

Vowel-dependent variation in Cantonese /s/ from an individual-difference perspective

Alan C. L. Yu^{a)}

Phonology Laboratory, Department of Linguistics, University of Chicago, Chicago, Illinois 60615, USA

(Received 16 November 2014; revised 2 March 2016; accepted 17 March 2016; published online 7 April 2016)

Individual variation is ubiquitous in the acoustic realization of human speech; however, little is known about the nature of individual differences in coarticulation. Through an in-depth case study of the temporal dynamics of vocalic influences on the acoustic realization of Cantonese /s/, this study demonstrates that coarticulatory effects may vary by the sex and self-reported autistic-like traits of the individual. These findings have significant implications for research in phonetics, phonology, and sound change. © 2016 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4944992>]

[CGC]

Pages: 1672–1690

I. INTRODUCTION

The acoustic realization of human speech is marked by substantial individual variation. While gender (Stuart-Smith, 2007) and dialects (Byrd, 1992) are two commonly assumed factors that modulate such variation, individual talkers of the same gender and dialect have nonetheless been shown to differ, for example, in vowel formant frequencies (Hillenbrand *et al.*, 1995), voice onset time (Allen *et al.*, 2003; Theodore *et al.*, 2009), as well as in frication centroid frequencies and skewness (Newman *et al.*, 2001). Individual variability may come from differences in vocal tract physiology, particularly related to the nature of sexual dimorphism of the vocal tract (Vorperian *et al.*, 2011), vocal tract size and shape (Peterson and Barney, 1952), idiosyncratic articulatory habits (Klatt, 1986), and behavioral/etiological factors (Sachs *et al.*, 1973; Ohala, 1994). Little is known regarding the nature of individual differences in coarticulation in speech, however. Previous studies on individual differences in speech have mainly focused on the overall variability in speech production without focusing on the nature of the contextual influences. To be sure, there are notable exceptions (Baker *et al.*, 2011; Grosvald, 2009; Grosvald and Corina, 2012; Harrington *et al.*, 2008; Kataoka, 2011); nonetheless, there remain many intriguing aspects regarding the nature of individual variability in coarticulation to be investigated. In particular, why individuals differ in the extent of coarticulation remains a question largely unexplored. This lacuna might reflect the belief that coarticulation is biomechanical in origin, and to the extent that humans share similar anatomical makeup and biomechanical tendencies, systematic variation in coarticulation across individuals is not expected. However, earlier studies on coarticulation have already cast doubt on the strict biomechanical interpretation of coarticulation. Previous studies have identified language-specific differences in coarticulation (Manuel, 1990), arguing for a role of contrastivity in constraining degrees of coarticulation (though, see Choi and Keating, 1991 and Beddor *et al.*, 2002

for findings that are not consistent with this theory). There is also evidence suggesting that coarticulation is, to a certain extent, planned (Whalen, 1990; Solé, 2007). Some further argue that coarticulation might be designed in part for the benefit of the listeners (Pycha, 2015; Scarborough, 2004, 2013; Wright, 2004). The degree of coarticulation might change with time, as suggested by Harrington *et al.* (2008), who found younger speakers of Southern British English, relative to the older speakers, exhibit stronger u-fronting; this phenomenon, they argued, could be linked synchronically to the fronting effects of a preceding anterior consonant.

Perhaps more intriguing are recent reports that found significant individual variability in perceptual compensation for coarticulation (Repp, 1981; Beddor, 2009; Yu and Lee, 2014). In particular, from speech perception theories that assume the objects of perception and production are one and the same (i.e., phonetic gestures of the vocal tract, e.g., Fowler, 2006), variability in perceptual compensation for coarticulation should find analogues in the production of coarticulated speech as well. The source of the individual variation in perceptual compensation remains a matter of debate, however. Some attribute variability in perceptual compensation to differences in perceptual grammar across individuals (Beddor, 2009). Others point to community-level sound change in progress (Harrington *et al.*, 2008), where individuals who are more advanced in phonologizing a coarticulatory pattern are compensating less than those who are comparatively less advanced in the sound change (see also Zellou and Tamminga, 2014). Of particular interest here is the link between inter-individual variability in processing context-dependent speech information and individual differences in cognitive processing style (i.e., psychological dimensions representing preferences and consistencies in an individual's particular manner of cognitive functioning, with respect to acquiring and processing information). Motivated in part by recent studies which show that autistic traits, or the broader autism phenotypes, are not restricted to individuals with clinical diagnoses of autism (Constantino and Todd, 2003; Robinson *et al.*, 2011; Lundström *et al.*, 2012), some scholars have explored autistic-like traits as an

