and a score of 1 indicates it was always visible. Correlation lines are plotted and were determined by Spearman's rho.

Examining the tongue gesture score also offers insight into the poor performance in the context of /u/ for the visual condition (see Figure 3). Figure 6, which plots the gesture score against each vowel context, clearly demonstrates that the visual cues for / θ / are indeed very weak

for /u/. This is either due to lip rounding obscuring the tongue or reduced interdentality in the /u/ context. Evidence for the latter comes from the acoustics of the stimuli and is addressed below.

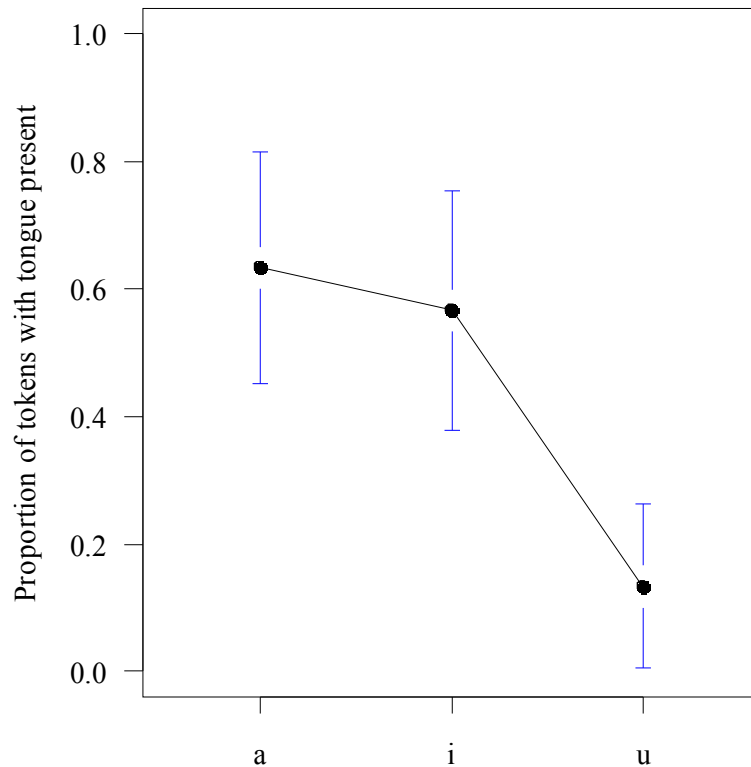


Figure 6. Tongue gesture score plotted for each vowel context, pooled across talkers. Error bars represent 95% confidence intervals.

5. Discussion

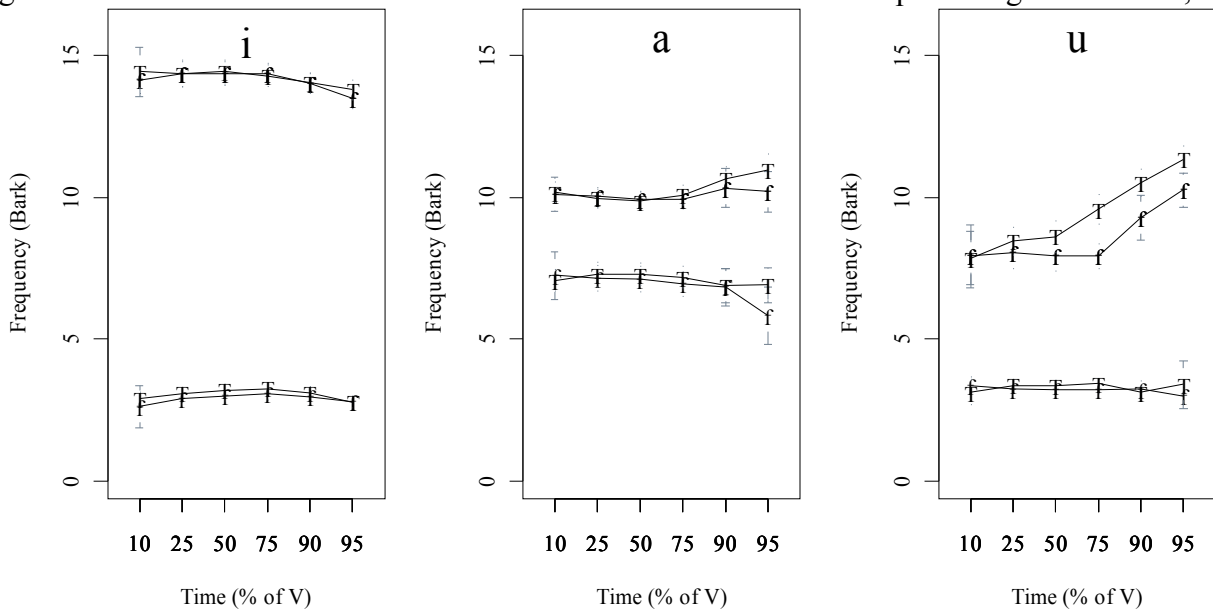
As noted previously, Catford (1982) disagrees with the classification of /θ/ as an interdental, based presumably on his experience with speakers of British varieties of English. Ladefoged and Maddieson (1996) find that the use of an actual visible interdental tongue gesture varies considerably across dialects. In this study, we find that the ten talkers who served as stimuli differ widely in their use of a visible tongue gesture. These talkers span a range from never using an interdental tongue gesture and receiving a tongue gesture score of 0.0 (e.g., talker

