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

GRANT MCGUIRE* and MOLLY BABEL**

**University of California at Santa Cruz*
***University of British Columbia*

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 cue variability within and across talkers. 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/ 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, DiCanio, 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

JLP 3 (2012), 251–272
DOI 10.1515/lp-2012-0014

1868-6346/12/03–251

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 such that /f/ regularly replaces /θ/, but /θ/ is not a frequent substitution for /f/. We argue that a multi-modal model of speech perception, coupled with the heretofore under-analyzed visible articulatory variability of /θ/, can account for observed patterns. We demonstrate that /θ/ has more variable cues, and, following Chang, Plauché, and Ohala (2001), we propose that this further destabilizes /θ/.

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) go so far as to 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/.

A common pattern is /θ/-fronting,[1] where /θ/ > /f/. This pattern 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 English. In the U.S., many varieties of African American English are described as having this phenomenon, usually combined with /θ/-stopping (/θ/ > /t/; Labov et al. 1968; Wolfram 1994; 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 intervocalically and word-finally (Labov et al. 1968).

To explain this variation and 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: Greek-acquiring children learn /θ/ earlier than English-acquiring children. The comparatively earlier acquisition of /θ/ in Greek is thought to reflect the fact that /θ/ is more frequent in Greek (Edwards and Beckman 2008). The perceptual weakness account exploits the fact that /f/ and /θ/ are indeed highly confusable perceptually (e.g., Miller and Nicely 1955), due primarily to their spectral similarity (Tabain 1998). It has, therefore, been suggested that the perceptual weakness argument has much more credibility (Jones 2002).

A perceptual weakness argument would suggest this sound change is listener-driven. Under such an account for θ > f, the sound change originates in the misperception or misinterpretation of a talker's intended production (Ohala 1981, 1993; Blevins 2004). Specifically, the /θ/ > /f/ change can be seen as a classic example of a context-free sound change according to Blevins (2004: 134). This pattern, she argues, is a sweeping sound change, having occurred as a synchronic pattern in many dialects of English and Italian and cannot be explained in terms of "general articulatory variability." The primary deficiency of a perceptual weakness account, however, is that it does not provide an explanation for the asymmetry. If the sounds are confusable, either should be the target of change; this is not the case. In fact, there are no attested non-pathological accounts of /f/-backing in English that we are aware of. Moreover, /f/ > /θ/ changes are extremely rare cross-linguistically with only a handful of cases attested, none of which are as systematic as the much more common /θ/ > /f/ pattern, found in Germanic, Athabascan, and Oceanic languages, among others (Rice 1989; Blevins 2004; Smith 2009).

While it is undeniable that /f/ and /θ/ present a weak perceptual contrast, this explanation does not adequately describe why it is /θ/ that is consistently targeted in sound change. In this paper we present the argument that /θ/ is more variable than /f/ in North American English and that it is this variation that would lead to its lack of stability in language. We demonstrate this variation in /θ/ through analyses of productions from the 10 talkers whose voices are used as stimuli in the perception experiment described below. Crucially, this variability in /θ/ exists both within and across talkers and across visible articulatory measurements and acoustic measurements. It is our view that this variability in /θ/ makes it even more vulnerable to sound change compared to /f/.

That visual cues play a role in disambiguating /f/ and /θ/ is not a new concept. In fact, Miller and Nicely (1955: 347) note that the "distinction between |f| and |θ| and between |v| and |edh| are among the most difficult for listeners to hear and it seems likely that in most natural situations the differentiation depends more on verbal context and on visual observation of the talker's lips than it does on the acoustic difference (sic)." Such a view is also advanced by Jones (2002), who notes that there may be a functional load difference in the phonemes.

Jongman, Wang, and Kim (2003) explored these possibilities in two speech perception studies. With respect to the visual saliency claim, they played audio only, visual only, and audio-visual CV stimuli produced by a female talker to English perceivers whose task was to identify the phone presented. Results showed 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, though considerably less reliably so than audio only. While highly suggestive of the account we and others have proposed, one key feature lacking in this study is the assessment of how variability in articulation of /θ/ affects perception of the contrast. There is good reason to believe that variability is present and that this variability is important to the perception of this contrast.

Specifically, 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 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).

Our focus on the variability of /θ/ and not /f/ is not accidental. Labiodental fricatives do, of course, vary in terms of whether the teeth make exo- or endolabial contact, but we do not know literature on dialect variation of this sort. Some varieties of Spanish spoken in the Americas, however, exhibit variation in /f/ such that the labiodental fricative is realized as /ɸ/, /x/, or /h/ (Resnick 1975); none of these varieties of Spanish have /θ/.

While the claim that variability itself may make a sound pattern unstable is novel, there is evidence that some types of variation adversely affect speech processing. For example, it is well established that processing multiple voices increases response times and error rates (Mullennix, Pisoni, and Martin 1988; Martin, Mullennix, Pisoni, and Summer 1989; Goldinger, Pisoni, and Logan 1991). More concretely, listeners exhibit an increase in response time when presented with a voice that has more variability in a contrast than with a voice that makes the same contrast with more separation between the two categories (Newman, Clouse, and Burnham 2001). Increasing the variability of stop VOT led to an increase in uncertainty in word recognition (Clayards, Tanenhaus, Aslin, and Jacobs 2008), contributing to the notion that variation can have negative implications for speech processing. Directly addressing perceptual asymmetries in consonant perception, Chang, Plauché, and Ohala (2001) argue that listeners are more likely to miss an acoustic cue, as opposed to mistakenly perceiving its existence. They offer this explanation for why, for example, [k^hi] is often misidentified as [t^hi], but [t^hi] is rarely labeled as [k^hi]. Chang and colleagues reason that [k^hi] and [t^hi] are acoustically very similar, but that [k^{h}i], with a longer front cavity, has a compact mid-frequency spectral peak that [t^{h}i] lacks.[2] The asymmetry arises when listeners miss this acoustic cue in [k^{h}i] given the assumption that listeners are more likely to miss a perceptual cue than to fabricate its existence in [t^{h}i].