^{a)}Electronic mail: aclyu@uchicago.edu

individual-difference dimension for indexing differences in cognitive processing style. While the clinical diagnosis of autism spectrum disorder (ASD) involves difficulties in social development and communication alongside the presence of unusually strong repetitive behavior or “obsessive” interests (American Psychiatric Association, 2013; ICD-10, 1994), cognitive theories of ASD have long argued that individuals with autism have different cognitive processing styles than neurotypicals. Individuals with ASD might show “detail-focused processing in which features are perceived and retained at the expense of global configuration and contextualized meaning” (Happé, 1999), while individuals with typical central coherence may parse incoming information for higher-level meaning often at the expense of memory for detail (Happé and Frith, 2006). Individuals with ASD also tend to have superior abilities with respect to the processing of low level perceptual information but exhibit difficulties with the integration of higher order information (Bonnell *et al.*, 2003; Mottron *et al.*, 2006). In the context of perceptual compensation for coarticulation, Yu (2010) found that the magnitude of perceptual compensation for the vocalic effect on sibilant perception is modulated by the listeners’ sex and autistic-like traits (Yu, 2010), as measured by the autism spectrum quotient (AQ) (Baron-Cohen *et al.*, 2001b), a short, self-administered scale for identifying the degree to which any individual adult of normal IQ may have traits associated with ASD. Yu (2010) reported that English-speaking females with low AQ are less likely to perceptually compensate for coarticulation (see also Turnbull, 2015).

In light of previous studies that reported an association between the perception of coarticulated speech and autistic-like traits, there might be an association between autistic-like traits and the production of coarticulated speech, assuming there is a linkage between perception and production (an assumption that is still a matter of great debate). However, the nature of the association between the production of coarticulated speech and autistic-like traits would differ depending on the nature of the relationship between the perception and production of coarticulated speech. If the magnitude of coarticulation in production mirrors the magnitude of perceptual compensation, as predicted by gestural accounts of perceptual compensation where the intended gesture is recovered by accessing tacit knowledge of the acoustic consequences of the candidate articulatory motor commands [i.e., motor theory of speech perception (Mann, 1980)], females with high AQ and male speakers in general are expected to exhibit stronger coarticulation since they have been shown to exhibit stronger perceptual compensation. Similar to the motor theory of speech perception, direct realism emphasizes the importance of gestural knowledge in speech perception and production. Specifically, direct realism argues that gestures referenced in perception and production are one and the same and perceptual compensation for coarticulation is a consequence of the listeners being attuned to the temporal overlap of the gestures in speech production and adjusting their percepts accordingly (Fowler, 2006). To the extent that direct realism makes any direct claims about individual variability in perception and production, we might expect a similarly direct

link between a speaker’s coarticulation magnitude in production and his/her compensatory response in perception.

The predicted relationship between autistic-like traits and the production of coarticulated speech would be quite different from the perspective of the so-called “C-CuRE” approach to perceptual compensation for coarticulation (Cole *et al.*, 2010; McMurray and Jongman, 2011). This approach assumes that the incoming acoustic cues are initially encoded veridically, but cues are recoded in terms of their divergence from expected values as different sources of variance are categorized. When such a listener-turn-speaker computes his/her production targets at the expectation-adjusted level, the production targets of a sound category are predicted to 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 may be 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 the auditory listeners of Repp (1981)]. Following this logic, females with high AQ and males in general are expected to exhibit less coarticulation on account of their strong compensatory responses to coarticulation in perception.

