

JOURNAL ARTICLE

# On the nature of the perception-production link: Individual variability in English sibilant-vowel coarticulation

Alan C. L. Yu

Phonology Laboratory, Department of Linguistics, University of Chicago, Chicago, IL, US  
[aclyu@uchicago.edu](mailto:aclyu@uchicago.edu)

---

This study aims to elucidate the nature of the perception–production link with respect to coarticulation by examining the production and perception of English sibilants before different vowels. A group of native speakers of American English were recorded reciting a set of /s/- and /ʃ/-initial words in different vocalic contexts and took part in an identification experiment designed to test their ability to adjust their perceptual expectation in light of the vocalic influence on the preceding sibilant. Significant correlations between the production and perception results were observed when by-subject estimates for context-relevant predictors (and their interactions) in the perception regression models were examined in relation to the by-subject estimates of the production regression models. These results suggest a positive correlation between how much an individual attends to context-specific variation in perception and how the sibilant contrast is realized in specific vocalic contexts. Ramifications of these findings are discussed for the nature of speech perception and production and the understanding of sound change.

---

**Keywords:** Perception-production link; perceptual compensation; coarticulation; sibilant; rounding; English

---

## 1 Introduction

Understanding the nature of individual variation in speech perception and production is increasingly important, particularly for research on sound change and propagation (Stevens & Harrington, 2014). Various scholars have argued in recent years that sound change actuation might come about as a result of interactions between individuals with different perceptual and/or articulatory targets for the ‘same’ sound category (Baker, Archangeli, & Mielke, 2011; Yu, 2013, 2016) or different tendencies to attach social meaning to linguistic differences (Garrett & Johnson, 2013). Beddor (2009), for example, argued that listeners can be accurate perceivers who attend to coarticulatory information available to them in the input signal but nonetheless have different perceptual weightings (or phonological grammars) in terms of how they use coarticulation to signal the presence (or not) of the coarticulation trigger in the signal. Various scholars, most prominently Ohala (1993b), have argued that listeners who fail to compensate for coarticulatory effects properly would lead to sound change. For example, oral vowels in nasal contexts (e.g., VN sequences) might be mistakenly reconstructed as nasal (e.g.,  $\tilde{V}N$ ) if listeners fail to take into account the effects of coarticulatory nasalization. Likewise, Beddor and Krakow (1998) suggested that if vowels in nasal contexts are perceived as partially nasalized, listeners might fail to disambiguate fully the spectral contribution of nasalization and tongue/jaw position in both contextual and distinct nasal vowels, fostering vowel height

shifts in both. Extending the logic that hypo- or hyper-corrective sound changes are due to listeners failing to properly take contextual information into account, some scholars have argued that listeners who compensate less, or not at all, for coarticulation could be more likely to initiate context-dependent sound changes (Yu, 2010, 2016; Yu & Lee, 2014). Identifying the underlying nature of individual variability in the perception and production of coarticulated speech might help explain why certain sound changes happen at all and, conversely, why sound changes are so rarely actuated even though the phonetic pre-conditions are always present in speech.

One common avenue to explore the nature of individual variation in speech perception and production is to examine how the two might be related. Studies using paradigms such as altered auditory feedback (e.g., Houde & Jordon, 2002; Katseff, 2011; Shiller, Sato, Gracco, & Baum, 2009) and phonetic imitation (e.g., Babel, 2012; Nielsen, 2011; Yu, Abrego-Collier, & Sonderegger, 2013) have generally found speakers to be quite adept at adjusting their production patterns in face of adjusted perceptual feedback or altered perceptual experiences, suggesting a close link between speech perception and production. Within the context of the perception and production of coarticulated speech, some studies have reported positive mapping between perception and production. Beddor and Krakow (1999), for example, compared the patterns of perceptual compensation for vowel-nasal coarticulation among English and Thai listeners and found that Thai listeners compensated for vowel-nasal coarticulation less than the English listeners (i.e., Thai listeners were more accurate in detecting vowel nasality in contexts where perceptual compensation is expected to reduce sensitivity to the presence of vocalic nasalization). They explained this difference by appealing to the fact that nasal coarticulation in Thai is less extensive than in English. Thai listeners who experience smaller degrees of contextual nasalization on a regular basis might come to expect less nasalization (and conversely be more sensitive to an unexpectedly high degree of nasalization) in the appropriate contexts. A similar cross-linguistic correlation was observed in the perception and production of vowel-to-vowel coarticulation (Beddor, Harnsberger, & Lindemann, 2002). Zellou and colleagues have recently demonstrated that American English speakers are able to imitate different degrees of vocalic nasalization from their own, suggesting further a positive correlation between perception and production (Zellou, Dahan, & Embick, 2017; Zellou, Scarborough, & Nielsen, 2016).

However, the correlation between the production and perception of coarticulation is not always consistently observed. Harrington, Kleber, and Reubold (2008), for example, found an age-based correlation between /u/-fronting and listeners' perceptual compensation for this fronting effect. Younger speakers of Southern British English compensated less for the /u/-fronting effect compared to the older speakers who exhibited stronger context-specific /u/-fronting. However, it is unclear how these findings inform the perception-production link concerning coarticulation since listeners who compensate less (i.e., the younger group) also have more fronted /u/ regardless of context. Thus, it is not clear if /u/ is less coarticulated in the fronting context or it is simply produced as more front in general. It also remains unknown as to whether speakers of the same age range exhibit similar perception-production link concerning /u/-fronting. Kataoka (2011), who examined /u/-fronting in alveolar contexts in speakers of California English between the age of 19 to 45, found no significant correlation between the production and perception of /u/-fronting. Others focused on long-distance vowel-to-vowel coarticulation and found that the magnitude of long-distance vowel-to-vowel coarticulation does not correlate with individuals' ability to discriminate coarticulated vowels in isolated contexts (Grosvald, 2009; Grosvald & Corina, 2012). However, since their perceptual measures did

not specifically test for the listeners' ability to utilize knowledge of coarticulation, the findings of Grosvald (2009) and Grosvald and Corina (2012) remain inconclusive. That is, there are no theoretical reasons to think that there should be a relationship between the ability to discriminate vowels (coarticulated or otherwise) in isolation (i.e., contextual information is not given) and to produce vowels in context. Thus the lack of a relationship between the results of the production and perception tasks reported in Grosvald (2009) and Grosvald and Corina (2012) does not constitute negative evidence against the relationship between the perception and production of coarticulation *per se* since they did not provide any independent measures of the listener's ability to perform perceptual readjustment to coarticulated speech. In a recent study, Zellou (2017) investigated the correlations between individual differences in the production of nasal coarticulation and patterns of perceptual compensation in American English based on the results of a production task, a paired discrimination task, and a nasality rating task. Individuals who produce less extensive anticipatory nasal coarticulation exhibit more veridical acoustic perception (indicating less attention to potential sources of coarticulation in the signal) than individuals who produce greater coarticulatory nasality in the paired discrimination task. However, listeners' nasal coarticulation in production did not predict results from the rating task. This inconsistency suggests potential task sensitivity in reifying the perception-production link. Clearly, further investigations into the perception-production link with regards to coarticulated speech are needed.

The perception-production link might be further complicated by other mediating factors. For example, studies concerning the processing of coarticulated information have found that coarticulatory information not only affects the classification and discrimination of speech, but also the temporal dynamics of speech processing (e.g., Beddor, Coetzee, Styler, McGowan, & Boland, 2018; Beddor, McGowan, Boland, Coetzee, & Brasher, 2013; Dahan, Magnuson, & Tanenhaus, 2001; Mahr, McMillan, Saffran, Weismer, & Edwards, 2015). The perception-production correspondence might be affected by individual variability in their general processing skills and styles. Beddor et al. (2018), for example, found that participants who produced an earlier onset of coarticulatory nasalization on vowels were more efficient users of nasality as listeners as that information unfolds over time. Furthermore, previous literature on the perception-production link outside of the coarticulated speech domain also suggests that the link might be more nuanced. For example, the distinctness of an individual's production of a contrast has been found to correlate with how well the individual discriminates that contrast (Newman, Clouse, & Burnham, 2001; Perkell, Guenther, et al., 2004; Perkell, Matthies, et al., 2004). However, it remains unclear how contrast distinctness might play a role in the perception and production of coarticulated speech. That is, how does coarticulation influence the context-specific distinctness of segmental contrasts in production and perception? To this end, the present study provides an opportunity to address this gap in the literature.

Current theories of the perception and production of coarticulated speech differ in their predictions on how the two may be related. Theories of perceptual compensation that see phonetic perception as a matter of simple auditory function (e.g., Lotto & Holt, 2006) predict a relationship whereby speech production is concerned with uncovering strategies for producing optimally perceivable acoustic signals. As such, the mapping between the two is imprecise due to the non-uniqueness of the mapping between the acoustic targets and vocal tract configurations. However, given that speech articulation is hypothesized to serve perceptual goals, one might expect the listeners who perceive the presence of coarticulation information in the speech signal to also produce speech that achieves a similar degree of coarticulation. From this perspective then, auditory theories would

predict a positive correlation between the perception and production of coarticulated speech. However, the nature of the production rendered could vary as long as the acoustic outputs achieve the desired percepts that resemble coarticulation.

Gestural theories of speech perception, such as the Motor Theory (Lieberman & Mattingly, 1985) and Direct Realism (Fowler, 2006), see speech perception as being guided by the recovery of gestures in the underlying signal. Such gestural knowledge might stem from an 'innate vocal tract synthesizer' (Lieberman & Mattingly, 1985) or some presumed universal function of perceiving in the world, as in the case of Direct Realism. Due to the universal nature of the presumed gestural knowledge, which helps explain why perceptual compensation is not language specific or unique to humans (Viswanathana, Magnusson, & Fowler, 2010), such theories generally provide little insights into the nature of individual variability in perceptual compensation. To the extent that variation is acknowledged, they are attributed to differences in listening modes (e.g., Fowler & Brown, 2000; Repp, 1981, and more discussion below). If the universalist assumption is relaxed and an individual's knowledge of coarticulation derives from his/her coarticulatory habits, gestural theories would predict that the magnitude of coarticulation in production should mirror the magnitude of perceptual compensation, as the objects referenced in perception and production are one and the same (i.e., phonetic gestures of the vocal tract).

A similarly direct link between the production and perception of coarticulation can be found in exemplar-based approaches to speech perception and production. Pierrehumbert (2002), for example, posits an explicit perception-production loop, where stored perceptual experiences are weighted by social and attentional factors and such perceptual exemplars are drawn upon to generate production targets. Context-specific perceptual experiences would lead to context-specific realizations of production targets. Sonderegger and Yu (2010) laid out an explicit model along this line, showing in particular that the listeners' perceptual compensatory responses to vowel-to-vowel coarticulation can be modeled effectively using a rational (in this case, Bayesian) model of speech perception. Crucially, the model assumes as inputs context-specific acoustic cues, such as means of F1 and F2 of the target vowels in different vocalic contexts in modeling the perception of vowel-to-vowel coarticulation. Thus, the acoustic measures, which were derived from a production study and presumably reflected the production targets of the speakers, helped explain the perceptual behaviors of the listeners. However, since the acoustic measures were not drawn from the same individuals who participated in the perceptual task, it is unclear if the perception-production link is evident at the individual level.

The approaches reviewed thus far assume that speech perception is veridical and that perceptual compensatory behaviors can be fruitfully modeled given proper articulatory/acoustic information and *vice versa*. There are, however, models that are more 'non-veridicality-centric,' that is, they assume that the input signal could be recoded before perception is registered. The so-called 'C-CuRE' approach to perceptual compensation for coarticulation (Cole, Lindebaugh, Munson, & McMurray, 2010; McMurray & Jongman, 2011), as well as models of sound change that rely on listener misperception as a driving force behind certain sound changes predict a negative correlation between the perception and production of coarticulated speech. C-Cure, for example, assumes that the incoming acoustic cues are initially encoded veridically, but cues are recoded (hence non-veridically stored) in terms of their differences from expected values, which can be specific to particular individuals, as different sources of variance are categorized. This approach predicts that individuals who engaged in expectation adjustment robustly in perception (i.e., those who compensate more



robustly) would have production targets of a sound category that have low variance and are relatively context-free (less coarticulated). The same sound category might conversely have a more diffused distribution and the production targets are more context-dependent (more coarticulated) for individuals who do not adjust for contextual information robustly in perception (i.e., they are more veridical in perception, perhaps similar to Repp's 1981 auditory listeners; see discussion below).

Finally, as noted in Whalen (1999), some see no necessary connection between perception and production. Perception can proceed with no knowledge of production, as is the typical position in the automatic speech recognition (ASR) literature (see Livescu, Jyothi, & Fosler-Lussier, 2016, for an alternative view).