m2) to always displaying a visible tongue gesture (e.g., talker f2) and being scored with a tongue gesture score of 1.0. This variation negatively affected listeners. Perceptual sensitivities to talkers differed widely in the Visual condition, whereas the degree of difference between talkers was slight in the Audio condition, and nonexistent in the Audio-visual condition (see Figure 4). The Audio and Audio-visual conditions in this experiment had the audio stimuli presented in the clear, with no background noise or any other type of degradation to the signal. Previous research has found that listeners exploit information in the visual channel in fricative identification when the audio channel has been compromised (Wang et al. 2008, 2009). We predict that listeners' categorization of the stimuli as /f/ or /θ/ in such environments would then more closely resemble the talker-specific pattern in the Visual Condition, resulting in lowered perceptual sensitivity to select talkers.

Variation in /θ/ production and the vowel environment results are also interrelated. The /u/ environment was found to produce the highest listener sensitivities for the Audio condition and the lowest for the Visual condition. Figures 6 and 7 show the first and second formant frequencies across the durations of the onset and coda vowels, respectively, in the tokens used in the task. The Time axis in these figures presents the Bark-scaled formant frequency averages at normalized time points from the vowels. The difference between the formant transitions going into and coming out of /f/ and /θ/ (represented as "T" in the figures) is greatest in the /u/ environment. The higher F2 onset of /θ/ in the /u/ environments is indicative of a more dentally articulated /θ/. This subtle place of articulation shift further differentiates the transition cues for /f/ and /θ/, and we suggest that listeners are making use of this information in the Audio condition. While the /u/ context was an enhanced environment in the Audio condition, it

garnered the lowest sensitivity scores in the Visual condition. While it is possible that the lip-rounding of /u/ effectively obscures any visible tongue gesture listeners would take as evidence of a /θ/ production, the high F2 onset of /θ/ in this context suggests that listeners struggled to accurately identify /θ/ in these instances because it was in fact a dentally, and not interdentially, articulated /θ/ which did not offer the necessary visual cues to make a /θ/ decision. What is interesting is that while listeners have access to both this useful acoustic information and the useless visual information in the Audio-visual condition, the pattern of listener responses in the /u/ environment more closely resembles that of the Audio condition. Listeners are able to disregard the uninformative visual cues in the process of labeling a stimulus as /f/ or /θ/ in the Audio-visual design.

Figure 7. Bark-scaled first and second formant transitions for the vowel preceding the fricative,



collapsed across both VC and the initial vowel in VCV contexts. The Time label on the x-axis presents formant frequency averages at normalized time points throughout the vowel. Time point 100%, for example, would be exactly at the vowel-consonant boundary. In the figures, “T” = /θ/ and “f” = /f/. Error bars represent 95% confidence intervals.