Summarizing, /f/ and /θ/ are highly similar and this similarity has been proposed as a way to account for /θ/ fronting to /f/, as in many other proposed listener-driven changes. This explanation cannot fully account for the asymmetry in the change and we propose that a bias towards /f/ originates in the greater visual saliency and stability of /f/. The utility of visual cues for disambiguating the contrast has been demonstrated before, but a convincing account has never been proposed for the pervasive θ > f bias in sound change. In the following experiment we address this hypothesis by replicating and extending Jongman and colleagues' (2003) study by using multiple talkers. Moreover, we 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).

2. General methods

2.1. *Audio and visual stimuli*

Five male and six female native speakers of American English with some phonetics training provided the audio-visual stimuli. The extra female subject (author MB) 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 /ɑ/, /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. Talkers 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. In spite of this coaching, talkers exhibited natural variation in the production of these vowels; for example, the formant dynamics of

/u/ varied according to the sibilant context (see Figures 7 and 8 below). VCV tokens were consistently produced by the eleven talkers with a H* accent on the initial vowel and L% on the second vowel. Audio was collected in this condition, but not used as an experimental stimulus.

Audio recordings were made in a sound-attenuated room immediately following the video session. Talkers 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) timed to present stimuli at approximately the same rate as in the previous elicitation. Audio and video recordings were synchronized using Adobe Premier Pro CS5. Initial synchronization was done such that the audio began with the first frame of visible articulation for each token. Unnatural alignments were corrected such that, ideally, the mid-point of audible frication was aligned with the central frame of visible articulation for each fricative.[3] While the participants in the following experiments were not asked explicitly about the naturalness of the tokens, they were asked in a post-experiment questionnaire, "Did the sounds seemed odd in any way?" The most common response to this was "no". The second most common answer was to note that the stimuli were generally not real words of English and/or that some sequences were phonotactically rare/nonexistent in English (e.g., [uθu]). A few subjects noted that some of the talkers had "accents." Only one subject commented in a way that indicates unnaturalness arising from the creation of the stimuli and responded, "Some were clearly edited and clipped." However, we take the overall preponderance of responses to indicate that the stimulus creation method did not introduce any problematic artifacts.

2.2. *Procedures, conditions, and participants*

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.

There were three experimental conditions: Audio, Visual, and Audio-Visual. In the Audio experiment only the audio stimuli were used. In the Audio-Visual experiment the synchronized audio-visual stimuli were used. For the Visual experiment the synchronized audio-visual stimuli were used, but the audio was muted. In all experiments the same female talker (author MB) 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: A total of 27 (18 female) undergraduates from the University of California at Santa Cruz participated. They were compensated with course credit for participation. Sixteen (8 female) participants were randomly sampled from this larger number for analysis in this study.
- Visual experiment: 16 undergraduates (10 female) from the University of British Columbia participated. They were compensated with $10 CAD for participation.
- Audio-Visual experiment: 16 undergraduates (9 female) from the University of California at Santa Cruz participated. They were compensated with course credit.

3. Results

3.1. *Condition and segmental effects*

Table 1 presents confusion matrices for the three conditions. 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

Table 1. *Confusion matrices for each condition.*

	Presented					
	Audio		Visual		Audio-Visual	
	f	θ	f	θ	f	θ
Responded f	3590	941	2980	775	3819	312
Responded θ	718	3366	529	2731	382	3864

Table 2. *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.*

	Sensitivity	Bias	Proportion correct
Audio only	1.43 (0.55)	−0.1 (0.44)	0.81 (0.10)
Visual only	1.35 (0.65)	−0.09 (0.47)	0.79 (0.13)
Audio-Visual	1.83 (0.6)	0 (0.31)	0.87 (0.1)

these results is approximately 3.65, and this can be considered near perfect perception. Table 2 reports summary statistics for each condition.

The overall design was a 2 (Fricative: /f θ/) × 3 (Vowel: /i a u/) × 3 (Syllable: CV, VC, or CVC) × 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 10 individual talkers were treated as variation for this purpose and are analyzed as a factor in the analysis reported in §3.2. 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 ($F[2,80] = 41.4$, $p < 0.001$), Vowel ($F[2,80] = 12.5$, $p < 0.001$), and Condition ($F[2,40] = 7.63$, $p < 0.01$) all returned as significant main effects. The ANOVA also revealed significant two-way Syllable × Condition ($F[4,80] = 7.00$, $p < 0.001$), Vowel × Condition ($F[4,80] = 82.30$, $p < 0.001$), and Syllable × Vowel ($F[4,160] = 5.96$, $p < 0.001$) interactions, as well as a three-way Syllable × Vowel × Condition interaction ($F[8,160] = 3.41$, $p < 0.01$).