Current theories concerning the nature of coarticulation in production offer additional rationale for the hypothesized link between variation in coarticulation and autistic-like traits. Theories that characterize speech as a dynamic balance between speaker- and listener-oriented forces [see, e.g., the H & H model of Lindblom (1990)] assume that variation in coarticulation results from differences in speech style, which may vary on a continuum of hyper- and hypo-speech depending on whether it is the need of the speaker or the listener that is emphasized. In contrast to this type of speaker-oriented approach to coarticulation, which equates variation in coarticulation as a side-effect of variation in articulatory efforts, some have recently argued that coarticulation might be designed in part for the benefit of the listeners (Pycha, 2015; Scarborough, 2004, 2013; Wright, 2004). That is, while coarticulation may diminish the acoustic distinctiveness of a segment by overlapping it with another, this overlap spreads the acoustic properties of a given segment across other segments, thus providing redundant temporally distributed cues for the spreading segment (Mattingly, 1981; Wright, 2004). To the extent that listeners can make use of coarticulatory cues efficiently, listeners’ comprehension would be enhanced. Listeners can also take advantage of the contextually predictable nature of coarticulation and use coarticulatory information to identify or predict other portions of the signal. If coarticulation is indeed for the benefit of the listeners (e.g., as a perceptually useful source of linguistic information to facilitate listeners’ perception, Pycha, 2015; Scarborough, 2004, 2013; Wright, 2004), rather than simply as a consequence of varying articulatory efforts under different communicative conditions (Lindblom, 1990; Lindblom *et al.*, 1995), speakers must be able to create, maintain, and update a detailed mental representation of their interlocutor’s knowledge, belief, and intentions in real time. Thus, we may extend the finding that speakers adjust

the degree of coarticulation dynamically across communicative conditions and hypothesize that intrinsic differences in theory-of-mind abilities across individuals might correlate with differences in coarticulation that are not specific to particular communicative contexts. As a person's level of autistic-like traits is negatively correlated with the individual's theory-of-mind abilities (Baron-Cohen *et al.*, 2001a), individuals with more autistic-like traits, and thus poorer theory-of-mind abilities, should exhibit less coarticulation in speech given their poorer ability to attribute complex mental states to others (cf. Turnbull, 2015). This work explores this association in detail.

Understanding the mechanisms underlying individual variation in speech production is increasingly important, particularly for research on sound change and propagation (Stevens and 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 *et al.*, 2011; Yu, 2013) or different tendencies to attach social meaning to linguistic differences (Garrett and Johnson, 2013). Some authors have argued that listeners' perceptual weights for coarticulatory effects might be reflected in their own productions (Beddor, 2009), while others hypothesize that listeners who compensate less for coarticulation are more likely to initiate sound change (Yu, 2010; Yu and Lee, 2014). However, sound change is only actualized when a change in perception is reflected in a change in production; therefore, establishing systematic individual differences in coarticulation (i.e., context-dependent variability) and identifying the underlying sources of such variability in speech production is an important step toward substantiating the individual-difference model of sound change and might help explain why sound change happens at all and, conversely, why sound change is so rarely actuated even though the phonetic pre-conditions are always present in speech.

To further this goal of understanding individual variability in coarticulated speech, this study focuses on the variable realization of /s/ in Cantonese. The reason for focusing on Cantonese /s/ is two-fold. To begin with, traditional descriptions of the Cantonese sibilant fricative vary quite drastically from author to author. This sibilant, which is voiceless and only found prevocally, has been variously transcribed by different researchers as [s] (Jones and Woo, 1912), [ç] (Chao, 1947), or [ʃ] (Wang, 1937). Most recently, it was described as varying between English alveolar [s] and post-alveolar [ʃ] (Bauer and Benedict, 1997). Hashimoto (1972) suggests the variability might be a matter of individual stylistic differences. One major reason for this variability might be due to the heavy vocalic influence on the realization of this sibilant. Some suggest that this sibilant takes on quality similar to [ç] or [ʃ] before /y/ and sometimes before /i/ (Bauer and Benedict, 1997), while others contend that the sibilant is similar to the English palatoalveolars before round vowels in general (Kao, 1971; Cheung, 1986). Hashimoto (1972), on the other hand, suggests that the sibilant is pronounced with some degree of palatalization before high front vowels. Still others consider [s] and [ʃ] to be in free variation

with each other (Pulleyblank, 1996). Despite these controversies, a detailed acoustic description of Cantonese /s/ in different vocalic environments remains elusive. This study aims to remedy this void in the literature.

Second, based on evidence from static palatography, Lee and Zee (2010) conclude that the Cantonese /s/ is a laminal alveolar articulation, produced with the tongue blade making contact on the alveolar ridge. When preceding high vowels /i/ and /y/, [s] shows a reduction in the width between the lateral contact on each side of the palatal region relative to the widths for [s] before lower vowels, even if the actual articulatory target region is not affected by the vowel context as much. Given that the underlying articulatory target of /s/ does not seem to differ across vocalic contexts, it suggests that the variable acoustic realization of the /s/ in Cantonese might be better characterized as the result of coarticulation from neighboring vowels, rather than as a categorical difference as would be captured by phonological rules such as /s/ → [ʃ]/_[+round]. From this point of view, a careful examination of the temporal dynamics of vocalic influence on the acoustic realization of /s/ is paramount. If the magnitude of vocalic influence on the acoustics of /s/ production varies across time, it will provide strong evidence for the coarticulatory interpretation of /s/ variation.