The importance for understanding the link between speech perception and production is all the more relevant to language and speech researchers in light of an increasing number of reported cases of variability in the perceptual compensation for coarticulation across individuals. Beddor (2009), for example, found a great deal of individual variability in the perception of nasalization in VNC sequences in American English. She suggested that the variability might stem from differences in perceptual grammar across individuals. Mann and Repp (1980) were first to report individual variability in perceptual identification of English sibilants in different vocalic contexts. Repp (1981) explained the individual variation in terms of two different strategies for listening to sibilant-vowel sequences. Some listeners are what he referred to as auditory listeners, who segregate the noise portion from the vocalic portion, while the others are phonetic listeners, where sibilant noise information is more integrated with the vocalic portion. Based on their findings, Yu and Lee (2014), who also observed individual variation in perceptual compensation for sibilant-vowel coarticulation, argued that the observed individual variability is more continuous than suggested by Repp's (1981) two-listening-mode model. Yu (2010) argued that the magnitude of perceptual compensation for the vocalic effect on sibilant perception might be modulated by the listener's gender and autistic-like traits, as measured by the Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001); the Autism Spectrum Quotient is a short, self-administered scale for identifying the degree to which any individual adult of normal IQ may have traits associated with Autism-Spectrum Disorder (ASD), where clinical diagnosis involves difficulties in social development and communication, alongside the presence of unusually strong repetitive behavior or 'obsessive' interests (American Psychiatric Association, 2013; I.C.D-10, 1994). In particular, he found females with a lower AQ score (i.e., less autistic-like) to be less likely to perceptually compensate for coarticulation. More recently, Yu (2016) found that speakers of Hong Kong Cantonese with less autistic-like traits exhibit more vocalic influence on /s/ production than those with more autistic-like traits. That study did not examine the participants' perceptual responses to coarticulatory information, however. The present study contributes to this line of inquiry by examining whether individual variation in perceptual responses to coarticulatory information might correlate with individual variation in coarticulation in production.

To examine the link between the perception and production of coarticulated speech, we report the results of two experiments examining (i) the identification of sibilants in different vocalic contexts and (ii) the acoustic realization of these sibilants in the corresponding contexts. Crucially, each participant took part in both experiments in order to allow for a direct examination of the correlation between the results of the two tasks. The investigation suggests that individual variability in the acoustic realization of /s/ and /ʃ/ in different vocalic contexts is indeed correlated with variability in how individuals respond to sibilants in different vocalic contexts perceptually.

## 2 Methods

### 2.1 Participants

Forty-two adult native speakers of American English (twenty-eight females), age ranged from 18 to 53 (median = 20,  $SD = 7$ ) with no reported history of speech, language, or hearing problems were recruited from the University of Chicago community and received a nominal fee or course credit for participating in the study.

### 2.2 Stimuli

For the perceptual task, the stimuli were those employed in Yu and Lee (2014) and can be found at <https://bit.ly/2CGcJRK>. Readers are referred to Yu and Lee (2014) for a detailed explanation on how the stimuli were created. Briefly, they created two 7-step  $/V_i s V_i - V_i f V_i/$  continua ( $V = /a/$  or  $/u/$ ) by mixing, using weighted average of waveforms,  $/s/$  and  $/f/$  taken from original  $/fa/$  and  $/fu/$  syllables produced by a male native speaker of American English. The natural  $/s/$  and  $/f/$  were included as endpoints of the 7-step series. The seven fricatives (synthesized and natural) were then cross-spliced with  $/a/$  and  $/u/$  taken from original  $/da/$  and  $/du/$  syllables produced by the same male speaker. The resulting tokens were then normalized for intensity and pitch.

The target stimuli for the production task were English words where the initial stressed syllable contained the onset consonant  $/s/$  or  $/f/$ , and one of the following vowels,  $/i/$ ,  $/u/$ , or  $/ae/$ . As noted above, we had chosen to employ the stimuli used in Yu and Lee (2014) to ensure maximal compatibility between earlier perceptual experiments and the current one. However, this methodological choice created several complications for the production task. To begin with, as the number of word-initial  $/s/$  and  $/f/$  minimal pairs in the  $/a/$  context is rather limited, we had to remedy this situation by using another low vowel,  $/ae/$ , as the environment for the production stimuli in order to preserve the contrast between low vs. high vowel contexts in the perceptual stimuli. In order to expand the empirical coverage of the production study, we also included the  $/i/$  context to examine the effect of lip rounding apart from vowel height. These methodological choices are admittedly not ideal since the vowel contexts examined in the perception and production tasks are not isomorphic. Future extensions of this experiment should aim for a more direct mapping between the perceptual and production stimuli. Each CV combination is instantiated by four distinct words. A total of 48 tokens (2 sibilants  $\times$  3 vowels  $\times$  4 words  $\times$  2 repetitions) were elicited from each participant. The list of target words is listed in **Table 1**.

### 2.3 Procedure

Participants performed a two-alternative forced choice (2AFC) identification task where the participants listened to a series of VCV sequences where C is one of the 7-step of the synthesized  $/s/ - /f/$  continuum and V is  $/a/$  or  $/u/$  and had to decide whether the fricative was  $/s/$  or  $/f/$ . Participants responded to six repetitions of each stimulus for a total of 84, randomly ordered, trials ( $= 7$  steps  $\times$  2 vowel contexts  $\times$  6 repetitions). The order of response options was counter-balanced across participants. Participants were given three

**Table 1:** Stimuli used in the production task.

Post-sibilant	$/s/$	$/f/$
$/i/$ :	CD, cedar, see, seeds	she, sheep, sheet, shield
$/ae/$ :	saddle, salad, sapphires, saxophone	champagne, shadow, shaft, shanks
$/u/$ :	soothsayer, soup, suitcase, Susan	shoe, shoelace, shoot, shoots

seconds to respond before the presentation of the next stimulus. Approximately 2% of the trials had no responses.

For the production task, each participant was digitally recorded in a quiet room individually at a sampling rate of 44,100 Hz with a Marantz PMD 670 solid-state recorder and a Shure SM10A head-mounted microphone, reading the target stimuli in a random order in a carrier phrase, “say \_\_ again.” The perception task always precedes the production task. Approximately 5% of the trials were lost due to mispronunciations or noise from participants touching the microphone.

Fricative segmentation involved the simultaneous consultation of the waveform and wideband spectrogram. Fricative onset was defined as the point at which high-frequency energy (roughly in the region above the second formant of the following vowel) first appeared on the spectrogram and/or the point at which the number of zero crossings rapidly increased. Frication offset was defined as the intensity minimum immediately preceding the onset of vowel periodicity.

The spectral properties of English sibilants have been extensively studied in the past (e.g., Blacklock, 2004; Iskarous, Shadle, & Proctor, 2011; Jongman, Wayland, & Wong, 2000; Whalen, 1981, 1991). We follow earlier reports, especially Shadle and Mair (1996) and Jongman et al. (2000), and analyzed the spectral properties of sibilant noise measured in terms of the spectral peak frequency, the first four spectral moments, and the total fricative duration. English /ʃ/ typically exhibits a midfrequency peak at around 2.5–3kHz, while /s/ displays a primarily spectral peak at around 4–5 kHz, although the location of the spectral peak is partly dependent on the speaker (Hughes & Halle, 1956) and vowel (Soli, 1981). Likewise, the first spectral moment (i.e., the spectral mean) also distinguishes well between /s/ and /ʃ/ in English (Jongman et al., 2000; Shadle & Mair, 1996) and across different vowel contexts (Nittrouer, 1995), gender (Nittrouer, 1995), and socio-economic classes (Stuart-Smith, 2007). Some report /s/ to be distinct from /ʃ/ in terms of having lower standard deviation (Jongman et al., 2000; Tomiak, 1990), although Li, Edwards, and Beckman (2009) found /s/ to have a more diffused shape (higher standard deviation) than /ʃ/ and /ç/. /ʃ/ is found to have a positive skewness, i.e., a concentration of energy in the lower frequencies. The sibilant /s/ also has a higher kurtosis (a more peaked distribution) than /ʃ/ in English (Jongman et al., 2000; Li et al., 2009). Shadle and Mair (1996) reported that the particularly high kurtosis value for /s/ around /u/ compared to other vowels is likely due to a whistly /s/ in rounded contexts.

A custom-made PRAAT script automatically extracted the spectral measurements. Similar to the measurement procedure described in Jongman et al. (2000), DFTs (frequency range: 500–12000 Hz) were calculated using a 40 ms Hamming window with preemphasis at 80 Hz, centered at eleven points (at 10% increments of the fricative’s duration from 0% to 100%) during the fricative. That is, each DFT is based on a window that spanned the preceding and following 20 ms of each measurement point. Measurements at 0%, 10%, 90%, and 100% were not included in the analysis. Spectral peak is defined in the script as the highest amplitude peak of the DFT. The same script also measured the duration of the sibilant.

The results of the experiments were analyzed in two different ways, at the group level and at the individual level. The perception-production link will be examined in the individual-level analysis section.

#### **2.4 Group-level perceptual results**

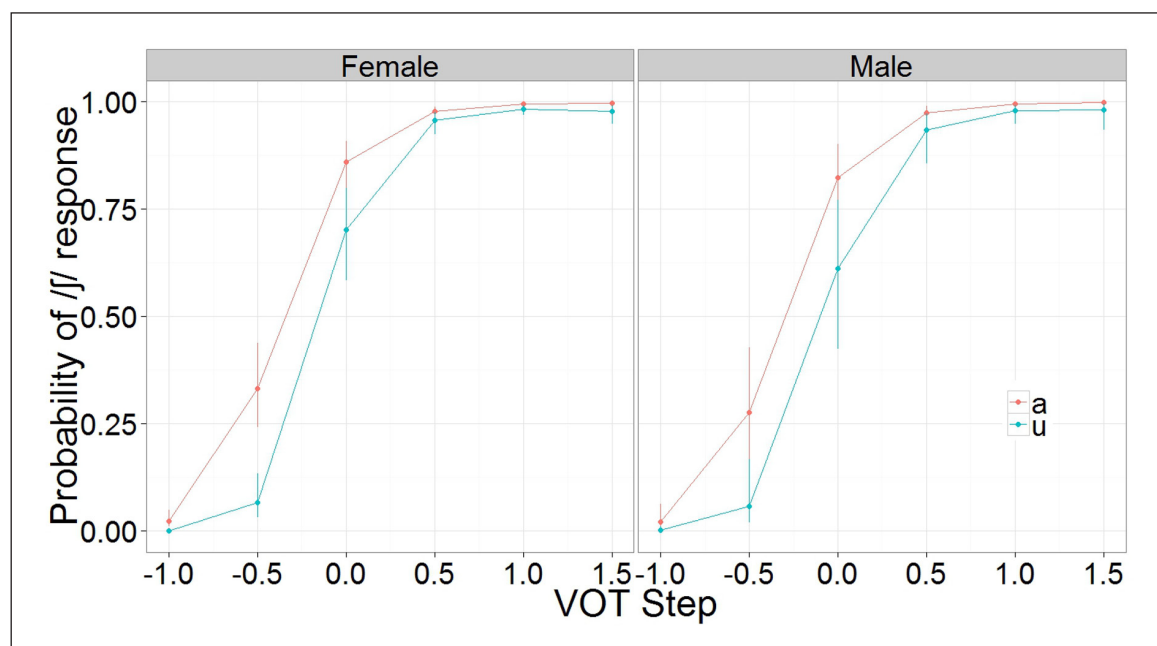
To examine the perceptual responses at the group level, participants’ perceptual responses (/ʃ/ response = 1, /s/ = 0) were modeled using logistic mixed-effects regression fitted in R, using the `lmer()` function from the `lme4` package (version 0.999999-2; Bates,

Maechler, & Bolker, 2011). The Wald's Z test, which describes how distant a coefficient estimate is from zero in terms of its standard error, was used to test the significance of estimates of the model.

A series of regression models were tested with three within-subject predictors (TRIAL indexed the order in which the stimuli were presented, VOWEL indexed the vocalic contexts [/u/ vs. /a/] and STEP indexed the seven steps along the /s-/ /j/ continuum) as well as a between-subject predictor of participant's biological SEX (female vs. male). Continuous variables (TRIAL and STEP) were centered and z-scored. SEX was sum-coded, while VOWEL was treatment-coded with /a/ as the baseline.

The final model included only four main predictors (TRIAL, VOWEL, STEP, and SEX), the two-way interactions between VOWEL and STEP, and their interactions with SEX. To account for the non-linearity of STEP, the quadratic term of STEP (i.e.,  $STEP^2$ ) was included in the model. By-subject random slopes were included for TRIAL, VOWEL, STEP, and  $STEP^2$ , as well as the interactions between VOWEL and STEP and between VOWEL and  $STEP^2$ , to allow for by-subject variability in the effect of each of the variables on /j/-identification. The model formula in lme4 style was: /j/-RESPONSE  $\sim$  TRIAL + VOWEL \* (STEP + I ( $STEP^2$ )) \* SEX + (1 + TRIAL + VOWEL \* (STEP + I ( $STEP^2$ )))|SUBJECT).