The main effects of this analysis are shown in the three panels of Figure 1. Tukey tests were used to explore the levels within these main effects. The first panel of Figure 1 illustrates the main effect of Condition. Post-hoc Tukey tests found that subjects were more sensitive to the contrast in the Audio-Visual condition than in both the Audio and Visual conditions ($p < 0.001$). The difference between the Audio and Visual conditions was not significant. 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, shown in the middle panel, revealed that both the CV and the VCV environments were more salient than the VC environment (CV and VC: $p < 0.001$; VCV and VC: $p < 0.01$). Finally, post-hoc testing of the main effect of Vowel revealed that subjects were more sensitive to the contrast in the context of /u/ compared to /a/ ($p < 0.05$); the difference between /u/ and /i/ was not significant.

These main effects demonstrate two main points. First, the Condition results replicate the intuitive finding of Jongman et al. (2003) that presenting both the audio and visual channels simultaneously improves recognition of the stimuli. Second, our findings for the Syllable environments follow the pattern expected from the typological facts; subjects are least sensitive when these fricatives occur in coda position and this is where /θ/ alternation patterns are most common. Our

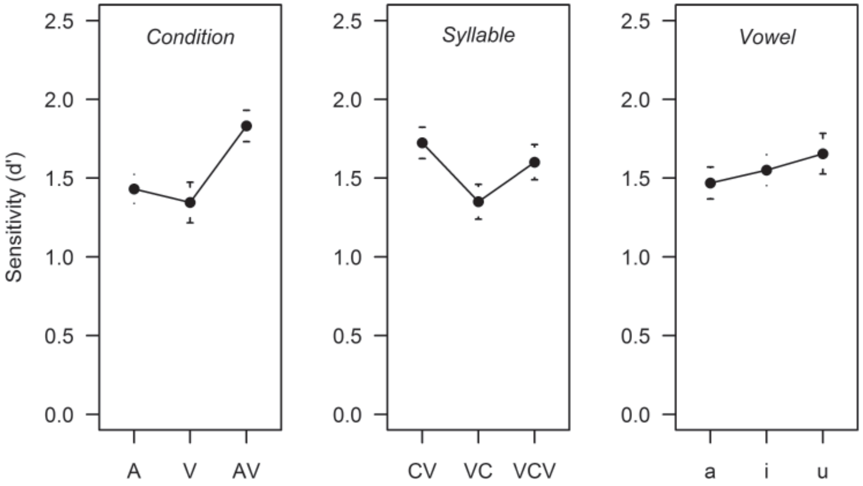

Figure 1. *Main effects of Condition (left panel), Syllable (middle panel), and Vowel (right panel). The plots present subjects'* d' *scores for each group within each factor. A higher* d' *value indicates that subjects were more sensitive to the f/θ contrast. Error bars represent 95% confidence intervals.*

results offer credence to the perceptual explanation for these patterns; 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. We argue below that this is perhaps due to talkers using a more dental, as opposed to interdental, articulation for /θ/ in the context of /u/. This place of articulation change leads to a higher F2 for /u/ in the context of /θ/. To our knowledge, this perceptual consequence is a new and unexpected 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 (alpha of 0.0125) showed that CV syllables were more salient than VC syllables in the Visual ($t[56] = 2.7947$, $p = 0.007$) and the Audio-Visual conditions ($t[87] = 4.1335$, $p < 0.001$), but not in the Audio Condition. The VCV environment was more salient than the VC environment for the Audio condition ($t[93] = 3.561$, $p < 0.001$). 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.

The Condition and Vowel interaction is shown in Figure 3 and demonstrates the highly divergent patterns across modalities. Post-hoc Tukey 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

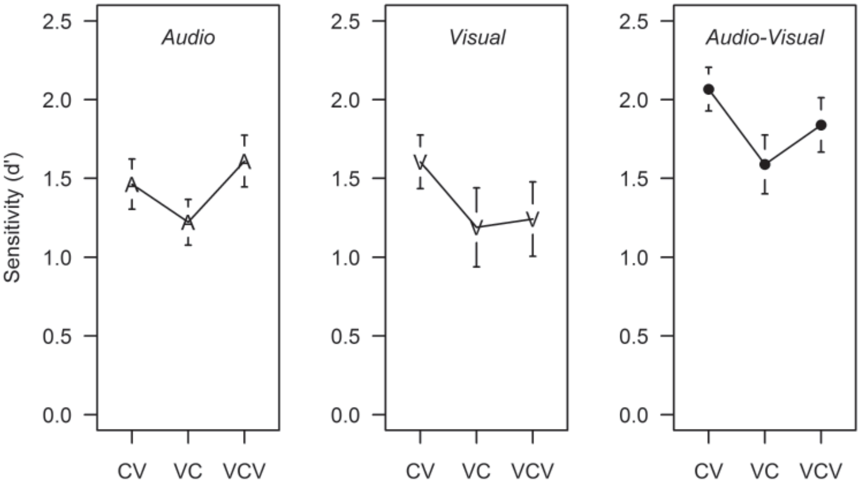

Figure 2. *Sensitivity by syllable position for each condition. The left panel is Audio only (A), the middle panel is Visual only (V), and the right panel is Audio-Visual (filled circle). Error bars represent 95% confidence intervals.*