The organization of this study is as follows: first, a detailed examination of the acoustic properties of Cantonese /s/ in different vocalic contexts is presented; second, potential sources of individual variability in coarticulation are explored by examining whether the realization of /s/ varies across the sexes and across individuals with different self-reported autistic-like traits. This study concludes with a discussion of the implications of our findings for phonetics and phonology, and for the study of sound change in general.

II. THE ACOUSTICS OF CANTONESE /s/

A. Materials

The recordings were originally obtained as part of a larger study of Cantonese phonetics and phonology. This study focuses on a set of ten target words in Cantonese: [sy] "book," [sy] "potato," [so] "comb," [so] "silly," [si] "poem," [si] "time," [se] "a little bit," [se] "snake," [sa] "sand," [san] "god." The symbol ˩ marks a high level tone, ˨ a low level tone, and ˨˨ a low falling tone.

B. Participants

105 native speakers of Hong Kong Cantonese (sixty-two females) with no reported history of speech, language, or hearing problems were recorded in Hong Kong as part of a larger study of Cantonese phonetics and phonology. Each subject received a nominal fee or course credit for participating in the study.

C. Procedure

Each participant was digitally recorded in a quiet room individually at a sampling rate of 44 100 Hz reading three blocks of the target stimuli, presented in one of two pseudo-randomized lists of target words in the carrier sentence,

[ŋɔ̃ | tok̃_]_ [pei | nei | tʰeŋ]] “I read_for you to hear.” A total of thirty target stimuli were analyzed from each participant. The stimuli were presented in traditional Chinese characters. All subjects also completed an online survey which included, among others, questions about the subject’s age, sex, second language knowledge, and the full 50-question AQ questionnaire (Baron-Cohen *et al.*, 2001b). Participants were given the option to complete either the Chinese or English version of the AQ.

Fricative segmentation involved the simultaneous consultation of waveforms and wideband spectrograms. Fricative onset was defined as the point at which high frequency energy 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.

To measure anticipatory vocalic influence as well as other details of fricative production involving tongue tip configuration, the spectral properties of sibilant noise in terms of seven spectral and duration parameters including the spectral peak frequency, the first four spectral moments, the relative frequency amplitude ratio, and the total fricative duration, were examined. The spectral peak is a commonly used feature for distinguishing place of articulation of sibilants. The overall spectral shape of a sibilant is determined by the size and shape of the oral cavity in front of the constriction. The longer this anterior cavity, the more defined the resulting spectrum (Stevens, 1998). /ʃ/ typically exhibits a midfrequency peak at around 2.5–3 kHz, which often corresponds to the F3 of the following vowel while /s/ displays a spectral peak at primarily around 4–5 kHz on account of a shorter anterior cavity, at least relative to /ʃ/. The location of the spectral peak is partly dependent on the speaker (Hughes and Halle, 1956) and the vowel context (Soli, 1981).

Another frequently invoked method for examining the spectral properties of fricative noise is spectral moment analysis, in which a power spectrum is modeled as a random probability distribution from which the first four moments (mean, variance, skewness, and kurtosis) are computed (Forrest *et al.*, 1988; Shadle and Mair, 1996; Jongman *et al.*, 2000). While the utility of moment analysis has been questioned for the discrimination of different place of articulation in English fricatives (Shadle and Mair, 1996), this method has been found to be useful for characterizing the differences between /s/ and /ʃ/ and the spectral characteristics of these sibilants in various vocalic contexts (Shadle and Mair, 1996). Mean and variance reflect the average energy concentration and range, respectively. The spectral mean is negatively correlated with the length of the front resonating cavity while the variance or standard deviation has been suggested to differentiate apical and laminal tongue postures. The first spectral moment has been shown to well distinguish between /s/ and /ʃ/ in English (Shadle and Mair, 1996; Jongman *et al.*, 2000), Aleut, Apache, Chickasaw, Gaelic, Hupa, Monotana Salish, and Toda (Gordon *et al.*, 2002) and show clear separation for sibilants in different vowel contexts (Nittrouer, 1995) and across 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 (English: Tomiak, 1990; Jongman *et al.*, 2000). Li *et al.* (2009), on the other hand, found English and Japanese /s/ to have a more diffused shape (higher standard deviation) than /ʃ/ in English and /ç/ in Japanese. Skewness refers to the overall slant of the energy distribution. A positive skewness suggests a negative tilt with a concentration of energy in the lower frequencies, as in the case of /ʃ/. Kurtosis measures the peakiness of the distribution and maybe useful for distinguishing fricatives with tongue posture differences (Li *et al.*, 2009). The higher the value, the more peaked the distribution. For example, /s/ has a high kurtosis compared to /ʃ/ in English (Jongman *et al.*, 2000; Li *et al.*, 2009) while, in Japanese, /s/ has a smaller kurtosis value than /ç/, which has a more compact and symmetrical distribution of energy around a single peak (Li *et al.*, 2009). Shadle and Mair (1996) report a particularly high kurtosis value for /s/ around /u/ compared to other vowels, suggesting a relationship between lip rounding and defined peaks. Spectral moments have been found to index sex differences. Females exhibit significantly higher spectral mean, variance and kurtosis and lower skewness than men (Jongman *et al.*, 2000).