In light of findings from previous literature, listeners exhibiting perceptual compensation for coarticulation are expected to respond /j/ less often before /u/ than before /a/. This is the pattern observed. **Figure 1** illustrates the probability of a /j/-response in different vocalic contexts. The regression model shows a main effect of VOWEL ( $\beta = -1.018$ ,  $z = -4.604$ ,  $p < 0.001$ ), suggesting that the cohort, as a whole, exhibits more /j/-responses before /a/ than before /u/. There are also significant main effects of STEP ( $\beta = 4.516$ ,  $z = 13.374$ ,  $p < 0.001$ ) and  $STEP^2$  ( $\beta = -0.981$ ,  $z = -4.396$ ,  $p < 0.001$ ), suggesting that, in the /a/ context, the pattern of change in the log-odds for /j/ identification across the sibilant continuum concaves mildly downward. Crucially, VOWEL significantly interacts with STEP ( $\beta = 1.093$ ,  $z = 2.958$ ,  $p < 0.01$ ) and  $STEP^2$  ( $\beta = -1.318$ ,  $z = -3.385$ ,  $p < 0.001$ ), suggesting that the log-odds of /j/-identification across the continuum concaves downward more severely in the /u/ context than in the



**Figure 1:** Model predictions of perceptual responses at the group level across females (left) and males (right). Each panel shows predicted probability of /j/-response in /a/ and /u/ contexts.



/a/ context. As illustrated in **Figure 1a**, the probability of /ʃ/-identification in the /a/ context was low at the /s/-end of the sibilant continuum and rose gradually toward the middle of the continuum. In the /u/ context, this rise in the /ʃ/-identification rose more sharply in the middle of the sibilant continuum than toward the two ends. No main nor interaction effect involving SEX was significant.

## 2.5 Group-level Production Results

A summary of six spectral measures for /s/ and /ʃ/ in different vocalic contexts is presented in **Table 2**. As expected, /s/ shows qualitatively lower spectral mean/peak frequency, higher standard deviation, less peakiness, and less negatively skewed in the /u/ context than in the other unrounded vowel contexts. There is less, if any, vowel-dependent variation in the spectral properties of /ʃ/. Sibilants are also found to be shorter before the low vowel /ae/ than before the high vowels (cf. Yu, 2016).

While the spectral information measured is commonly analyzed separately for the purpose of elucidating the acoustic properties of fricatives, the mapping between individual spectral measures and their perceptual correlates is not clear, especially since some of the measures are highly correlated with each other (e.g., spectral mean, peak frequency, and skewness). In an effort to reduce the dimensionality of the mapping between spectral cues and perceptual responses and the complexity of the correlation analysis in the context of the examination of the perception-production link in the next section, rather than analyzing the spectral measures individually, an integrated cue-combination approach was taken such that the five spectral measures (centroid frequency, standard deviation, kurtosis, skewness, and peak frequency; see **Table 2** for a summary) were first submitted to a principal component analysis (PCA) to obtain linear combinations of these spectral variables that would capture the maximum variation. This analytic procedure has been successfully applied to the analysis of sibilants elsewhere (e.g., Yu, 2016).

**Table 2:** Descriptive statistics for peak frequency and the four moments measured at 50% of the sibilant noise interval, and duration of that interval in different vocalic contexts. Values presented are averages (and standard deviations in parentheses) across 42 participants and four lexical items per participant.

SIBILANT TYPE	/s/			/ʃ/		
	/ae/	/i/	/u/	/ae/	/i/	/u/
SPECTRAL MEAN	8355.85 (980.16)	8414.95 (948.03)	7992.27 (979.76)	4632.39 (553.54)	4674.06 (568.51)	4751.32 (549.05)
PEAK FREQUENCY	8319.59 (1264.20)	8366.70 (1246.80)	7853.40 (1528.81)	3715.18 (589.72)	3753.31 (681.57)	3715.36 (756.50)
STANDARD DEVIATION	1280.81 (226.01)	1275.54 (231.70)	1476.20 (303.43)	1542.65 (248.94)	1545.69 (256.69)	1626.33 (246.94)
KURTOSIS	1.77 (1.56)	1.83 (1.54)	0.89 (1.48)	2.07 (2.17)	1.81 (2.09)	1.19 (1.99)
SKEWNESS	-0.36 (0.60)	-0.37 (0.61)	-0.25 (0.63)	1.32 (0.56)	1.24 (0.55)	1.06 (0.56)
DURATION	142.23 (26.49)	178.34 (39.23)	175.29 (35.72)	150.86 (33.12)	191.66 (33.03)	189.83 (33.84)

The specifics of the PCA are as follows. Spectral peak frequency, spectral mean, and spectral standard deviation, which are all in Hertz, and kurtosis, which is unitless but negatively skewed, were log-transformed (natural log). Skewness was not transformed in any way since it is already unitless and normally distributed. Since the kurtosis values can be negative, all kurtosis values were increased by 3 to ensure that they were positive prior to log-transformation. The spectral measures (measurements at the first and last two measurement points were not included in the analysis to avoid measurement problems at the edges of the sibilant), not including sibilant duration, were analyzed using the `prcomp()` function in R, which performs a principal components analysis on the given data matrix. All acoustic parameters were centered and z-scored for the PCA.

The relative weightings and proportion of variance for each component are summarized in **Table 3**. The optimal linear combination (PC1), which accounts for about 59% of the variance, and the 2nd component (PC2), which accounts for approximately 32% of the variance, were selected as independent variables for the analysis below as the first two components collectively account for more than 90% of the variance. PC1 has strong loadings for skewness and log-transformed peak frequency and spectral mean, which are all spectral measures that characterize the concentrations of spectral energy. Higher spectral mean and peak values and left-skewness generally correspond to a more /s/-like percept (Jongman et al., 2000). PC2, on the other hand, is dominated by standard deviation and kurtosis, which pertain to the energy levels across different frequencies of the spectrum. Lower standard deviation and higher kurtosis generally correspond to a more /s/-like percept (Jongman et al., 2000).

The results of the two linear-transformed components, obtained using the `predict()` function in R, were modeled separately using linear mixed-effects regression fitted in R, using the `lmer()` function from the `lme4` package (version 0.999999-2; Bates et al., 2011). Both components were tested for the following task-level effects: TRIAL order (1–45), SYLLABLE count (1–3), sibilant DURATION, sibilant TYPE (/s/ vs. /ʃ/), VOWEL (/i/, /ae/, /u/), measurement POSITION (1–7), the participant’s biological SEX (female vs. male), and the two- and three-way interactions between TYPE, VOWEL, POSITION, and SEX. In order to take into account the temporal dynamics of the vocalic influence, the quadratic term for POSITION (i.e., POSITION<sup>2</sup>) was included in the analysis. The number of syllables of each stimulus, which might affect the realization of the target sibilant, was taken into account since syllable count was not controlled for in stimulus selection in an effort to maintain the prosodic position of the target sibilants. Continuous

**Table 3:** The cumulative proportion of variance accounted for and loadings from the PCA of sibilants from spectral measures.

	PC1	PC2	PC3	PC4	PC5
Standard deviation	1.709	1.261	0.567	0.363	0.186
Proportion of Variation	0.584	0.318	0.064	0.026	0.007
Cumulative Proportion	0.584	0.902	0.967	0.993	1.000
log (Peak Frequency)	0.569	-0.001	0.054	0.557	-0.602
log (Spectral mean)	0.572	-0.095	0.021	0.251	0.775
Skewness	-0.538	0.192	-0.235	0.766	0.179
log (Standard deviation)	-0.244	-0.647	0.694	0.199	0.018
log (Kurtosis+3)	0.000	0.732	0.678	0.009	0.068

variables, including POSITION, were centered and z-scored. The VOWEL variable was treatment-coded such that the /ae/ context was taken to be the baseline. Thus the first contrast, VOWEL<sub>p</sub>, compares the influence of /ae/ to the influence of the high vowels /i/, while the second contrast, VOWEL<sub>u</sub>, encodes the rounding contrast between /ae/ and /u/. The models also included by-subject random intercepts to allow for subject-specific variation in the specific spectral component as well as by-subject random slopes for each of the main predictors. The model formula in `lme4` style for the first two principal components of the spectral measures (PC) was  $PC \sim \text{TRIAL} + \text{DURATION} + \text{SYLLABLE} + (\text{SIBILANT TYPE} + \text{VOWEL} + \text{POSITION} + \text{SEX})^3 + (\text{SIBILANT TYPE} + \text{VOWEL} + \text{I}(\text{POSITION}^2) + \text{SEX})^3 + (1 + \text{TRIAL} + \text{DURATION} + \text{SYLLABLE} + \text{SIBILANT TYPE} + \text{VOWEL} + \text{POSITION} + \text{I}(\text{POSITION}^2))|\text{SUBJECT}$ .

The residuals of the initial fit of each model were examined and were found to deviate strongly from normality. As a result, residuals which were more than 2.5 standard deviations from the mean were trimmed, which amounted to no more than 2.6% of the data for each principal component modeled, and the models were refitted to the trimmed data set. The new models had residual distributions much closer to normality, and it is the refitted models that are reported below. The estimates for all predictors in the analysis of the first two principal components of the spectral measures can be found in **Table 4**.

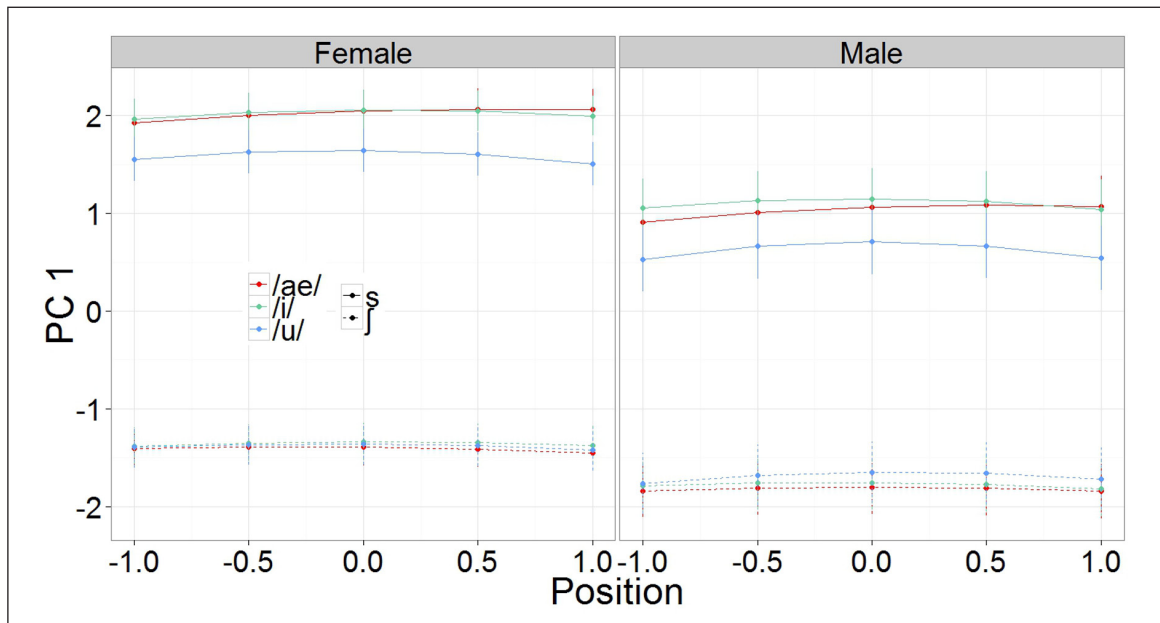
Recall that PC1 has the strongest loadings for peak frequency, spectral mean, and skewness; an increase in PC1 corresponds to an increase in peak frequency and spectral mean and a decrease in skewness (thus less tilt toward the left). Previous reports suggest that, relative to the spectrum of /ʃ/, the spectrum of /s/ has higher spectral mean and spectral peak, and is more negatively skewed. Taken together, we expect a higher PC1 values for /s/ than for /ʃ/. Given that lip rounding and protrusion results can lead to a more /ʃ/-like /s/ acoustically, PC1 values are expected to be lower before /u/ than before /ae/ and /i/. These predictions are indeed borne out. **Figure 2** illustrates the distribution of PC1 for /s/ and /ʃ/ in different vocalic contexts. The PC1 values for /s/ and /ʃ/ differ significantly, as indicated by a main effect of SIBILANT TYPE ( $\beta = 1.58$ ,  $t = 27.17$ ,  $p < 0.001$ ). The difference in PC1 between /s/ and /ʃ/ varies across measurement points both linearly (TYPE  $\times$  POSITION:  $\beta = 0.04$ ,  $t = 8.16$ ,  $p < 0.001$ ) and non-linearly (TYPE  $\times$  POSITION<sup>2</sup>:  $\beta = -0.01$ ,  $t = -2.08$ ,  $p < 0.05$ ). As illustrated in **Figure 2**, this difference is primarily driven by the fact that the trajectory of PC1 changes across measurement points for /s/ has a more downwardly concave shape than that of /ʃ/.