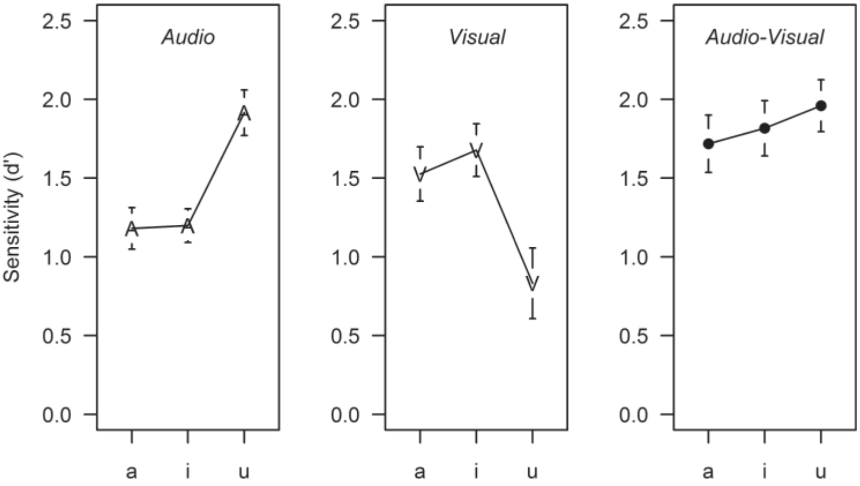

Figure 3. *Sensitivity by vowel environment for each condition. The left panel is Audio only (A), the middle panel is Visual only (V), and the right panel is Audio-Visual (filled circle). Error bars represent 95% confidence intervals.*

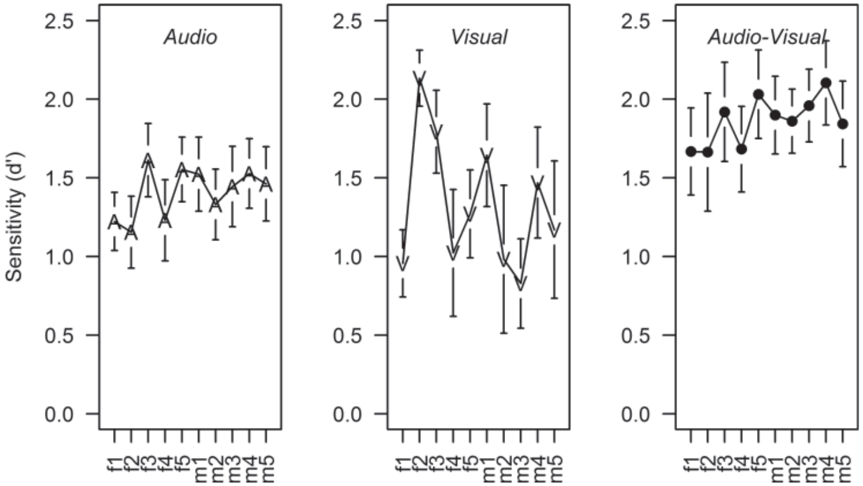

Figure 4. *Listener sensitivity to the f/θ contrast for each talker in each condition. The left panel shows the Audio (A) condition, the middle panel is the Visual (V) condition, and the right panel shows the Audio-Visual (filled circle) condition. Error bars represent 95% confidence intervals.*

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 subjects were the least sensitive. There were no differences across the vowel contexts for the Audio-Visual condition.

3.2. *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[9,342] = 9.74, $p < 0.001$) and Condition (F[2,38] = 8.01, $p < 0.01$) returned as significant. The interaction between Talker × Condition was significant as well (F[18,342] = 6.94, $p < 0.001$). 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 pattern in response to each talker is 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.

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 (corrected to an alpha of 0.0125). In the Audio condition, talkers f3 and f5 are significantly more salient than f2 (f3 and f2: $t[30] = 2.98$, $p = 0.003$; f5 and f2: $t[29.63] = 2.76$, $p = 0.005$). However, in the Visual condition, talker f2 provides the most salient contrast. Note that in the Audio and Audio-Visual conditions, f2 is one of the least salient talkers. In the panel illustrating the Audio-Visual differences by talker, the error bars overlap a great deal. There were no differences between any pair of talkers in the Audio-Visual condition, demonstrating that with both audio and visual channels available to subjects, all talkers produce an equally distinguishable contrast as measured by *d′*, statistically speaking.

The high degree of variability among talkers in the visual condition is of interest here. It is well known that English speakers can vary in the production of /θ/ with both dental and interdental types being observed (Catford 1982; Ladefoged and Maddieson 1996). For /f/, there is the possibility of endo- and exo-labial variation present. In order to capture the variation present, the videos were analyzed as follows. All /θ/ productions were classified as either dental (0) or interdental (1). All /f/ productions were scored as endolabial (0) or exolabial (1). Lip movement was quantified by measuring relative distances on facial landmarks for all fricative productions. To this end, the frames corresponding to the midpoint of the vowel (first vowel in VCV syllables) and the midpoint of the fricative were extracted from each token and four vertical measurements (in pixels) were taken from along the center-line of the face: (1) tip of the nose, (2) bottom of the upper lip, (3) top of the lower lip, (4) bottom of the chin. From these measures relative distances were calculated between the lower and upper lip and the lower lip and the nose. Each distance for the consonant portion of each token was subtracted from the vowel portion to quantify the amount of movement for the lips in each production and then averaged for each talker by fricative. This provides relative measures for each talker's variability in terms of distance between lips and movement of the lower lip relative to the nose (a static landmark).