To assess the degree of “palatalization” (i.e., the tongue posture difference), a measure of the so-called amplitude ratio (ampRatio), defined as the difference in dB between the F2 amplitude and the amplitude of the most prominent peak above the F2 region (Li *et al.*, 2007), was included. Specifically, the average amplitude within the F2 region was estimated by taking the F2 at the onset of the following vowel and defining a 1000 Hz band around that peak. The amplitude ratio was then calculated by subtracting the average amplitude within the F2 region from the average amplitude ratio of a 1000 Hz band centered on the highest amplitude peak above the F2 region. The amplitude ratio is expected to be higher when the tongue posture is flatter and more palatal due to the lack of resonance in the F2 region, while it should be smaller (i.e., higher F2 amplitude) if the front and the back cavities are coupled. Li *et al.* (2007), for example, found /ç/ to have a higher amplitude ratio than /s/ in Japanese.

A custom-made PRAAT script automatically extracted the spectral measurements, taken from the entire frequency range (0–12 000 Hz), using a 40 ms Hamming window with pre-emphasis at 80 Hz, centered at ten points (at 11.11% increments of the fricative’s duration from 0% to 100%) during the fricative. Measurements at 0% and 100% were not included in the analysis. The same script also measured the duration of the sibilant, the word duration, and the onset F2 of the following vowel. The setting in PRAAT used to estimate the onset F2 was an LPC analysis specified for five formants calculated with a window length of 0.025 ms over a range from 0 to 5000 Hz for male and 0 to 5500 Hz for female productions.

D. Statistical analysis

The seven acoustic measures (centroid frequency, standard deviation, kurtosis, skewness, peak frequency, ampRatio, and sibilant duration) were modeled separately using linear

mixed-effects regression fitted in R, using the `lmer()` function from the `lme4` package (Bates *et al.*, 2011). Because of the heavily skewed distribution of the kurtosis values, the kurtosis measure was log-transformed [i.e., $\log(k+3)$, where k is the actual kurtosis value] for further statistical analysis. The value 3 was added to all kurtosis values to ensure that they are positive prior to log-transformation.

All acoustic measures were tested for the effects of trial order (TRIAL; 1–30) and vowel quality (i, e, a, o, y). Vowel quality was contrast-coded. The two main contrasts are `VOWELRound` [round (o, y) vs unround (i, e, a)] and `VOWELHigh` [high (i, y) vs mid (e, o)]. Previous reports suggested that the allophonic deviation from /s/ primarily occurs in the /y/ context. This hypothesis is tested by comparing /y/ with /i/, its unrounded counterpart (`VOWELy/i`), and with /o/, its mid rounded and back counterpart (`VOWELy/o`). For the non-duration-based measures, sibilant DURATION was also included as a main predictor. The spectral measures (i.e., excluding sibilant duration) were also tested for the effects of measurement position (POSITION: 0–7), as well as the interaction between POSITION and VOWEL to examine the dynamic aspects of the vocalic influence.