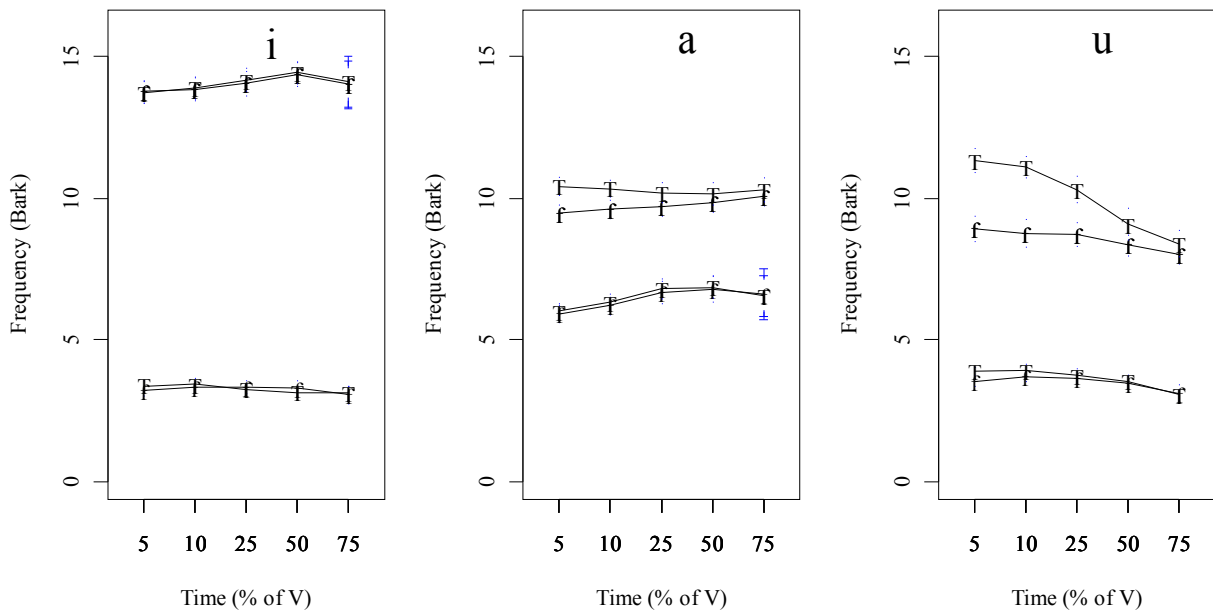


Figure 8. Bark-scaled first and second formant transitions for the vowel following the fricative, collapsed across both CV and the second vowel in VCV contexts. The Time label on the x-axis presents formant frequency averages at normalized time points throughout the vowel. Time point 0%, for example, would be exactly at the consonant-vowel boundary. In the figures, “T” = /θ/ and “f” = /f/. Error bars represent 95% confidence intervals.

The acoustic motivations for the Syllable context effects are also observable from the formant data in Figures 7 and 8, where it is clear that the difference in second formant frequency transitions out of /f/ and /θ/ are greater when the fricative is in onset position. We found that listeners were least sensitive to the fricative contrast in coda position. This result echoes previous

work arguing that acoustic cues are more robust in transitions out of a consonant (Fujimura et al. 1978, Ohala 1990). In finding the coda position to be the most perceptually challenging environment, we also support descriptions of more /θ/-fronting and /θ/-stopping in coda positions (Wells 1982, Dubois and Horvath 1998).

6. Conclusion

The series of experiments described above offers evidence that /θ/ is a particularly unstable segment because of the asymmetry in the audio and visual cues provided by native speakers. This hypothesis is an improvement over previous phonetically-motivated arguments for the $\theta > f$ bias. The acoustic fact that these sounds are spectrally very similar (Harris 1958, Ladefoged and Maddieson 1996, Tabain 1998) predicts that these sounds would be perceptually confusable (Miller and Nicely 1955). While true, this observation makes no headway into explaining why the sounds behave asymmetrically. Labeling /θ/ as articulatorily difficult (Wells 1982, Kjellmer 1995) accounts for the asymmetry, if that indeed were the case. Our argument makes use of audio and visual channels, both of which are involved in speech perception (e.g., McGurk and McDonald 1976, Summerfield 1979), and reasons, particularly, as to why the bias in perception would be targeted to /θ/.