The distance measures and interdentality coding were then used to predict the perceivers' sensitivity to each talker. The exo- and endolabial measures were not included as all talkers produced all /f/s as endolabials, except for a single token from M2. Stepwise linear regression models were fitted for each condition (A, V, AV), with each having talker *d′* (varying by condition) as the dependent variable. The predictor variables were degree of inderdentality (Inter), change in inter-lip distance of /f/ (fLipDist), change in inter-lip distance /θ/ (thLipDist), change in lower lip position for /f/ (fLLip), and change in lower lip position for /θ/ (thLLip). Only the model for the V condition was significant ($F[2,7] = 21.3$, adjusted $r^2 = 0.82$, $p < 0.01$) and its results are in Table 3.

Table 3. *Predictor variables for Visual stepwise linear regression model. Inter refers to the interdentality of /θ/ and fLLip refers to the change in inter-lip distance of /f/. Note that while fLLip was selected by the model, its contribution is not significant.*

	Estimate	Std. error	*t*-value	*p*-value
(Intercept)	0.56	0.14	3.89	<0.01
Inter	0.93	0.24	3.8	<0.01
fLLip	0.03	0.01	2.29	0.06

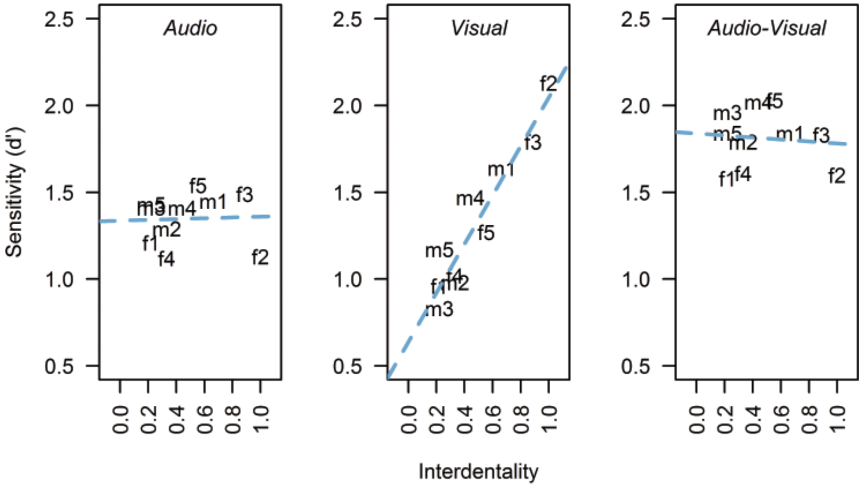

Figure 5. *The average /θ/ interdentality gesture score for each talker plotted against average* d′ *sensitivity values for each talker by condition. An interdentality score of 0 indicates that the talker's tongue was never visible during /θ/ production and a score of 1 indicates it was always visible.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

The results of the visible articulatory measurements point towards an interesting conclusion. The primary driver of the effect was the interdentality of a talker's /θ/s – the more frequently interdental a talker's productions were, the more sensitive viewers were to that talker's contrasts. This is apparent in Figure 5 where there is a strong relationship between the interdentality measure and listener sensitivity in the Visual Condition (the middle panel), but not in the Audio or Audio-Visual Conditions. A secondary effect is the amount of lower lip movement. Here the greater the amount of movement between /f/ production and the vowel, the more sensitive the perceivers were. While neither of these results may be particularly unexpected, it does demonstrate that there is talker variation in the visible productions of these fricatives. This variation is especially notable in /θ/, and there are saliency consequences as a result.

4. Acoustic characteristics of the audio stimuli

In the previous section we demonstrated articulatory variability across our 10 talkers which contributed to variation in subjects' sensitivity to the f/θ contrast for the talkers. In this section we report on acoustic qualities of both the fricatives and the vowels in our stimuli.

4.1. *Fricative acoustics*

A post-hoc analysis of the frication noise was done to establish its role in the results. This analysis generally followed those reported by Gordon et al. (2002) and Tabain (1998). Measures taken included all four spectral moments (mean, deviation, skewness, kurtosis) and duration. All tokens were high pass filtered at 10 kHz[4] (following Tabain 1998), and the spectral measures were calculated using FFT power spectra having 1,024 points, centered on the middle of the fricative with a Gaussian window. Separate ANOVAs were run with each of the measures, each having the factors Fricative and Talker. In this analysis, the interaction between Fricative and Talker would point to differences by Talker for each fricative. Table 4 summarizes the findings.

Table 4. *ANOVA summary tables for the four spectral moments and duration.*

Measure	Factors	Df	Sum Sq.	Mean Sq.	*F*	*p*
Centroid	Fricative	1	2,581,879	2,581,879	6.99	<0.01
	Talker	9	31,175,338	3,463,926	9.38	<0.001
	Fricative*Talker	9	5,870,841	652,316	1.77	0.09
	Residuals	160	59,067,173	369,170		
Standard Deviation	Fricative	1	125,252	125,252	1.34	0.25
	Talker	9	6,056,634	672,959	7.19	<0.001
	Fricative*Talker	9	884,908	98,323	1.05	0.4
	Residuals	160	14,976,476	93,603		
Skewness	Fricative	1	0.3	0.3	1.63	0.2
	Talker	9	14.68	1.63	8.76	<0.001
	Fricative*Talker	9	2.4	0.27	1.43	0.18
	Residuals	160	29.8	0.19		
Kurtosis	Fricative	1	7.91	7.91	4.97	<0.05
	Talker	9	92.77	10.31	6.48	<0.001
	Fricative*Talker	9	19.2	2.13	1.34	0.22
	Residuals	160	254.69	1.59		
Duration	Fricative	1	0	0	0.11	0.74
	Talker	9	0.17	0.02	7.55	<0.001
	Fricative*Talker	9	0.01	0	0.46	0.9
	Residuals	160	0.39	0		