There is a main effect of VOWEL<sub>u</sub> ( $\beta = -0.14$ ,  $t = -3.88$ ,  $p < 0.001$ ), showing that /ae/ and /u/ exert significantly different effects on PC1 such that PC1 is lower before /u/ than before /ae/. VOWEL<sub>u</sub> interacts significantly with TYPE ( $\beta = -0.24$ ,  $t = -20.01$ ,  $p < 0.001$ ); the difference in effects between /ae/ and /u/ is larger in /s/ than in /ʃ/. The vocalic effect also varies temporally. The difference between the effects of /ae/ and /u/ on PC1 varies across measurement points both linearly (VOWEL<sub>u</sub>  $\times$  POSITION:  $\beta = -0.03$ ,  $t = -4.06$ ,  $p < 0.001$ ) and nonlinearly (VOWEL<sub>u</sub>  $\times$  POSITION<sup>2</sup>:  $\beta = -0.06$ ,  $t = -6.15$ ,  $p < 0.001$ ). Moreover, the sibilant-specific effects of vowel also differ across measurement positions. **Figure 2** shows that the difference in vocalic influence between the onset and offset of /s/ is larger than the difference between those of /ʃ/. The significant three-way interactions between TYPE, VOWEL<sub>u</sub>, and POSITION ( $\beta = -0.05$ ,  $t = -6.72$ ,  $p < 0.001$ ) and between TYPE, VOWEL<sub>p</sub>, and POSITION<sup>2</sup> ( $\beta = -0.03$ ,  $t = -3.13$ ,  $p < 0.01$ ) suggest that there are sibilant-specific differences in terms of the effects of /ae/ and /u/ on PC1 across measurement points; the temporally-dependent difference in vocalic influence on PC1 is more pronounced on /s/ than on /ʃ/.

**Table 4:** Estimates for all predictors in the analysis of the first two principal components of the spectral measures. The variable, VOWEL, was treatment-coded with the /ae/ context as the baseline.

	PC1		PC2	
	Coef (SE)	t-value	Coef (SE)	t-value
INTERCEPT	-0.023 (0.070)	-0.324	0.317 (0.096)	3.313***
TRIAL	0.018 (0.009)	1.974*	0.003 (0.018)	0.160
SYLLABLE COUNT	0.006 (0.008)	0.802	-0.012 (0.019)	-0.631
DURATION	0.012 (0.013)	0.909	0.123 (0.033)	3.678***
SEX	0.349 (0.070)	5.009***	0.053 (0.092)	0.581
POSITION	0.031 (0.011)	2.790**	0.045 (0.020)	2.231*
POSITION <sup>2</sup>	-0.049 (0.009)	-5.758***	-0.053 (0.020)	-2.652**
POSITION × SEX	-0.008 (0.011)	-0.727	-0.015 (0.019)	-0.772
POSITION <sup>2</sup> × SEX	0.005 (0.009)	0.629	0.018 (0.018)	1.010
SIBILANT TYPE	1.577 (0.058)	27.166***	-0.095 (0.099)	-0.960
TYPE × POSITION	0.042 (0.005)	8.157***	-0.057 (0.011)	-5.293***
TYPE × POSITION <sup>2</sup>	-0.012 (0.006)	-2.083*	-0.052 (0.013)	-4.143***
TYPE × SEX	0.143 (0.057)	2.526*	0.304 (0.094)	3.246**
TYPE × POSITION × SEX	0.001 (0.003)	0.377	0.021 (0.007)	3.030**
TYPE × POSITION <sup>2</sup> × SEX	0.004 (0.004)	1.079	-0.014 (0.008)	-1.734
VOWEL <sub>i</sub>	0.050 (0.026)	1.937	-0.103 (0.069)	-1.507
VOWEL <sub>i</sub> × POSITION	-0.032 (0.008)	-4.117***	-0.006 (0.016)	-0.384
VOWEL <sub>i</sub> × POSITION <sup>2</sup>	-0.015 (0.009)	-1.685	-0.047 (0.019)	-2.467*
VOWEL <sub>i</sub> × TYPE	-0.001 (0.011)	-0.105	0.030 (0.024)	1.235
VOWEL <sub>i</sub> × TYPE × POSITION	-0.037 (0.007)	-5.271***	-0.025 (0.015)	-1.656
VOWEL <sub>i</sub> × TYPE × POSITION <sup>2</sup>	-0.010 (0.008)	-1.253	-0.012 (0.017)	-0.703
VOWEL <sub>i</sub> × SEX	-0.018 (0.024)	-0.761	-0.077 (0.055)	-1.393
VOWEL <sub>i</sub> × TYPE × SEX	-0.021 (0.008)	-2.586**	0.045 (0.017)	2.704**
VOWEL <sub>i</sub> × POSITION × SEX	0.018 (0.008)	2.348*	0.009 (0.016)	0.533
VOWEL <sub>i</sub> × POSITION <sup>2</sup> × SEX	0.004 (0.009)	0.393	0.048 (0.019)	2.536*
VOWEL <sub>u</sub>	-0.142 (0.037)	-3.881***	-0.756 (0.098)	-7.739***
VOWEL <sub>u</sub> × POSITION	-0.033 (0.008)	-4.063***	-0.128 (0.017)	-7.560***
VOWEL <sub>u</sub> × POSITION <sup>2</sup>	-0.057 (0.009)	-6.146***	-0.020 (0.020)	-1.022
VOWEL <sub>u</sub> × TYPE	-0.236 (0.012)	-20.014***	-0.128 (0.025)	-5.174***
VOWEL <sub>u</sub> × TYPE × POSITION	-0.049 (0.007)	-6.721***	-0.030 (0.015)	-1.955
VOWEL <sub>u</sub> × TYPE × POSITION <sup>2</sup>	-0.026 (0.008)	-3.134**	-0.011 (0.018)	-0.647
VOWEL <sub>u</sub> × SEX	-0.042 (0.033)	-1.271	0.065 (0.081)	0.798
VOWEL <sub>u</sub> × TYPE × SEX	0.018 (0.008)	2.181*	0.056 (0.017)	3.246**
VOWEL <sub>u</sub> × POSITION × SEX	-0.010 (0.008)	-1.178	0.019 (0.017)	1.128
VOWEL <sub>u</sub> × POSITION <sup>2</sup> × SEX	0.021 (0.009)	2.279*	-0.004 (0.019)	-0.209

Note: \*\*\* =  $p < 0.001$ ; \*\* =  $p < 0.01$ ; \* =  $p < 0.05$ .  $p$ -values were obtained using normal approximation which has the assumption that the  $t$  distribution converges to the  $z$  distribution as degrees of freedom increase (see Mirman, 2014, for details).



**Figure 2:** Model predictions of PC1 values for /s/ and /ʃ/ in the contexts of /i/, /ae/, and /u/. Error bars present the 95% confidence intervals.

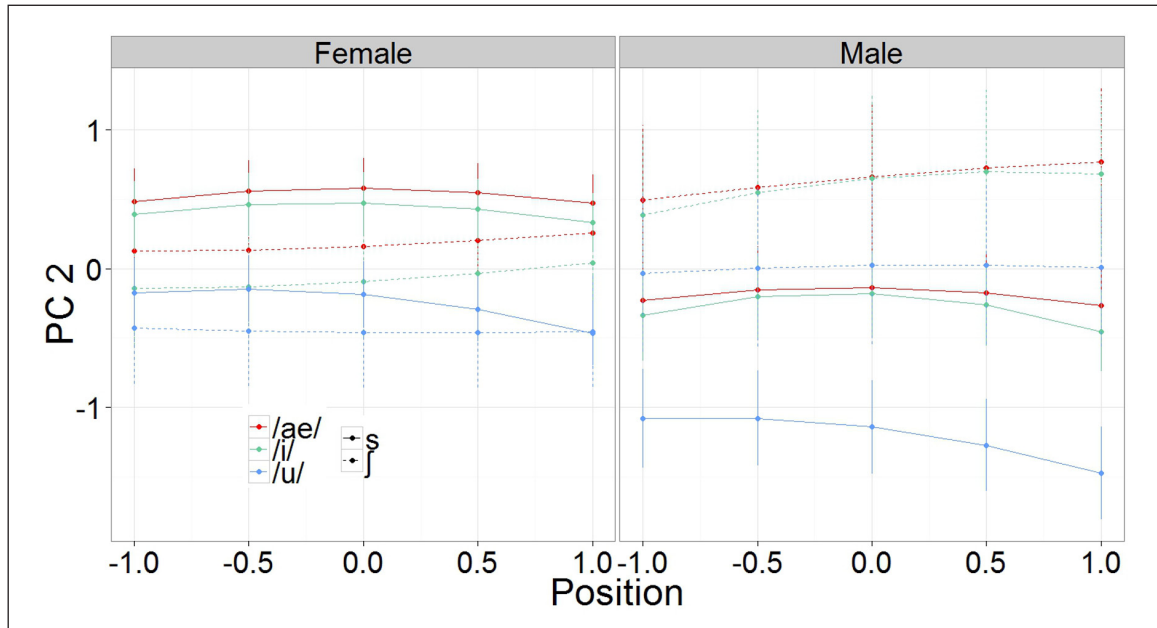
In terms of sex-specific effects, there is a main effect of SEX on PC1 ( $\beta = 0.35$ ,  $t = 5.01$ ,  $p < 0.001$ ); in general, females have higher PC1 than males. SEX also interacts significantly with other factors. In particular, the interaction between SEX and TYPE ( $\beta = 0.14$ ,  $t = 2.53$ ,  $p < 0.05$ ) indicates that there is a larger difference between /s/ and /ʃ/ in females than in males. The sibilant-specific effect of vocalic rounding on PC1 is smaller in females than in males (VOWEL<sub>u</sub> × TYPE × SEX:  $\beta = 0.02$ ,  $t = 2.18$ ,  $p < 0.05$ ). The significant three-way interaction between VOWEL<sub>u</sub>, POSITION<sup>2</sup>, and SEX indicates that the rounding effect on PC1 across measurement points has a more downward concave trajectory in males than in females.

While there is not a significant main effect of VOWEL<sub>i</sub>, suggesting that the influences of /ae/ and /i/ on PC1 do not differ markedly, there is a significant two-way interaction between VOWEL<sub>i</sub> and POSITION, indicating that the rise in PC1 across measurement positions is steeper before /ae/ than before /i/. This difference is driven by /s/, however, as indicated by the significant three-way interaction with TYPE ( $\beta = -0.04$ ,  $t = -5.27$ ,  $p < 0.001$ ). Furthermore, this vocalic difference in PC1 steepness across measurement positions is stronger in males (VOWEL<sub>i</sub> × POSITION × SEX:  $\beta = 0.02$ ,  $t = 2.35$ ,  $p < 0.05$ ). Finally, while the influences of /ae/ and /i/ do not differ across sibilant type, PC1 is higher before /i/ than /ae/ for male /s/, but not for female ones ( $\beta = -0.02$ ,  $t = -2.59$ ,  $p < 0.01$ ).

With respect to the second principal component (PC2), recall that PC2 has strong loadings for standard deviation and kurtosis; an increase in PC2 corresponds to a decrease in standard deviation and an increase in kurtosis (i.e., more peaky distribution). Given that /s/ has been shown to have lower standard deviation and greater kurtosis than /ʃ/, /s/ is expected to have higher PC2 than /ʃ/. Likewise, as /s/ is more /ʃ/-like before /u/, /s/ is expected to have lower PC2 before /u/ than before the other vowels. These predictions are only partially borne out.

**Figure 3** illustrates the distribution of PC2 for /s/ and /ʃ/ in different vocalic contexts. While there is not a significant main effect of TYPE, TYPE interacts significantly with POSITION ( $\beta = -0.06$ ,  $t = -5.29$ ,  $p < 0.001$ ) and POSITION<sup>2</sup> ( $\beta = -0.05$ ,  $t = -4.14$ ,  $p < 0.001$ ), suggesting that PC2 for /s/ and /ʃ/ have different slopes across measurement





**Figure 3:** Model predictions of PC2 values for /s/ and /ʃ/ in the contexts of /i/, /ae/, and /u/. Error bars present the 95% confidence intervals.

points and PC2 has a more downward concave trajectory for /s/ than for /ʃ/. TYPE also interacts with SEX significantly ( $\beta = 0.3$ ,  $t = 3.25$ ,  $p < 0.01$ ), suggesting that there is a large PC2 difference between /s/ and /ʃ/, but this difference is mostly observed in males, not females. In particular, male speakers have higher PC2 for /ʃ/ than /s/, while female speakers show the opposite trend. TYPE  $\times$  SEX also significantly interacts with the linear term of POSITION ( $\beta = 0.02$ ,  $t = 3.03$ ,  $p < 0.01$ ), indicating that, while PC2 for /s/ and /ʃ/ diverge linearly across measurement positions for the male speakers, they converge for the females. There is also a significant effect of DURATION ( $\beta = 0.12$ ,  $t = 3.68$ ,  $p < 0.001$ ); an increase in sibilant duration leads to an increase in PC2 (i.e., a decrease in standard deviation and an increase in kurtosis).