Specifically, the results of this study indicate that /θ/ identification is more variable, both in Audio and Visual conditions: variability in the Audio condition involved acoustically-suggested shifts in the place of articulation of /θ/ to a more dental fricative in the /u/ context; variability in the Visual condition revolved around talker variability in the production of /θ/ and

the visibility of the /θ/ tongue gesture across vowel contexts. We propose that it is this feature – variability – which contributes to its volatility across time and offers an explanation for the observed synchronic and diachronic asymmetries in its patterning. While variability is typically not an argument for lack of stability in a system, Clayards and colleagues (2008) find that wider probability distributions give way to increased perceptual uncertainty. In the case of f-θ, listeners must be prepared to face both unpredictable inter-talker variability in /θ/ production along with the perhaps slightly more predictable intra-talker variability in /θ/ production across contexts. Failure to perceive either an audio or visual /θ/ cue will most likely lead to the sound being categorized as /f/ based on both their acoustic and visual phonetic similarities. Listeners, being more likely to miss a cue than fabricate one's existence (Chang et al. 2001), will then identify the missed /θ/ as an /f/.

Generally, our argument is similar to Kiparsky's (1995) model of phonological change where phonetic variation gives way to sound change through perceptual reinterpretation. Our approach differs from that of Kiparsky in that at least perceptual uncertainty, if not perceptual reinterpretation, occurs in adult listeners and is not limited to language learners. Talker-driven theories of sound change, by their nature, do not consider visual perceptual cues as factors in sound change, while listener-driven theories have focused solely on auditory cues as the source of misperception by listeners. Considering that the variability of /θ/ is rooted in phonetic cue quality across both the audio and visual channels, these data demonstrate the need to heed multi-modal phonetic information when theorizing about sound change, acquisition, and the typological distribution of sounds.

Acknowledgements

Author order is alphabetical. The authors would like to thank audiences at LabPhon 12 (Albuquerque), the University of California, Berkeley, University of California, Santa Cruz, and the University of British Columbia for their feedback on various stages of this work and especially to Jaye Padgett for comments on an early draft of this paper. Thanks to Jennifer Abel, Ryan Bennett, Meaghan Delaney, Ronald Sprouse, and, especially, Travis Whitebread. This work has been inspired by discussions with Keith Johnson, Andrew Garrett, and John Ohala. A special thank you goes to our research participants as well.

References

- Bender, Byron, Ward Goodenough, Frederick Jackson, Jeffrey Marck, Kenneth Rehg, Ho-min Sohn, Stephen Trussel, & Judith Wang (2003): Proto-Micronesian Reconstructions 1. *Oceanic Linguistics* 42 (1), 1-110.
- Blevins, Juliette. (2004). *Evolutionary Phonology: The emergence of sound patterns*. Cambridge University Press.
- Bloomfield, Leonard. (1933). *Language*. New York: Henry Holt.
- Bradlow, Anne R., Gina M. Torretta, & David B. Pisoni (1996). Intelligibility of normal speech I: Global and fine-grained acoustic–phonetic talker characteristics, *Speech Communication*, 20 (3-4), 255-272.
- Catford, J.C. (1982). *Fundamental problems in phonetics*. Midland, ON: Midland Books.
- Chang, Steve, Madelaine C. Plauché, & John J. Ohala. (2001). Markedness and Consonant Confusion Asymmetries. In Elizabeth Hume & Keith Johnson (Eds.), *The Role of Speech Perception in Phonology*, 79-101. Academic Press.
- Clayards, Meghan, Michael Tanenhaus, Richard Aslin, & Robert Jacobs (2008). Perception of speech reflects optimal use of probabilistic speech cues, *Cognition*, 108 (3), pp. 804-809.
- Cook, Eung-Do. (2004). *A grammar of Dëne Sų́líné (Chipewyan)*. Winnipeg, MB: Algonquian and Iroquoian Linguistics.
- Dubois, S. & B. M. Horvath (1998). Let's tink about dat: interdental fricatives in Cajun English. *Language Variation and Change*, 10(3), 245-262