As can be seen in Table 4, no interactions rose to significance at $p < 0.05$. The fricatives were significantly different when summarized across all talkers for Centroid (/f/ = 12,928, /θ/ = 12,689) and Kurtosis (/f/ = 0.696, /θ/ = 1.12), but no other measures returned with a main effect for fricative. The talkers varied significantly for each measure. Thus, we found talker differences in the overall acoustics of the productions, two measures that separated the fricatives across talkers, but none of the measures captured the talker specific differences in contrast (if, indeed, there are any).

4.2. *Vowel acoustics*

It has long been recognized that the adjacent vowel transitions provide valuable information about the identity of nonsibilant fricatives (Harris 1958). To understand the role of these formant transitions in subjects' performance, we analyzed the first and second formant frequencies across the duration of the vowels. F1 and F2 of the vowels were extracted from a series of Gaussian windows with a 2.5 ms step size. This was done 5%, 10%, 25%, 50%, 75%, 90%, and 95% of the way through each vowel. Outliers were hand-corrected and values were converted to the Bark scale. Figure 6 presents these values for the VC context and the initial vowel of the VCV sequence. Figure 7 provides the formant data for the second vowel of the VCV sequence and the CV environment.

In the analysis above, we found an unanticipated connection between sensitivity to the fricative contrast and vowel environment. Specifically, the /u/ environment

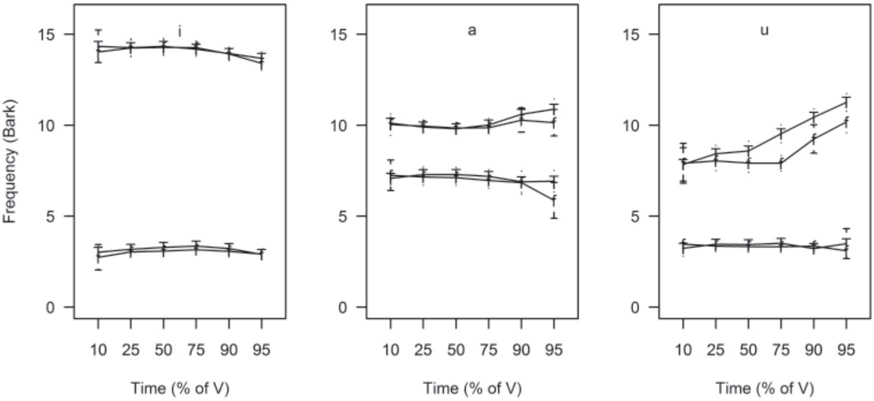

Figure 6. *Bark-scaled first and second formant transitions for the vowel preceding the fricative, collapsed across both VC and the initial vowel in VCV contexts. The Time label on the x-axis presents formant frequency averages at normalized time points throughout the vowel. Time point 100%, for example, would be exactly at the vowel-consonant boundary. In the figures, "T" = /θ/ and "f" = /f/. Error bars represent 95% confidence intervals.*

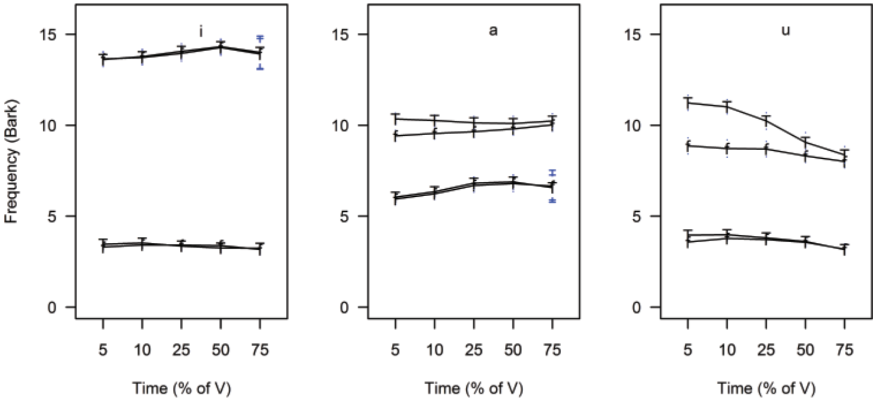

Figure 7. *Bark-scaled first and second formant transitions for the vowel following the fricative, collapsed across both CV and the second vowel in VCV contexts. The Time label on the x-axis presents formant frequency averages at normalized time points throughout the vowel. Time point 0%, for example, would be exactly at the consonant-vowel boundary. In the figures, "T" = /θ/ and "f" = /f/. Error bars represent 95% confidence intervals.*

was found to produce the highest listener sensitivities for the Audio condition and the lowest for the Visual condition. This seems to be related to substantial variation in the production of /θ/ across different vowel environments. 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 subjects are making use of this information in the Audio condition. While the /u/ context was an enhanced environment for the Audio condition, it had the lowest sensitivity scores in the Visual condition. While it is possible that the lip-rounding of /u/ effectively obscures any visible tongue gesture subjects would take as evidence of a /θ/ production, the high F2 onset of /θ/ in this context suggests that subjects struggled to accurately identify /θ/ in these instances because it was in fact a dentally, and not interdentally, articulated /θ/ which did not offer the necessary visual cues to make a /θ/ decision. What is interesting is that for this /u/ context in which subjects have access to both the useful acoustic information and the uninformative visual information in the Audio-Visual condition, the pattern of listener responses in the /u/ environment more closely resemble that of the Audio condition. Subjects are able to disregard the uninformative visual cues in the process of labeling a stimulus as /f/ or /θ/ in the Audio-Visual design.