Following Iskarous *et al.* (2011), a growth-curve model (Singer and Willett, 2003) was used to analyze the effects of vocalic context on the temporal trends in the acoustic dimensions during the production of /s/. In particular, the linear and quadratic terms of POSITION were used to predict the variability with time [i.e., POSITION, which goes from measurement point 0 (11.11%) to point 7 (88.88%)]. Only linear and quadratic effects were used since higher orders were found to be consistently insignificant. If a main effect parameter increases with time, POSITION will have a positive linear coefficient. If there is a significant uniform curvature with respect to time, it will have a negative quadratic coefficient if it is concave downward (i.e., the parabola has a vertex on top); an upward concave curvature (i.e., the vertex is at the bottom) would have a positive quadratic coefficient. In the models below, there is generally a slow down of the rate of change toward the offset of the sibilant as the vertexes are always to the right of zero along the x axis.

To reduce multicollinearity between predictors, continuous variables (not including POSITION) were centered and z-scored. The models also included by-subject random intercepts to allow for subject-specific variation in each of the acoustic measures as well as by-subject random slopes for VOWEL, and the linear and quadratic effects of POSITION, if relevant, to allow for by-subject variability in the effects of vowels as well as the measurement positions on the acoustic measure. Models with by-subject random slopes for the interaction between VOWEL and POSITION did not converge consistently across acoustic parameters and were therefore not included in the final analysis. By-subject random slopes for TRIAL and DURATION that did not correlate with the by-subject intercept were also included to account for by-subject variability in the trial order effect and the effect of sibilant duration. Finally, by-item random intercepts were also included to allow for word-specific variation in the acoustic measure. A series of log-likelihood tests confirmed that the inclusion of the by-subject random intercepts and

random slopes, as well as the by-item random intercepts, are significant ($p < 0.01$) in all the regression models for the sibilant acoustic measures. The model formula in `lme4` style for the non-duration-based spectral measures (M) was $M \sim \text{TRIAL} + \text{DURATION} + \text{VOWEL} * (\text{POSITION} + \text{POSITION}^2) + (0 + \text{DURATION} + \text{TRIAL} | \text{SUBJECT}) + (1 + \text{VOWEL} + \text{POSITION} + \text{POSITION}^2 | \text{SUBJECT}) + (1 | \text{Word})$. The model for sibilant duration was $\text{DURATION} \sim \text{TRIAL} + \text{VOWEL} + (0 + \text{TRIAL} | \text{SUBJECT}) + (1 + \text{VOWEL} | \text{SUBJECT}) + (1 | \text{Word})$. POSITION^2 represents the quadratic term of POSITION. 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 3% of the data for each acoustic measure 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.

E. Results

Table I provides a descriptive summary of the moment analysis as well as peak frequency, `ampRatio`, and sibilant duration. Models for each spectral measure are discussed in detail below. A summary of the main and interaction effects of all predictors for each acoustic measure is given in Table II. Only significant predictors, determined at the $p < 0.05$ level are reported; 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).

1. Spectral mean

The average spectral mean at sibilant onset (i.e., 11% of the sibilant duration) is 7613 Hz. The regression analysis reveals a significant effect of TRIAL, suggesting that the spectral mean increases by 37 Hz across every nine trials (i.e., one standard deviation of TRIAL; $\beta = 37.12$, $t = 3.01$, $p < 0.001$). There is a main linear effect of measurement position ($\beta = 306.13$, $t = 12.49$, $p < 0.001$), suggesting there is a 306 Hz increase in spectral mean between each measurement position from the sibilant onset to its offset. There is also a significant negative curvature effect of POSITION ($\beta = -26.73$, $t = -11.01$, $p < 0.001$). Together, the by-POSITION trend of the spectral mean is the sum of a linear function with a positive slope and a quadratic function with downward concavity, indicating that the spectral mean has a slowing downward trend toward the offset of the /s/. No significant effect of sibilant duration is observed.

The top left panel of Fig. 1 illustrates the various vocalic effects on the spectral mean. Visual inspection suggests that the spectral mean exhibits different profiles, both in terms of frequency range and temporal dynamics, before rounded and unrounded vowels. This is confirmed by the regression model. While there is not a significant main effect of `VOWELRound`, `VOWELRound` interacts with POSITION both linearly ($\beta = -64.24$, $t = -6.87$, $p < 0.001$) and quadratically ($\beta = -6.15$, $t = -4.82$, $p < 0.001$). Taken together, these findings show that, relative to the unrounded vowel contexts,

TABLE I. Descriptive statistics for the moment analysis as well as peak frequency, amplitude ratio, and sibilant duration in different vocalic contexts.