- Edwards, Jan & Mary E. Beckman. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development, *Language Learning and Development*, 4(2), 122–156.
- Eilers, Rebecca, Wesley R. Wilson, & John M. Moore. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research*, 20, 776-780.
- Fujimura O., M.J. Macchi, & L. A. Streeter (1978). Perception of stop consonants with conflicting transitional cues: a cross-linguistic study. *Language and Speech*, 21 (4), 337-346.
- Grant, K. W., B.E. Walden, & P.F. Seitz. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America*, 103, 2677–2690.
- Green, K. P., P. K. Kuhl, A. M. Meltzoff, & E.B. Stevens. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception and Psychophysics*, 50, 524-536.
- Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1, 1-7.
- Howe, D. & S. Fulop. (2005). Acoustic features in Athabaskan. *Unpublished manuscript*.
- Hugdahl, K., M. Ek, T. Rintee, J. Tuomainen, C. Haaral, & K. Hämäläinen. (2004). Blind individuals show enhanced perceptual and attentional sensitivity for identification of speech sounds. *Cognitive Brain Research*, 19, 28–32.

- Johnson, Keith, Christian DiCanio, & Laurel McKenzie. (2007). The Acoustic and Visual Phonetic Basis of Place of Articulation in Excrescent Nasals. *UC Berkeley Phonology Lab Annual Report*, <http://linguistics.berkeley.edu/phonlab/annual_report/documents/2007/JohnsonDicianioMackenzie.pdf>, 529-561.
- Jones, Mark J. (2002). More on the “instability” of interdental fricatives: Gothic *þliuhan* ‘flee’ and Old English *flēon* ‘flee’ revisited. *Word* 53 (1), 1-8.
- Jongman, Allard (1989). Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America*, 85, 1718-1725.
- Jongman, Allard, Yue Wang, & Brian Kim. (2003). Contributions of sentential and facial information to perception of fricatives. *Journal of Speech, Language, and Hearing Research*, 46, 1367-1377
- Kerswill, Paul (2003). Dialect levelling and geographical diffusion in British English. In Peter Trudgill, David Britian, & Jenny Cheshire (Eds.), *Social dialectology: in honour of Peter Trudgill*, 223-244. John Benjamins Publishing.
- Kiparsky, Paul. (1995). The Phonological Basis of Sound Change. In John Goldsmith (Ed.), *The Handbook of Phonological Theory*, 640-670. Blackwell Publishing.
- Kjellmer, Göran. (1995). Unstable fricatives: On Gothic *þliuhan* ‘flee’ and Old English *flēon* ‘flee’. *Word* 46. 207-23.
- Krauss, M. & J. Leer. (1981). Athabaskan, Eyak, and Tlingit sonorants. *Alaska Native Language Center Research Papers* 5.

- Labov, William, Paul Cohen, Clarence Robins, & John Lewis. 1968. *A study of the non-standard English of Negro and Puerto Rican speakers in New York City*. Final report, Cooperative Research Project 3288. Vols. I and II.
- Ladefoged, Peter & Ian Maddieson. (1996). *The Sounds of the World's Languages*. Cambridge, MA: Blackwell Publishers.
- Leer, J. (1996). The historical evolution of the stem syllable in Gwich'in (Kutchin /Loucheux) Athabaskan. In *Athabaskan language studies: Essays in honor of Robert W. Young*, eds. Eloise Jelinek, Sally Midgette, Keren Rice and Leslie Saxon, 193-234. Albuquerque: University of New Mexico Press.
- Levitt, Andrea, Peter W. Jusczyk, Janice Murray, & Guy Carden. (1988). Context effects in Two-Month-Old Infants' Perception of Labiodental/Interdental Fricative Contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 14 (3), 361-368.
- Maddieson, I., & K. Precoda. (1990). *UPSID-PC. The UCLA Phonological Segment Inventory Database*.
- Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Lawrence Erlbaum Assoc., Inc.
- Ménard, L., S. Dupont, S. Baum, & J. Aubin. (2009). Production and perception of French vowels by congenitally blind and sighted adults. *Journal of the Acoustical Society of America*, 126 (3), 1406-1414.
- McGurk, H. and J. MacDonald (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