The acoustic motivations for the Syllable context effects are also observable from the formant data in Figures 6 and 7, 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 subjects 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, our results also indicate that descriptions of more /θ/-fronting and /θ/-stopping in coda positions (Wells 1982; Dubois and Horvath 1998) may be rooted in increased misperception in this environment.

5. Discussion and conclusion

The goal of this study was to examine the role of visual information in the perception of /f/ and /θ/ across multiple talkers in an attempt to understand how visual information may be implicated in sound change involving the two sounds. 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, while 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 10 talkers who served as stimuli differ in their use of a visible tongue gesture. These talkers span a range from never using an interdental tongue gesture and receiving an interdentality score of 0.0 (e.g., talker m2) to always displaying a visible tongue gesture (e.g., talker f2) and being scored with an interdentality measure of 1.0. This variation negatively affected subjects. 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). As expected from descriptive work about the environments under which we find more θ-fronting or θ-stopping, subjects were least sensitive to the f/θ contrast in VC syllables. Our perception results are specific to a population which has limited exposure to /θ/ substitution patterns. We note that our model talkers and subjects are from populations that speak varieties of North American English with categorically stable /θ/ realizations. Our population likely has some exposure to African American English, which exhibits both /θ/ > f and θ-stopping patterns, but we do not believe the extent of this exposure is very great.

The series of experiments described above offers evidence that /θ/ is a particularly unstable segment because of the asymmetry in the reliability of the auditory and visual cues provided by 10 speakers. This hypothesis is an improvement over previous phonetically-motivated arguments for the θ > f bias. The 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, such an observation makes no headway into explaining why the sounds behave asymmetrically in terms of diachronic phonological change. 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 MacDonald 1976; Summerfield 1979), and provides an explanation for why the bias in perception would be targeted to /θ/.

An analysis of our stimuli indicated that /θ/ production is more variable both within and across-talkers. Within talkers, /θ/ production varies across vowel environments; in /u/ contexts, the formant transitions suggest that /θ/ is articulated more dentally. This is useful in the Audio and Audio-Visual contexts where the more dental articulation gives way to an F2 transition that is higher than the lower F2 transition for /f/; subjects exhibited heightened sensitivity in this context. This /u/ environment became the most detrimental for subjects in the visual condition as subjects seemed to rely on the interdentality of the /θ/ to make their decision and in a dental articulation the visible tongue gesture needed to log a /θ/ response was not present. Across talkers, /θ/ is most variable because, in addition to this context-specific vowel pattern of interdentality, talkers vary with respect to how dental or interdental their productions of /θ/ are. This fact leads to the large amount of variability in talker sensitivity in the Visual Condition.

Variability for /θ/ was found in the acoustic and visible articulatory domains. We propose that it is this feature of /θ/ – 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 – speakers, of course, are constantly facing an infinitely variable signal – variability and category overlap within talkers has been shown to degrade perceptual performance (Newman et al. 2001). Clayards and colleagues (2008) find that wider probability distributions give way to increased perceptual uncertainty. In the case of f~θ, speakers must be prepared to face both unpredictable cross-talker variability in /θ/ production along with the perhaps slightly more predictable within-talker variability in /θ/ production across vocalic 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. In the face of this variability, it is likely that speakers will miss the information necessary to log a /θ/ response. Speakers are more likely to miss a cue than to fabricate the existence of a cue (Chang et al. 2001), and will therefore identify the missed /θ/ as an /f/.

Of course, an /f/ bias in English could be accounted for by the fact that /f/ is nearly 5 times as frequent as /θ/ (SUBTLEX$_{US}$; Brysbaert and New 2009). While this frequency difference may bias English listeners to misperceive /θ/ as /f/ more than the reverse, this frequency effect cannot account for the cross-linguistic pattern that θ > f, while cases of f > θ are vanishingly rare. As overviewed in the introduction, however, given the fact that descriptions of /θ/ production differ across dialects and languages, it is likely that acoustic and visible articulatory variability exists in all languages with /θ/. This variability in the auditory and visual domains for an acoustically non-robust sound will allow for its instability in the sound systems across languages.

One question we can ask is why not have articulatory stability in the production of /θ/? Why would a system allow /θ/ to vary in these ways? Our response to such a question addresses the within-talker and cross-talker variability in turn. The within-talker variability is contextual variability conditioned by vowel environment. While languages vary in the extent to which coarticulatory properties carry over (Hombert et al. 1979; Manuel 1990; Beddor et al. 2002), we simply expect some variability in /θ/ due to the vocalic context. The cross-talker variability is an example of the type of articulatory difference that results in a small acoustic difference; for a contrast that does not carry a very high functional load, it is likely that listeners rarely take notice. Within- and cross-talker variability of this sort has been especially well documented for North American English /ɹ/ (e.g., Delattre and Freeman 1968).