In terms of vocalic effects, it is as predicted that there is a main effect of VOWEL<sub>u</sub> ( $\beta = -0.76$ ,  $t = -7.74$ ,  $p < 0.001$ ) such that PC2 is lower before /u/ than before /ae/. VOWEL<sub>u</sub> interacts with POSITION ( $\beta = -0.13$ ,  $t = -7.56$ ,  $p < 0.001$ ), indicating that the PC2 trajectories diverge before /ae/ and /u/. In particular, PC2 has a downward trend before /u/ while an upward trend before /ae/. VOWEL<sub>u</sub> also interacts with TYPE ( $\beta = -0.13$ ,  $t = -5.17$ ,  $p < 0.001$ ), showing that the PC2 for /s/ is lower before /u/ than for /ʃ/. Moreover, there is also a three-way interaction between VOWEL, TYPE, and SEX ( $\beta = 0.06$ ,  $t = 3.25$ ,  $p < 0.01$ ), indicating that the vocalic difference in PC2 between /s/ and /ʃ/ is smaller for females than males.

No main effect of VOWEL<sub>i</sub> is observed. There is a significant three-way interaction between VOWEL<sub>i</sub>, TYPE, and SEX ( $\beta = 0.05$ ,  $t = 2.7$ ,  $p < 0.01$ ), suggesting that there is a larger /ae/-/i/ difference between /s/ and /ʃ/ in females than in males. A significant three-way interaction between VOWEL<sub>i</sub>, POSITION<sup>2</sup>, and SEX ( $\beta = 0.05$ ,  $t = 2.54$ ,  $p < 0.05$ ) indicates that the PC2 before /i/, relative to PC2 before /ae/, has a more downward concave trajectory in males than in females.

Overall, the production results suggest that, while males show less distinctness between /s/ and /ʃ/ in PC1 (i.e., in terms of spectral mean, peak frequency, and skewness) relative to females, they exhibit more distinctness in PC2 (in terms of standard deviation and kurtosis) for the same contrast. In terms of vocalic influence, males appear to exhibit more vocalic influences than females.

Thus far, we have reviewed the significant effects of vocalic coarticulation on the acoustic realization of /s/ and /ʃ/ in English. The vocalic effects are generally stronger on /s/ than on /ʃ/ and they are stronger toward the sibilant offset than the onset. The perceptual results also indicate that the participants, as a group, exhibit significant effects of perceptual compensation. That is, consistent with previous literature, participants recorded fewer /ʃ/ responses in the /u/ context than in the /a/ context. However, the fact that the inclusion of by-subject random slopes for contextual factors (i.e., the main factors of VOWEL in both the perceptual and production models and the interaction of VOWEL with STEP in the perceptual model) significantly improves model likelihood suggest there exist a high degree of individual variability in the perception and production of sibilants in different vocalic contexts. The next section explores the nature of this variation in detail.

## 2.6 The perception-production link: A closer look at individual variation

This section considers the nature of individual variability in the perception and production of coarticulated speech by examining the link between them. In this study, we explore the connection between the perception and production of coarticulated speech by fitting individual regression models for each subject's perceptual and production responses and examining the correlation between the by-subject estimates (i.e., the coefficients) for the coarticulatory context-related predictors in the perceptual and production regression models. The regression models were fitted using the `ddply()` function in the `plyr` package (Wickham, 2012). The perceptual responses were modeled in terms of Firth's Bias-Reduced logistic regressions using the `logistf()` function in the `logistf` package (Heinze, Ploner, Dunkler, & Southworth, 2016) to avoid problems of separation. The production measures (PC1 and PC2) were modeled using linear regressions. The models included the fixed factors already mentioned above. Model formulas are /ʃ/-RESPONSE  $\sim$  TRIAL + VOWEL \* STEP and PC  $\sim$  TRIAL + DURATION + SIBILANT TYPE \* VOWEL \* (POSITION + POSITION<sup>2</sup>). **Table 5** summarizes the predictors whose model estimates were used in the correlation study. With 68 possible correlations (4 perception estimates  $\times$  17 production estimates), the alpha level with Bonferroni correction is 0.0007.

**Theoretical predictions:** Before diving into the results of the correlation analysis, it is worth noting that, despite the admittedly exploratory nature of the correlation analysis, it is important to consider what correlations might be expected *a priori*. As reviewed in

**Table 5:** Summary of predictors whose model estimates were used in the correlation study. The variable VOWEL has two contrasts, VOWEL<sub>i</sub> and VOWEL<sub>u</sub>.

Perception	Production
VOWEL	VOWEL
STEP	POSITION
VOWEL $\times$ STEP	POSITION <sup>2</sup>
VOWEL $\times$ STEP <sup>2</sup>	TYPE
	VOWEL $\times$ POSITION
	VOWEL $\times$ POSITION <sup>2</sup>
	TYPE $\times$ POSITION
	TYPE $\times$ POSITION <sup>2</sup>
	TYPE $\times$ VOWEL
	TYPE $\times$ VOWEL $\times$ POSITION
	TYPE $\times$ VOWEL $\times$ POSITION <sup>2</sup>

the Introduction, the existing literature reported conflicting claims about the perception-production link. Some reported no relationship, while others found indirect or inconsistent evidence. The inconsistency in previous research findings might stem from the fact that the mapping between perception and production is not direct. As shown above, many factors (and the interactions between them) are responsible for explaining the variances in the perception and production results.

From a purely theoretical perspective, perception models that do not assume non-veridical encoding of percepts predict a positive relation between perception and production. In its most basic form, one might expect the magnitude of perception compensation to find the analog in the degree of coarticulation reflected in production. That is, for example, the coefficients for VOWEL from both the perception and production models would be expected to correlate positively. More nuanced models, such as Pierrehumbert (2002) and Sonderegger and Yu (2010), which see perceptual responses as reflective of the frequency distribution of the acoustic cues that are indexed to different contextual information, including coarticulated contexts, would predict that the context-dependent realization of /s/ and /ʃ/ (as indexed by the TYPE × VOWEL interaction in the production model) would positively correlate with the context-dependent classification of the /s/-/ʃ/ continuum (as indexed by the STEP × VOWEL interaction in the perceptual model). That is, the greater the degree of context-dependent variation is observed in production, the more context-sensitivity is expected in the listeners' classification of the sibilant continua. For example, given the distinctness between /s/ and /ʃ/ is reduced in the /u/ context, one might expect listeners to exhibit less certainty (e.g., a shallower classification function) in classifying the sibilant continuum in the /u/ context. Given that the vocalic influence, particularly the effect of /u/, is stronger toward the sibilant offset than at the onset, to the extent that listeners are sensitive to the temporal dynamics of vocalic influence on the spectral quality of the sibilant, a correlation is expected between the coefficients for the VOWEL × POSITION<sup>(2)</sup> interactions from the production models and the coefficients for the VOWEL or VOWEL × STEP<sup>(2)</sup> term of the perception models. Models that assume non-veridical encodings, such as C-Cure and perceptual models implicitly assumed in listener-misperception models of sound change (see more discussion below), would predict the opposite correlations. That is, individuals with strong coarticulation in production should exhibit the least compensatory response or less context-sensitivity.

**Results of the correlation analysis:** Table 6 summarizes the correlation results. While several correlations show *p* values below 0.05, only one correlation is significant at the alpha-adjusted level, namely, the correlation between the estimates for the TYPE × VOWEL<sub>u</sub> interaction in the production models and the estimates for the STEP × VOWEL<sub>u</sub> in the perceptual models ( $r = 0.51$ ,  $p = 0.0005$ ; Bonferroni-corrected alpha = 0.0007). Figure 4 shows a scatterplot illustrating the negative relationship between the STEP × VOWEL estimates (the x-axis) on the one hand, and the estimates for the TYPE × VOWEL<sub>u</sub> interaction (the y-axis) on the other.

As the variables being correlated are estimates for interactions between predictors, it would not be feasible to interpret the nature of the correlations without first examining the nature of individual variability concerning a given interaction between predictors within each regression model. To this end, we first focus on the STEP × VOWEL estimates. The left column of Figure 5 shows the mean percentage of /ʃ/ response in /a/ and /u/ contexts by individuals in the 1st (top panel) and 4th (bottom panel) quartiles of the STEP × VOWEL estimates. Individuals in the 4th quartile of the STEP × VOWEL estimates (i.e., data points toward the right end of the x-axis in Figure 4) show the expected patterns of perceptual compensation for coarticulation, where the identification function for /ʃ/ responses in the /u/ context appears to the right of the corresponding identification function in the

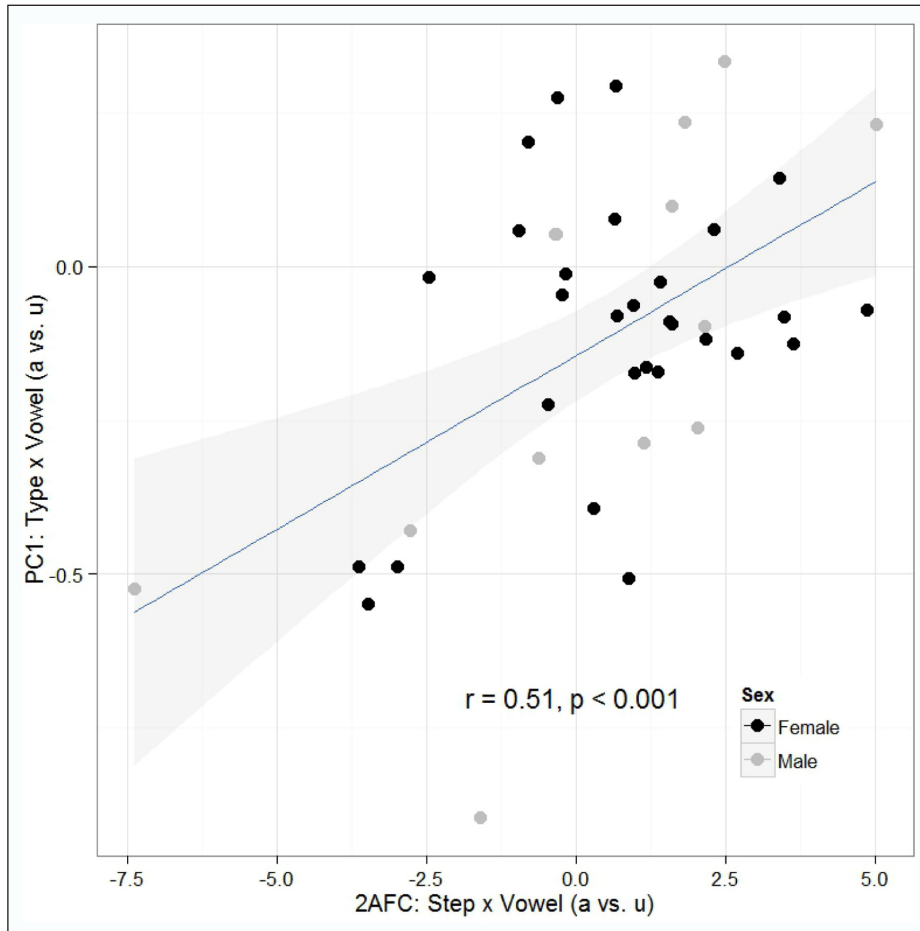
**Table 6:** Correlations between coefficients from the perception (in columns) and production (in rows) regression models.

	V(OWEL)	STEP	STEP <sup>2</sup>	V × STEP	V × STEP <sup>2</sup>
VOWEL <sub>i</sub>	-0.08	0.06	-0.23	-0.22	0.11
VOWEL <sub>u</sub>	-0.09	-0.27	-0.15	0.22	0.38*
POSITION	0.19	0.02	-0.04	0.17	-0.21
POSITION <sup>2</sup>	0.15	-0.23	-0.36*	0.12	0.27
TYPE	0.23	0.02	0.19	-0.06	-0.15
TYPE × VOWEL <sub>i</sub>	0	0.05	-0.06	0.02	-0.15
TYPE × VOWEL <sub>u</sub>	-0.02	-0.49**	-0.06	<b>0.51***</b>	0.2
TYPE × POSITION	-0.13	-0.14	-0.11	0.25	-0.07
TYPE × POSITION <sup>2</sup>	0.01	-0.01	-0.08	-0.17	0.16
VOWEL <sub>i</sub> × POSITION	0.03	0.08	0.13	0.02	0.02
VOWEL <sub>u</sub> × POSITION	-0.13	0.15	-0.09	-0.28	0.08
VOWEL <sub>i</sub> × POSITION <sup>2</sup>	0	-0.04	0.26	0.07	-0.17
VOWEL <sub>u</sub> × POSITION <sup>2</sup>	0.06	-0.05	0.19	-0.03	-0.2
TYPE × VOWEL <sub>i</sub> × POSITION	-0.06	0	-0.13	-0.06	0.12
TYPE × VOWEL <sub>u</sub> × POSITION	-0.08	0.05	-0.13	-0.2	-0.01
TYPE × VOWEL <sub>i</sub> × POSITION <sup>2</sup>	-0.02	0.01	0.04	-0.01	-0.05
TYPE × VOWEL <sub>u</sub> × POSITION <sup>2</sup>	0.08	0.02	0.03	0	-0.12

Note: \* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ . The correlation that reached the alpha level with Bonferroni correction of  $p < 0.0007$  is highlighted.