Vowel	N	Mean	Standard deviation	Skewness	Kurtosis	Peak	ampRatio	Duration
a	105	8215(702)	1749(239)	-0.13(0.46)	0.21(0.83)	8011(1108)	51(3.6)	161(26)
e	105	8328(735)	1697(243)	-0.20(0.47)	0.37(0.91)	8127(1139)	52(3.5)	170(26)
i	105	8539(722)	1618(224)	-0.28(0.48)	0.61(1.02)	8434(1115)	50(3.3)	202(29)
o	105	7924(787)	2140(295)	-0.40(0.69)	0.22(1.52)	7760(1518)	51(3.8)	170(27)
y	105	8062(872)	2079(374)	-0.67(0.74)	1.04(2.19)	8086(1553)	49(4)	203(34)

the linear trend of the spectral mean is less positive (i.e., spectral mean does not rise from the sibilant onset to its offset as much before rounded vowels as before unrounded ones) and the curved trend is more downward concave before rounded vowels.

Concerning the effects of vowel height, Fig. 1 suggests that the higher the neighboring vowel, the higher the spectral mean. This observation is borne out in the regression model. There is a significant main effect of $VOWEL_{High}$ ($\beta = 235.49$, $t = 6.58$, $p < 0.001$). Unlike the effect of vocalic rounding, the vowel height effect is not temporally dependent, as indicated by the lack of interaction between $VOWEL_{High}$ and POSITION.

With respect to the claim that /s/ only alternates in the /y/ context, this does not appear to be the case judging from Fig. 1. As in the general effect of vocalic rounding, while there is not a significant main effect of $VOWEL_{y/i}$, there are significant interactions between $VOWEL_{y/i}$ and POSITION both linearly ($\beta = -84.60$, $t = -5.85$, $p < 0.001$) and quadratically ($\beta = -4.53$, $t = -2.29$, $p < 0.05$), suggesting that spectral mean rises less steeply and the curvature of spectral mean is more downward concave before /y/ than before /i/. Interestingly, there is a significant main effect of $VOWEL_{y/o}$ ($\beta = 231.87$, $t = 4.25$, $p < 0.001$), suggesting that the spectral mean is actually higher before /y/ than before /o/. As a lower spectral mean corresponds to a more /ʃ/-like percept, the idea that /s/ is more [ʃ]-like only before /y/ (Bauer and Benedict, 1997) is not evident from the data. Rather, the spectral mean difference before /y/ and /o/ is likely a reflex

of the main vowel height effect since the /y/ ~ /o/ difference in spectral mean (~232 Hz) is comparable to the general vowel height difference (~235 Hz) observed above. Also, like the vowel height effect, the /y/ ~ /o/ effect is not temporally dependent.

2. Standard deviation

The average standard deviation is approximately 2245 Hz at the sibilant onset. Standard deviation increased as the experiment progressed ($\beta = 12.85$, $t = 2.31$, $p < 0.05$). There is a significant effect of sibilant duration; standard deviation decreases as sibilant duration increases ($\beta = -30.53$, $t = -3.77$, $p < 0.001$). Concerning the temporal trends, standard deviation decreases across measurement points ($\beta = -204.71$, $t = -23.98$, $p < 0.001$) and has an upward concave curved trend ($\beta = 18.36$, $t = 21.14$, $p < 0.001$).

The top right panel of Fig. 1 illustrates the vocalic effects on standard deviation. The standard deviation is significantly higher before rounded vowels than before unrounded ones ($\beta = 284.91$, $t = 7.35$, $p < 0.001$) and this rounding effect interacts linearly with POSITION ($\beta = 45.5$, $t = 10.04$, $p < 0.001$), suggesting that the fall in standard deviation before rounded vowels is shallower than before unrounded vowels. The standard deviation is also significantly lower before high vowels than before mid vowels ($\beta = -34.8$, $t = -2.21$, $p < 0.05$). $VOWEL_{High}$ significantly interacts with POSITION both linearly ($\beta = -20.6$, $t = -4.15$, $p < 0.001$) and quadratically ($\beta = 3.75$, $t = 5.53$, $p < 0.001$),

TABLE II. Main and interaction effects of all predictors are shown for significant effects ($p < 0.05$). POSITION² represents the quadratic term of POSITION.