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. Even in degraded listening conditions, however, listeners spend more time gazing at a talker's eyes than their lips. 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. Speakers use visual articulatory cues when necessary, but do not exhaustively exploit the visual signal. The presence of visual information does have subtle effects on phonetic systems. For example, 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 (Ménard et al. 2009). We argue that visual information also has subtle effects on the shape of sound inventories across time.

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. We would suggest that multi-modal phonetic cues also be considered in discussions of acquisition and typological distributions of sounds.

Acknowledgments

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, Michael McAuliffe, Ronald Sprouse, Devin Tankersley, Eric Vatikiotis-Bateson, and, especially, Travis Whitebread. This work has been inspired by discussions with Keith Johnson, Andrew Garrett, and John Ohala. This paper was improved immeasurably by the comments from two anonymous reviewers and James Scobbie. A special thank you goes to our research participants as well.

Correspondence e-mail address: gmcguir1@ucsc.edu

Notes

1. As noted by Jones (2002), this term is less than ideal as the "fronting" requires a completely different set of articulators rather than a single articulation that is moved forward. However, we use this as it is the most common in the literature.
2. We are not advocating a model of speech perception where listeners attend solely to singular cues in the process of recognizing meaningful linguistic units in the speech signal. However, individual acoustic and visible articulatory cues often provide valuable information about a sound's identity.
3. Examples from the practice talker are available here: http://people.ucsc.edu/~gmcguir1/McGuireBabelAV/.
4. An identical analysis was run with a 100 Hz high-pass filter with similar results. In that analysis, only Talker effects were significant for all measures (all $p < 0.01$) and Fricative was significant for Kurtosis ($p < 0.05$).

References

Beddor, Patrice Speeter, James D. Harnsberger, & Stephanie Lindemann. 2002. Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics* 30(4). 591–627.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: 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.

Brysbaert, Marc, & Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41(4). 977–990.

Catford, John Cunnison. 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. New York: Academic Press.

Clayards, Meghan, Michael Tanenhaus, Richard Aslin, & Robert Jacobs. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108(3). 804–809.

Delattre, Pierre, & Freeman, Donald C. 1968. A dialect study of American *r*'s by x-ray motion picture. *Linguistics* 44. 29–68.

Dubois, Sylvie, & Barbara 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.

Fujimura Osamu, Marian 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.

Goldinger, Stephen D., David B. Pisoni, & John S. Logan. 1991. On the nature of talker variability effects in recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* 17(1). 152–162.

Gordon, Matthew, Paul Barthmaier, & Kathy Sands. 2002. A cross-linguistic acoustic study of fricatives. *Journal of the International Phonetic Association* 32. 141–174.

Harris, Katherine Safford. 1958. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech* 1. 1–7.

Hombert, Jean-Marie, John J. Ohala, & William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55. 37–58.

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*, 529–561. http://linguistics.berkeley.edu/phonlab/annual_report/documents/2007/JohnsonDicanioMackenzie.pdf

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, 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.

Kjellmer, Göran. 1995. Unstable fricatives: On Gothic *þliuhan* 'flee' and Old English *flēon* 'flee'. *Word* 46. 207–223.

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.

Macmillan, Neil A., & C. Douglas Creelman. 2005. *Detection Theory: A User's Guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Manuel, Sharon. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America* 88. 1286–1298.

Martin, Christopher S., John W. Mullennix, David B. Pisoni, & Walter V. Summers. 1989. Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(4). 676–684.

McGurk, Harry, & John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264. 746–748.

Ménard, Lucie, Sophie Dupont, Shari Baum, & Jérôme 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.

Miller, George A. & Patricia E. Nicely. 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27(2). 338–352.

Mullennix, John W., David Pisoni, & Christopher S. Martin. 1989. Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America* 85(1). 365–378.

Newman, Rochelle S., Sheryl A. Clouse, & Jessica L. Burnham. 2001. The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America* 109(3). 1181–1196.

Ohala, John J. 1981. The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick, & Mary Frances 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 John Kingston & Mary 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.

Resnick, Melvyn. 1975. *Phonological Variants and Dialect Identification in Latin American Spanish*. Mouton: The Hague.

Rice, Keren. 1989. *A Grammar of Slave*. Berlin: Mouton de Gruyter.

Rickford, John R. 1999. *African American Vernacular English: Features, Evolution, Educational Implications*. Malden, MA: Wiley Blackwell.

Schneider, Walter, Amy Eschman, & Anthony Zuccolotto. 2007. E-Prime: User's Guide, Version 2.0. Psychology Software Tools.

Shadle, Christine H., Andre Moulinier, Christian U. Dobelke, & Celia Scully. 1992. Ensemble averaging applied to the analysis of fricative consonants. Paper presented at 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 Monique Dufresne, Fernande Dupuis, & Etleva Vocaj (eds.), *Historical Linguistics 2007: Selected papers from the 18th International Conference on Historical Linguistics, Montreal, 6–11, August 2007*, 20–36. Amsterdam: John Benjamins Publishing.

Summerfield, Quentin. 1979. Use of visual information for phonetic perception. *Phonetica* 36. 314–331.

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 & Nicholas W. Bankson (eds.), *Child Phonology: Characteristics, Assessment, and Intervention with Special Populations*, 227–244. New York: Thieme Medical Publishers.