/a/ context. The pattern differs for the individuals in the 1st quartile (top panel). While individuals in the 1st quartiles (data points toward the left end of **Figure 4**) exhibited some degree of perceptual compensation, as evidenced by the differences in identification functions across vowel contexts at the crossover point (i.e., when identification rate is at 50%), we see that the /f/-identification function in the /u/ context shows a shallower rise than the corresponding identification function in the /a/ context, suggesting that the identification of /s/ and /f/ is more gradient in the /u/ context than in the /a/ context. Listeners in this quartile are also less certain in their identification of sibilant in the /u/ context in general. That is, listeners are less likely to identify a sibilant as /f/ at the /f/-end of the sibilant continuum and they are more likely to identify a sibilant as /f/ even at the /s/-end of the continuum. Crucially, this uncertainty is most acute in the /u/ context, not generally, suggesting that this is unlikely to be a general difference in classification strategies across individuals (cf. Kong & Edwards, 2016).

The production patterns mirror the perceptual patterns. The TYPE × VOWEL<sub>u</sub> interaction captures the way the vocalic environment influences the production of the contrast between /s/ and /f/. The larger the STEP × VOWEL estimate in perception (i.e., more uncertainty between /s/ and /f/ in the /u/ context), the less distinct the contrast in production is between /s/ and /f/ in the /u/ context relative to the /ae/ context. As shown in the right column of **Figure 5**, which shows the corresponding mean PC1 values for /s/ and /f/ in the contexts of /i/, /ae/, and /u/, individuals in the 1st quartile of the STEP × VOWEL estimates (top panel) show a weaker distinction (i.e., a smaller PC1 difference) between /s/ and /f/ in the /u/ context, compared to the individuals in the 4th quartile (bottom panel).



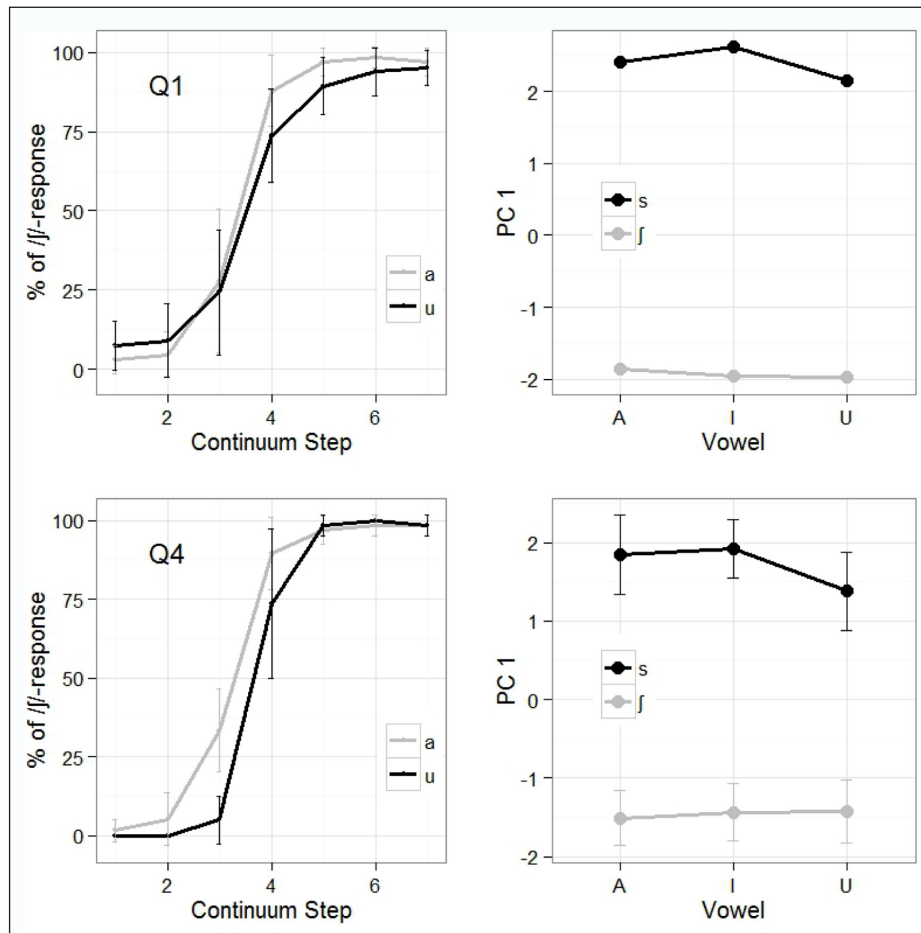
**Figure 4:** The correlation between estimates for the STEP  $\times$  VOWEL interaction in the /f/-response models on the x-axis and estimates for the TYPE  $\times$  VOWEL<sub>u</sub> interaction in the PC1 models on the y-axis. The sex of participant is indicated by the color of the symbol (black for females and gray for males). The line and shading show a linear fit and 95% confidence intervals (CIs).

## 2.7 Discussion

In the preceding section, we observed significant correlations between individual-level estimates for predictors in the perceptual and production regression models. More specifically, our results point to a relationship between the distinctness in the acoustic realization of /s/ and /ʃ/ across vocalic contexts and the categorical perception of /s/ and /ʃ/ in the corresponding vocalic environments. In particular, the less distinct the acoustic differences are between /s/ and /ʃ/ in a particular vowel context (in this case, primarily in the /u/ context), the less categorical the perceptual responses are in the corresponding vocalic context. Simply put, the participants are less certain about the identity of the sibilant in the /u/ context if their own productions of /s/ and /ʃ/ are not so distinct in that vocalic context. This uncertainty in the /u/ context extends even to the endpoints of the continuum where the sibilants should be most distinct. These findings are most consistent with perceptual models that assume veridical encoding, such as the auditorist, the gesturalist, and the exemplar-based approaches to speech perception since there is a positive relationship between the perception and production of coarticulation.

Our results resonate most strongly with the perception-production loop models such as Pierrehumbert (2002) and Sonderegger and Yu (2010). Specifically, the context-specific distinctness in the acoustic realization of the sibilants is reflected in shifts in context-specific category boundaries as well as the categoricity of the perceptual responses. This finding echoes previous findings that show a correlation between how distinct an





**Figure 5:** The mean percentage of /j/ response in /a/ and /u/ contexts (left column), and the corresponding mean PC1 for /s/ and /j/ in the contexts of /i/, /ae/, and /u/ by individuals in the 1st and 4th quartiles of the STEP  $\times$  VOWEL estimates.

individual produces a contrast and how well that individual discriminates the contrast (Newman et al., 2001; Perkell, Guenther, et al., 2004; Perkell, Matthies, et al., 2004). Our findings extend these earlier findings, demonstrating the context-specificity of such a contrast-based correlation.

Our findings are not consistent with predictions of perceptual models that assume non-veridical encoding, such as C-Cure (Cole et al., 2010, cf. Yu, 2016; Yu & Lee, 2014) and models of sound change that rely on listener misperception as a driving force behind certain sound changes. Such models predict a negative relationship between the perception of coarticulated speech. That is, assuming that all listeners engage in expectation adjustment somehow, albeit to varying degrees, listeners engaging in weak expectation adjustment (i.e., the so-called ‘misperceivers’) would register a heavily coarticulated /su/ as /ju/ while listeners who engage in more robust compensation would recode the same co-articulated /su/ as a non-coarticulated /su/. Since the perceptual exemplar space for the so-called ‘misperceivers’ would have more /ju/-like exemplars than individuals who engage in more robust expectation adjustment, the production targets of the ‘misperceivers’ should be more [ju]-like than those of the individuals who engage in more robust expectation adjustment. Thus the more robust one engages in expectation adjustment, the less vocalic influence is expected in one’s production.

A noteworthy aspect of our findings is the fact that the correlation between perception and production of coarticulation is much more nuanced than previous studies have generally assumed. That is, the correlations between the general effects of vocalic contexts

on sibilant perception (i.e., the magnitude of the VOWEL effect in perception) and the coarticulatory effects of vowels on sibilants (i.e., the VOWEL<sub>u</sub> and VOWEL<sub>l</sub>) did not turn out to be significant, echoing, for example, Kataoka's (2011) findings concerning the lack of a correspondence between the extent of coarticulatory fronting of /u/ and the extent of perceptual compensation for /u/-fronting. This fact is particularly striking since the vocalic context is a significant predictor in both group-level models for the perception and production data. Our findings offer potential insights into why previous studies that focused primarily on the extent of the coarticulatory effects on speech perception and production failed to observe a perception-production link. The present findings suggest that what is important is not the extent of coarticulation *per se*, but rather the effects that coarticulation has on the realization of the segments targeted. Such findings are consistent with recent studies demonstrating that listeners have fine-grained sensitivity to acoustic-phonetic cues needed to track their distributions (Clayards, Michael K. Tanenhaus, & Jacobs, 2008), as evidenced by listeners' sensitivity to within-category differences in, for example, reaction time (Pisoni & Tash, 1974), patterns of eye movements (McMurray, Tanenhaus, & Aslin, 2002), neural patterns of activities (Blumstein, Myers, & Rissman, 2005), as well as in category formation (Maye, Werker, & Gerken, 2002). Thus if one's articulation produces acoustic-phonetic cues for sibilants that result in more overlapping variances in certain coarticulated contexts (e.g., in the /u/ context) and not others (e.g., in the /ae/ context), that person's perceptual responses will exhibit more uncertainty in contexts where there is more overlapping variances.

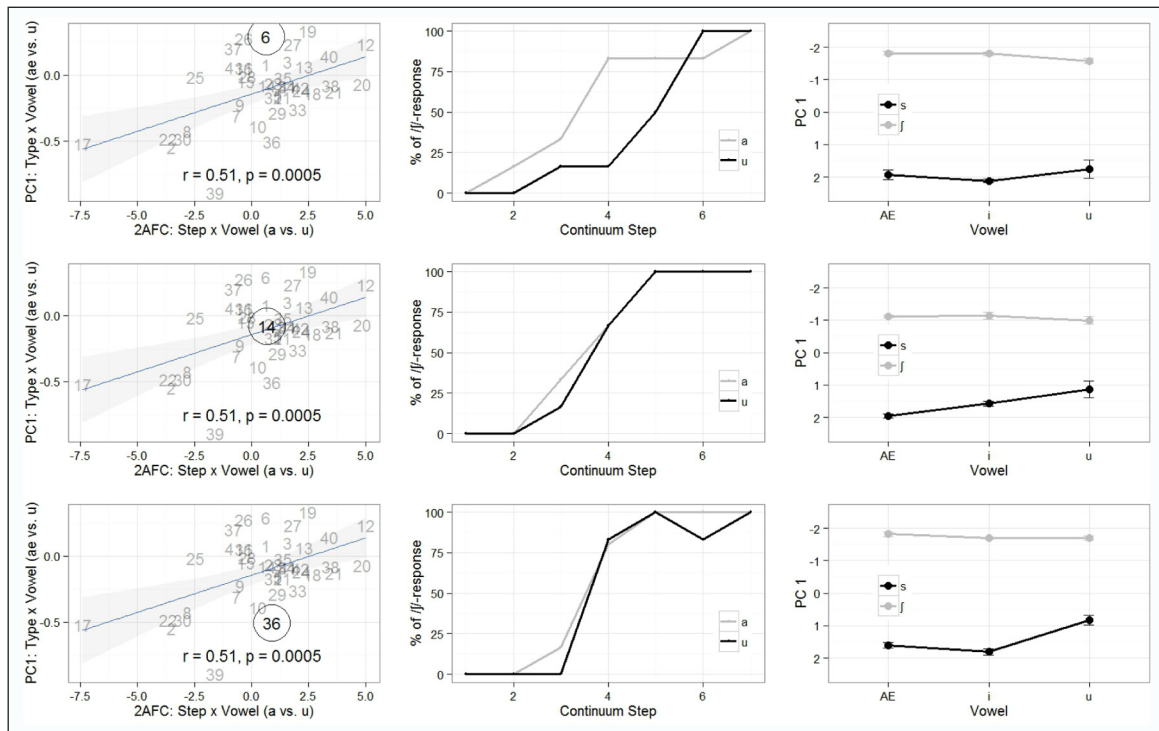
Further research is needed to ascertain the underlying mechanisms responsible for these individual differences in the perception-production linkage. Individuals might differ in the type of coarticulated speech input they encounter, which could result in different distributions of context-specific perceptual exemplars (cf. Pierrehumbert, 2002) or different understanding of how articulatory gestures are coordinated in different vocalic contexts, to the extent the universalist understanding of gestural knowledge can be relaxed. Variability in coarticulatory experiences need not stem from source variability *per se*, however. Individual differences in cognitive processing style (Yu, 2010) might also affect how individuals analyze coarticulated speech even if the input source is the same. As noted in the introduction, one type of cognitive processing style difference that has been linked to variation in the perception and production of coarticulation is autistic-like traits. While this study does not investigate this factor explicitly, this variable is unlikely to be the type of cognitive processing style difference that mediates the link between the perception and production of coarticulated speech since the effect of AQ is found only within females and not among males (at least according to the findings of Yu, 2010). Individuals might also vary in terms of oral-motor skills. From an auditorist perspective, for example, individuals might differ in how well they are at uncovering strategies for producing optimally perceivable acoustic signals. However, existing evidence for individual variation in oral-motor skills remain scant (cf. Diehl, Preston, & Bennetto, 2011; Iverson & Thelen, 1999; Mahler, 2012). Future research might also elucidate potential lurking variables not tied to knowledge of coarticulation (e.g., individual variation in cognitive sensitivity) that might account for, if only partially, the observed link between perception and production.