- Miller, G. A. & P. E. Nicely. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27 (2): 338-352.
- Mills, A.E. (1987). The development of phonology in the blind child, in *Hearing by Eye: The Psychology of Lip-Reading*, edited by B. Dodd and R. Campbell (Lawrence Erlbaum Associates, London), pp. 145–163.
- Narayan, Chandan R. (To appear). Developmental perspectives on phonological typology and sound change, In A.C.L. Yu (Ed.), *Origins of Sound Patterns: Approaches to Phonologization*, Oxford University Press.
- Ohala, John. J. (1981). The listener as a source of sound change. In C. S. Masek, R. Hendrick, & M. Miller (Eds.), *Papers from the Parasession on Language and Behavior*, 178-203. Chicago Linguistics Society.
- Ohala, John. J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the grammar and the physics of speech*, 258-275. Cambridge: Cambridge University Press.
- Ohala, John. J. (1993). The phonetics of sound change. In Charles Jones (Ed.), *Historical Linguistics: Problems and Perspectives*, 237-278, London: Longman.
- Osthoff, Hermann & Karl Brugman. (1878). Vorwort. *Morphologische Untersuchungen*, 1:iii-xx.
- Paul, Hermann. (1880). *Principien der Sprachgeschichte*. 1st Edition. Max Niemeyer.
- Paul, Hermann. (1920). *Principien der Sprachgeschichte*. 5th Edition. Halle: Niemeyer.
- Rice, Keren. (1989). *A grammar of Slave*. Berlin, New York: Mouton de Gruyter.

- Rickford, John R. (1999). *African American Vernacular English: Features, Evolution, Educational Implications*. Wiley Blackwell.
- Robert-Ribes, J., J. L. Schwartz, T. Lallouache, & P. Escudier. (1998). Complementarity and synergy in bimodal speech: Auditory, visual and audiovisual identification of French oral vowels in noise. *Journal of the Acoustical Society of America* 103, 3677–3689.
- Rosenblum, L. D. & H. M. Saldaña. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 22(2), 318-331.
- Rosenblum, L.D., M.A. Schmuckler, & J.A. Johnson. (1997). The McGurk effect in infants. *Perception & Psychophysics*, 59(3), 347-357.
- Schneider, Walter, Amy Eschman, & Anthony Zuccolotto. (2007). E-Prime: User's Guide, Version 2.0. Psychology Software Tools.
- Shadle, C. H., A. Moulinier, C.U. Dobelke, & C. Scully (1992). Ensemble averaging applied to the analysis of fricative consonants. *Second International Conference on Spoken Language Processing*. Banff, AB.
- Smith, Bridget. (2009). Dental fricatives and stops in Germanic: Deriving diachronic processes from synchronic variation. In M. Dufresne, F. Dupuis, & E. Vocaj (Eds.), *Historical Linguistics 2007: Selected papers from the 18th International Conference on Historical Linguistics, Montreal, 6-11, August 2007*, 20-36, John Benjamins Publishing.
- Stuart-Smith, Jane. (2004). *Phonetics and philology: sound change in Italic*. Oxford University Press.

- Sumbly W.H. & I. Pollack. (1954) Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America* 26, 212-215.
- Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica*, 36, 314-331.
- Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip reading* (pp.3-51). London: Erlbaum.
- Tabain, Marie. (1998). Non-sibilant fricatives in English: Spectral Information above 10 kHz. *Phonetica*, 55(3), 107-130.
- Vatikiotis-Bateson, Eric, Inge-Marie Eigsti, Sumio Yano, & Kevin Munhall (1998). Eye movement of perceivers during audiovisual speech perception, *Perception and Psychophysics*, 60 (6), 926-940.
- Wang, Yue, Dawn M. Behne, & Haisheng Jiang. (2008). Linguistic experience and audio-visual perception of non-native fricatives, *Journal of the Acoustical Society of America*, 124 (3), 1716-1726.
- Wang, Yue, Dawn M. Behne, & Haisheng Jiang. (2009). Influence of native language phonetic system on audio-visual speech perception, *Journal of Phonetics*, 37, 344-356.
- Wells, J.C. (1982). *Accents of English, vols 1-3*. London: Cambridge University Press.
- Wolfram, Walt. (1994). The phonology of a socio-cultural variety: The case of African American Vernacular English. In John E. Bernthal and Nicholas W. Bankson (eds), *Child*

Phonology: Characteristics, Assessment, and Intervention with Special Populations, 227-244 . New York: Thieme Medical Publishers.