The fact that the correlation between perception and production is modest deserves some attention as well. The strength of the correlation hovers around 0.5, suggesting that there remains a sizable portion of the variance not explained by the correlation. Various factors might have contributed to this state of affairs. Given the perception task was administered prior to the production one, participants might have become hypersensitive to the /s/-/ʃ/ distinction and hyperarticulated in the production task, potentially limiting

the degree of coarticulation. In addition, the lack of filler items might have further heightened the participants' hyperarticulation tendencies. The reliance of original /da/ and /du/ stimuli to create the target stimuli in the perceptual study might bias listeners toward a more alveolar percept (i.e., /s/). Also, the fact that the production data but not the perceptual judgments were obtained in lexical contexts might have constrained the strength of the correlation between perception and production. Finally, as noted earlier, the perception and production tasks targeted different low vowels (i.e., /a/ in the perceptual task and /ae/ in the production one), which might have further constrained the nature of the perception-production correlation. To this end, it is worth noting that Jongman et al. (2000), who examined English fricatives in different vocalic contexts, including /a/ and /ae/, reported no significant difference between the effects of /a/ and /ae/ on fricative noise duration, spectral peak location, noise amplitude, and the various spectral moment measures they examined. The only acoustic difference observed concerns F2 onset frequency, at 1820 Hz before /ae/ and 1512 Hz before /a/. Given that F2 onset frequency was not a spectral measure included in the acoustic analysis conducted, this difference is unlikely to have affected the results of the perception-production correlation analysis in any significant way. Finally, as summarized succinctly in Beddor et al. (2018), successful communication depends on the listeners being malleable and able to perceptually adapt to a diverse set of variances, including phonetic context, speaker, speaking rate, novel experiences, and others. To the extent that listeners are able to adapt efficiently, the production system may not need to be similarly malleable. As such, the perception-production link is unlikely to be a perfect relation. In any event, despite these potential limitations, a significant correlation between the perceptual and production results is nonetheless found, which can be taken as further evidence of the robustness of the correlation observed.

Given that some of the participants fall outside the 95% confidence intervals (CIs) of the correlation trend lines, as illustrated by the scatterplots in **Figure 4**, further attention to the perception and production patterns of these individuals might yield fruitful insights. To this end, consider the relationship between the perceptual and production results of the three participants highlighted in **Figure 6**. All three participants have similar STEP  $\times$  VOWEL estimates ((P)articipant 6 = 0.651 (top), P14 = 0.676 (mid), P36 = 0.880 (bottom)), but they have wildly different TYPE  $\times$  VOWEL<sub>u</sub> estimates. P36's TYPE  $\times$  VOWEL<sub>u</sub> estimate ( $\beta = -0.507$ ) is below the trend line, showing a strong convergence between /s/ and /ʃ/ in PC1 values in the /u/ context, relative to the /ae/ context. P6's TYPE  $\times$  VOWEL<sub>u</sub> estimate ( $\beta = 0.294$ ) is above the trend line, showing almost no vocalic influence on PC1 at all. P14 falls within the 95% CIs (in fact, almost exactly on the trend line;  $\beta = -0.079$ ), and shows a modicum of vocalic influence on PC1.

The type of individual variation exemplified by participants 6, 14, and 36 is unlikely to be explainable by general mapping principles between perception and production. Further research is needed to ascertain the mechanism underlying this type of individual variability. To be sure, it would be useful to ascertain first just how stable is this type of variation across individuals. While there is evidence for stable individual differences across perceptual tasks tapping into listeners' knowledge of coarticulation (Yu & Lee, 2014), the stability of perceptual compensation, or more aptly, context-dependent perceptual response, across time remains to be shown. To the extent that the variability observed is stable, it points to a need to further explore other factors that might influence perception and production independently and together. To this end, it should be noted here that, while SEX interacts significantly with the acoustic realization of /s/ and /ʃ/ and modulates the influence of vocalic contexts, the two sexes do not appear to differ in the way perception and production are correlated with each other. To be sure, the sex-based



**Figure 6:** The perception-production relationships in three participants with similar STEP × VOWEL estimates. Each row illustrates the perceptual-production patterns of one participant. The participant targeted is circled in the scatterplot; the line and shading show a linear fit and 95% confidence intervals (CIs).

differences observed in the production study are particularly noteworthy from this perspective. While previous studies have identified sex-based differences in the acoustic realization of sibilants, until recently, little is known about the sex-specific nature of the vocalic influence on sibilant production. Such sex-based differences in sibilant realization might stem from potential sociolinguistic differences in articulatory strategies beyond basic variation in male and female vocal physiology (Strand, 1999; Stuart-Smith, 2007), including differences in terms of sexual orientation (Munson & Babel, 2007) or style construction (Podesva, Roberts, & Campbell-Kibler, 2001). In this study, males appear to exhibit a greater degree of vocalic influence in /s/ realization. However, no sex-based difference is observed in the perceptual responses.

Another source of potential variances might stem from contrast-related differences in articulation. Perkell, Matthies, et al. (2004), for example, show that, while there is generally a substantial contact of the underside of the tongue tip with the lower alveolar ridge during the production of /s/ but not /ʃ/, the degree of acoustic contrast between /s/ and /ʃ/ among a gender-balanced cohort of 20 native speakers of American English is related to their use of contact contrastively and in their discriminative performance. The most distinct sibilant productions were from participants who used contact in producing /s/ but not /ʃ/ and who had high discrimination scores, while the participants who did not use the contact differentially to produce the sibilants would produce the least distinct sibilant and would also discriminate synthetic sibilants less well. Individual variation in vocalic influence on the realization of the sibilant contrast might come about as a result of how individuals vary in whether the vocalic context influences the use of contact in producing the /s/ and /ʃ/ contrast.

Finally, as noted earlier, an increasing number of studies has argued for the importance of understanding individual variation in perception and production as a means to

understand sound change actuation (Baker et al., 2011; Beddor, 2009; Dimov, Katseff, & Johnson, 2012; Garrett & Johnson, 2013; Mielke, Baker, & Archangeli, 2016; Stevens & Harrington, 2014; Yu, 2010, 2013, 2016; Zellou, 2017). As coarticulation-induced variation in speech is often assumed to be a major source of phonetic precursors to sound change and sound patterns (Ohala, 1993a, 1993b), our findings suggest that some individuals within the same speech community are more advanced in reifying context-specific variation in speech production than others and this progression is mirrored in the individuals' perceptual behavior as well. Specifically, some individuals exhibit a greater reduction in contrast between /s/ and /ʃ/ in certain vowel contexts than others. This individual variability in context-dependent contrast reduction is reflected in how individuals perceive sibilants in the relevant contexts. Individuals whose sibilants are less distinct in the /u/ context are also less certain in their classification of sibilants in that context. Such findings are reminiscent of recent findings concerning the progression of sound change and categoricity in perception. In particular, Pinget (2015) investigated labiodental devoicing and labial stop devoicing in word onset position across various dialect regions in Dutch-speaking regions where these two instances of sound change in progress are at different stages of completion; fricative devoicing is more advanced than stop devoicing. She found that regions where devoicing was most advanced in production turned out to also be regions where perception was the least categorical. Taken together, Pinget's findings and the findings of the current study suggest that a reduction in contrast could lead to eventual innovation of a new sound pattern. In the present case, the contrast between /s/ and /ʃ/ in American English might eventually be partially or completely neutralized before a rounded vowel. While there remain major gaps in our understanding of the relationship between individuals who exhibit substantial context-dependent variation in speech perception and production and their social profiles within a speech community, to the extent that such individuals (the proverbial 'innovators' in change) become leaders within their community of practice, or have strong influence on such leaders themselves, their patterned variation might propagate throughout their respective communities.

### 3 Conclusion

This study establishes significant individual variability in the perception and production of /s/ and /ʃ/ in English across vocalic contexts. The variability is not random, however. There is a significant correlation between how individuals categorize sibilants in context-specific ways and how they realize their sibilants in the corresponding contexts. The present findings not only further the understanding of coarticulation in speech perception and production, they also have significant implications for research on sound change and language variation and change in general. Further research is needed to identify the causal mechanism behind the perception-production link identified in this study.

### Acknowledgements

This work was partially supported by National Science Foundation Grants BCS-0949754 and BCS-1827409. Special thanks go to the anonymous reviewers as well as the handling editors for their useful comments and criticisms. Many thanks also to the audiences at the Sound Change in Interacting Human Systems at University of California, Berkeley, and at the Linguistics department colloquia at the University of Arizona, Cornell University, Hong Kong University, McGill University, New York University, University of Toronto, Stanford, the Ohio State University and UCLA. All errors are of course my own.

I also thank the Linguistics in Open Access foundation (<https://www.lingoa.eu/>) for financial support of open access publication of *Laboratory Phonology*.



## Competing Interests

The author has no competing interests to declare.

## References

- American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association. DOI: <https://doi.org/10.1176/appi.books.9780890425596>
- Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177–189. DOI: <https://doi.org/10.1016/j.wocn.2011.09.001>
- Baker, A., Archangeli, D., & Mielke, J. 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, 23(3), 347–374. DOI: <https://doi.org/10.1017/S0954394511000135>
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. 2001. The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males, females, scientists and mathematicians. *Journal of Autism & Developmental Disorders*, 31, 5–17. DOI: <https://doi.org/10.1023/A:1005653411471>
- Bates, D., Maechler, M., & Bolker, B. 2011. lme4 [Computer software manual]. (R package version 0.999375-38).
- Beddor, P. S. 2009. A coarticulatory path to sound change. *Language*, 85(4), 785–832. DOI: <https://doi.org/10.1353/lan.0.0165>
- Beddor, P. S., Coetzee, A., Styler, W., McGowan, K., & Boland, J. 2018. The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language*, 94, 1–38. DOI: <https://doi.org/10.1353/lan.2018.0051>
- Beddor, P. S., Harnsberger, J., & Lindemann, S. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30, 591–627. DOI: <https://doi.org/10.1006/jpho.2002.0177>
- Beddor, P. S., & Krakow, R. A. 1998. Perceptual confusions and phonological change: How confused is the listener? In: Bergen, B. K., Plauché, M. C., & Bailey, A. C. (eds.), *Proceedings of the twenty-fourth annual meeting of the Berkeley Linguistics Society*, 320–334. Berkeley Linguistics Society. DOI: <https://doi.org/10.3765/bls.v24i1.1235>
- Beddor, P. S., & Krakow, R. A. 1999. Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America*, 106(5), 2868–2887. DOI: <https://doi.org/10.1121/1.428111>
- Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. 2013. The time course of perception of coarticulation. *Journal of the Acoustic Society of America*, 133(4), 2350–2366. DOI: <https://doi.org/10.1121/1.4794366>
- Blacklock, O. S. 2004. *Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods* (Unpublished doctoral dissertation). School of Electronics and Computer Science, University of Southampton, Southampton, UK.
- Blumstein, S. E., Myers, E. B., & Rissman, J. 2005. The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, 17(9), 1353–1366. DOI: <https://doi.org/10.1162/0898929054985473>
- Clayards, M., Michael, K., Tanenhaus, R. N. A., & Jacobs, R. A. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108, 804–809. DOI: <https://doi.org/10.1016/j.cognition.2008.04.004>
- Cole, J., Lindebaugh, G., Munson, C. M., & McMurray, B. 2010. Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics*, 38, 167–184. DOI: <https://doi.org/10.1016/j.wocn.2009.08.004>

- Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. 2001. Subcategorical mismatch and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16(5/6), 507–534. DOI: <https://doi.org/10.1080/01690960143000074>
- Diehl, J. J., Preston, J., & Bennetto, L. 2011. Diadochokinetic rate and accuracy in autism spectrum disorders. In: *International meeting for autism research: Diadochokinetic rate and accuracy in autism spectrum disorders*.
- Dimov, S., Katseff, S., & Johnson, K. 2012. Social and personality variables in compensation for altered auditory feedback. In: Sabater, M. J. S., & Recasens, D. (eds.), *The initiation of sound change: Perception, production, and social factors*, 185–210. Philadelphia: John Benjamins. DOI: <https://doi.org/10.1075/cilt.323.15dim>
- Fowler, C. A. 2006. Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, 68(2), 161–177. DOI: <https://doi.org/10.3758/BF03193666>
- Fowler, C. A., & Brown, J. M. 2000. Perceptual parsing of acoustic consequences of velum lowering from information of vowels. *Perception & Psychophysics*, 62(1), 21–32. DOI: <https://doi.org/10.3758/BF03212058>
- Garrett, A., & Johnson, K. 2013. Phonetic biases in sound change. In: Yu, A. C. L. (ed.), *Origins of sound change: Approaches to phonologization*, 51–97. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199573745.003.0003>
- Grosvald, M. 2009. Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics*, 37(2), 173–188. DOI: <https://doi.org/10.1016/j.wocn.2009.01.002>
- Grosvald, M., & Corina, D. 2012. Perception of long-distance coarticulation: An event-related potential and behavioral study. *Applied Psycholinguistics*, 33, 55–82. DOI: <https://doi.org/10.1017/S0142716411000105>
- Harrington, J., Kleber, F., & Reubold, U. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America*, 123(5), 2825–2835. DOI: <https://doi.org/10.1121/1.2897042>
- Heinze, G., Ploner, M., Dunkler, D., & Southworth, H. 2016. logistf [Computer software manual]. (R package version 1.22).
- Houde, J. F., & Jordon, M. I. 2002. Sensorimotor adaptation of speech 1: Compensation and adaptation. *Journal of Speech, Language and Hearing Research*, 45, 295–310. DOI: [https://doi.org/10.1044/1092-4388\(2002/023\)](https://doi.org/10.1044/1092-4388(2002/023))
- Hughes, G. W., & Halle, M. 1956. Spectral properties of fricative consonants. *Journal of the Acoustical Society of America*, 28, 303–310. DOI: <https://doi.org/10.1121/1.1908271>
- I.C.D-10 (ed.). 1994. *International classification of diseases* (10th ed.). Geneva, Switzerland: World Health Organisation.
- Iskarous, K., Shadle, C. H., & Proctor, M. I. 2011. Articulatory-acoustic kinematics: The production of American English /s/. *Journal of the Acoustical Society of America*, 129(2), 944–954. DOI: <https://doi.org/10.1121/1.3514537>
- Iverson, J., & Thelen, E. 1999. Hand, mouth and brain. *Journal of Consciousness Studies*, 6, 19–40.
- Jongman, A., Wayland, R., & Wong, S. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, 108(3), 1252–1263. DOI: <https://doi.org/10.1121/1.1288413>
- Kataoka, R. 2011. *Phonetic and cognitive bases of sound change* (Unpublished doctoral dissertation). University of California, Berkeley.

- Katseff, S. 2011. Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech*, 55(2), 295–308. DOI: <https://doi.org/10.1177/0023830911417802>
- Kong, E. J., & Edwards, J. 2016. Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, 59(1), 40–57. DOI: <https://doi.org/10.1016/j.wocn.2016.08.006>
- Li, F., Edwards, J., & Beckman, M. E. 2009. Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, 37(1), 111–124. DOI: <https://doi.org/10.1016/j.wocn.2008.10.001>
- Liberman, A., & Mattingly, I. G. 1985. The motor theory of speech perception revised. *Cognition*, 21(1), 1–36. DOI: [https://doi.org/10.1016/0010-0277\(85\)90021-6](https://doi.org/10.1016/0010-0277(85)90021-6)
- Livescu, K., Jyothi, P., & Fosler-Lussier, E. 2016. Articulatory feature-based pronunciation modeling. *Computer Speech & Language*, 36, 212–232. DOI: <https://doi.org/10.1016/j.csl.2015.07.003>
- Lotto, A. J., & Holt, L. L. 2006. Putting phonetic context effects into context: Commentary on Fowler (2005). *Perception & Psychophysics*, 68, 178–183. DOI: <https://doi.org/10.3758/BF03193667>
- Mahler, B. A. 2012. *Comparing motor speech skills of children with high functioning autism versus those of typically developing children using diadochokinetic tasks* (Unpublished master's thesis). The Ohio State University.
- Mahr, T., McMillan, B. T., Saffran, J. R., Weismer, S. E., & Edwards, J. 2015. Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, 142, 345–350. DOI: <https://doi.org/10.1016/j.cognition.2015.05.009>
- Mann, V. A., & Repp, B. H. 1980. Influence of vocalic context on perception of the [f]-[s] distinction. *Perception & Psychophysics*, 28, 213–228. DOI: <https://doi.org/10.3758/BF03204377>
- Maye, J., Werker, J. F., & Gerken, L. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101–B111. DOI: [https://doi.org/10.1016/S0010-0277\(01\)00157-3](https://doi.org/10.1016/S0010-0277(01)00157-3)
- McMurray, B., & Jongman, A. 2011. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, 118, 219–246. DOI: <https://doi.org/10.1037/a0022325>
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. 2002. Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2), B33–B42. DOI: [https://doi.org/10.1016/S0010-0277\(02\)00157-9](https://doi.org/10.1016/S0010-0277(02)00157-9)
- Mielke, J., Baker, A., & Archangeli, D. 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language*, 92(1), 101–140. DOI: <https://doi.org/10.1353/lan.2016.0019>
- Mirman, D. 2014. *Growth curve analysis and visualization using r*. Boca Raton, Florida: Chapman and Hall/CRC.
- Munson, B., & Babel, M. 2007. Loose lips and silver tongues, or, projecting sexual orientation through speech. *Language and Linguistic Compass*, 1(5), 416–449. DOI: <https://doi.org/10.1111/j.1749-818X.2007.00028.x>
- Newman, R. S., Clouse, S. A., & Burnham, J. L. 2001. The perceptual consequences of within-talker variability in fricative production. *Journal of the Acoustical Society of America*, 109, 1181–1196. DOI: <https://doi.org/10.1121/1.1348009>
- Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132–142. DOI: <https://doi.org/10.1016/j.wocn.2010.12.007>

- Nittrouer, S. 1995. Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America*, 97, 520–530. DOI: <https://doi.org/10.1121/1.412278>
- Ohala, J. J. 1993a. Coarticulation and phonology. *Language and Speech*, 36(2, 3), 155–170.
- Ohala, J. J. 1993b. The phonetics of sound change. In: Jones, C. (ed.), *Historical linguistics: Problems and perspectives*, 237–278. London: Longman Academic.
- Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. 2004. The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America*, 116(4), 2338–2344. DOI: <https://doi.org/10.1121/1.1787524>
- Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Guenther, F. H., et al. 2004. The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language and Hearing Research*, 47, 1259–1269. DOI: [https://doi.org/10.1044/1092-4388\(2004/095\)](https://doi.org/10.1044/1092-4388(2004/095))
- Pierrehumbert, J. 2002. Word specific phonetics. In: Gussenhoven, C., & Warner, N. (eds.), *Laboratory phonology vii*, 101–139. Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110197105.101>
- Pinget, A.-F. 2015. *The actuation of sound change*. Utrecht, the Netherlands: LOT.
- Pisoni, D. B., & Tash, J. 1974. Reaction times to comparisons within and across category. *Perception and Psychophysics*, 15(2), 285–290. DOI: <https://doi.org/10.3758/BF03213946>
- Podesva, R., Roberts, S., & Campbell-Kibler, K. 2001. Sharing resources and indexing meanings in the production of gay styles. In: Campbell-Kibler, K., Podesva, R., Roberts, S., & Wong, A. (eds.), *Language and sexuality: Contesting meaning in theory and practice*, 175–189. Stanford, CA: CSLI.
- Repp, B. H. 1981. Two strategies in fricative discrimination. *Perception & Psychophysics*, 30(3), 217–227. DOI: <https://doi.org/10.3758/BF03214276>
- Shadle, C. H., & Mair, S. 1996. Quantifying spectral characteristics of fricatives. In: *ICSLP 96. Proceedings of the Fourth International Conference on Spoken Language Processing*, 1521–1524. IEEE. DOI: <https://doi.org/10.1109/ICSLP.1996.607906>
- Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. 2009. Perceptual recalibration of speech sounds following speech motor learning. *Journal of the Acoustical Society of America*, 125(2), 1103–1113. DOI: <https://doi.org/10.1121/1.3058638>
- Soli, S. D. 1981. Second formants in fricatives: Acoustic consequences of fricative–vowel coarticulation. *Journal of the Acoustical Society of America*, 70, 976–984. DOI: <https://doi.org/10.1121/1.387032>
- Sonderegger, M., & Yu, A. C. L. 2010. A rational account of perceptual compensation for coarticulation. In: Ohlsson, S., & Catrambone, R. (eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 375–380. Austin, TX: Cognitive Science Society.
- Stevens, M., & Harrington, J. 2014. The individual and the actuation of sound change. *Loquens*, 1(1), e003. DOI: <https://doi.org/10.3989/loquens.2014.003>
- Strand, E. A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Psychology*, 18, 86–99. DOI: <https://doi.org/10.1177/0261927X99018001006>
- Stuart-Smith, J. 2007. Empirical evidence for gendered speech production: /s/ in Glaswegian. In: Cole, J., & Hualde, J. I. (eds.), *Laboratory Phonology 9*, 65–86. New York: Mouton de Gruyter.




- Tomiak, G. R. 1990. *An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents* (Unpublished doctoral dissertation). SUNY Buffalo.
- Viswanathana, N., Magnusona, J. S., & Fowler, C. A. 2010. Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance* Volume, 36(4), 1005–1015. DOI: <https://doi.org/10.1037/a0018391>
- Whalen, D. H. 1981. Effects of vocalic formant transitions and vowel quality on the English [s]–[ʃ] boundary. *Journal of the Acoustical Society of America*, 69(1), 275–282. DOI: <https://doi.org/10.1121/1.385348>
- Whalen, D. H. 1991. Perception of the English /s-/ʃ/ distinction relies on friction noises and transitions, not on brief spectral slices. *Journal of the Acoustical Society of America*, 90(4), 1776–1785. DOI: <https://doi.org/10.1121/1.401658>
- Whalen, D. H. 1999. Three lines of evidence for direct links between production and perception in speech. In: Ohala, J. J., Hasegawa, Y., Ohala, M., Granville, D., & Bailey, A. (eds.), *Proceedings of the 14th ICPhS*. San Francisco: The Regents of the University of California.
- Wickham, H. 2012. plyr [Computer software manual]. (R package version 1.8).
- Yu, A. C. L. 2010. Perceptual compensation is correlated with individuals' 'autistic' traits: Implications for models of sound change. *PLoS One*, 5(8), e11950. Retrieved from. DOI: <https://doi.org/10.1371/journal.pone.0011950>
- Yu, A. C. L. 2013. Individual differences in socio-cognitive processing and the actuation of sound change. In: Yu, A. C. L. (ed.), *Origins of sound change: Approaches to phonologization*, 201–227. Oxford, UK: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199573745.003.0010>
- Yu, A. C. L. 2016. Vowel-dependent variation in Cantonese /s/ from an individual-difference perspective. *Journal of Acoustical Society of America*, 139(4), 1672–1690. DOI: <https://doi.org/10.1121/1.4944992>
- Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. 2013. Phonetic imitation from an individual-difference perspective: Subjective attitude, personality, and 'autistic' traits. *PLOS ONE*, 8(9), e74746. DOI: <https://doi.org/10.1371/journal.pone.0074746>
- Yu, A. C. L., & Lee, H. 2014. The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study. *Journal of the Acoustical Society of America*, 136(1), 382–388. DOI: <https://doi.org/10.1121/1.4883380>
- Zellou, G. 2017. Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, 61(1), 13–29. DOI: <https://doi.org/10.1016/j.wocn.2016.12.002>
- Zellou, G., Dahan, D., & Embick, D. 2017. Imitation of coarticulatory vowel nasality across words and time. *Language, Cognition and Neuroscience*, 32(6), 776–791. DOI: <https://doi.org/10.1080/23273798.2016.1275710>
- Zellou, G., Scarborough, R. A., & Nielsen, K. 2016. Phonetic imitation of coarticulatory vowel nasalization. *The Journal of the Acoustical Society of America*, 140(5), 3560–3575. DOI: <https://doi.org/10.1121/1.4966232>



**How to cite this article:** Yu, A. C. L. 2019 On the nature of the perception-production link: Individual variability in English sibilant-vowel coarticulation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10(1): 2, pp. 1–29. DOI: <https://doi.org/10.5334/labphon.97>

**Submitted:** 14 June 2017    **Accepted:** 04 December 2018    **Published:** 06 February 2019

**Copyright:** © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 