	Mean	Standard deviation	Skewness	Kurtosis	Peak	AmpRatio	Duration
INTERCEPT	7612.86	2245.02	-0.20	1.00	7332.56	42.32	179.86
DURATION	—	-30.53	—	0.03	—	—	NA
TRIAL	37.12	12.85	-0.03	—	51.91	0.20	-1.58
POSITION	306.13	-204.71	-0.05	0.04	384.43	4.46	NA
POSITION ²	-26.73	18.36	—	—	-32.46	-0.43	NA
$VOWEL_{Round}$	—	284.91	-0.35	—	463.23	-2.46	9
$VOWEL_{Round}$: POSITION	-64.24	45.50	-0.07	0.04	-183.27	0.80	NA
$VOWEL_{Round}$: POSITION ²	-6.15	—	0.01	-0.01	—	-0.09	NA
$VOWEL_{High}$	235.49	-34.80	-0.15	0.04	403.33	1.20	30.93
$VOWEL_{High}$: POSITION	—	-20.60	—	0.04	—	-1.19	NA
$VOWEL_{High}$: POSITION ²	—	3.75	—	-0.005	—	0.07	NA
$VOWEL_{y/i}$	—	335.10	-0.43	—	333.58	-1.26	—
$VOWEL_{y/i}$: POSITION	-84.60	35.35	-0.05	0.03	-133.05	—	NA
$VOWEL_{y/i}$: POSITION ²	-4.53	—	0.01	-0.01	—	-0.05	NA
$VOWEL_{y/o}$	231.87	—	-0.24	0.08	364.89	2.90	32.14
$VOWEL_{y/o}$: POSITION	—	-31.13	—	0.04	—	-2.05	NA
$VOWEL_{y/o}$: POSITION ²	—	5.13	—	0.004	—	0.13	NA

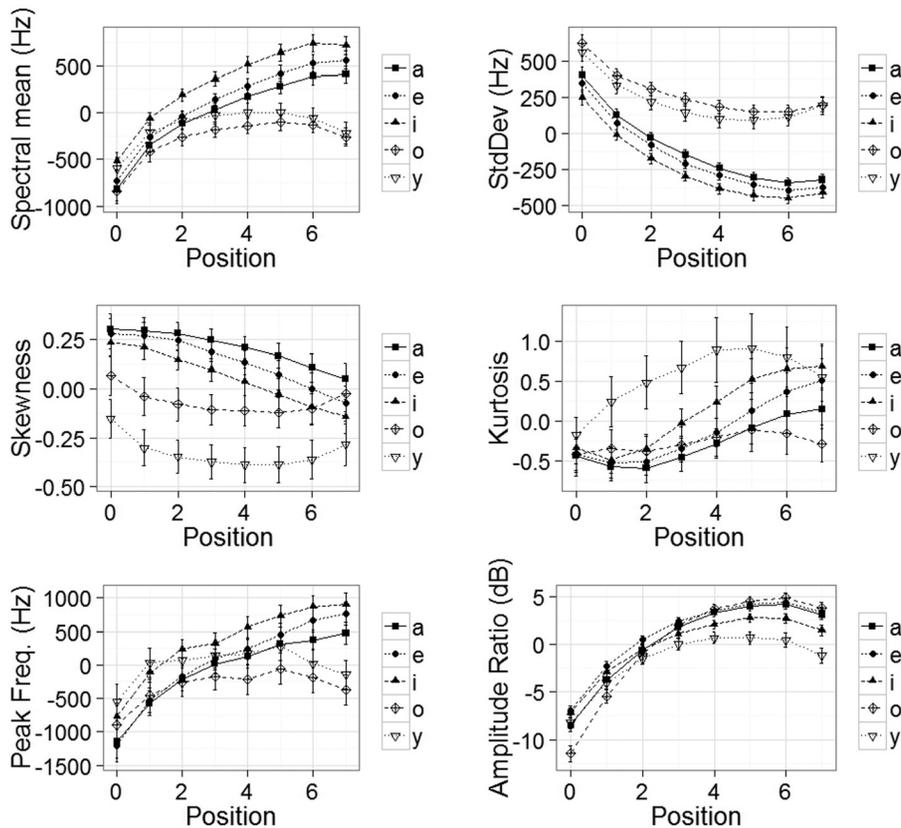


FIG. 1. Six spectral parameters (peak frequency, spectral mean, standard deviation, skewness, kurtosis, and ampRatio) in different vocalic contexts plotted at 1/9 increment of the /s/ interval. Position 0 is at 11.11% of the sibilant while position 7 is the 88.88%. All values are subject-normalized by centering relative to each subject's mean value during all /s/ tokens.

suggesting that standard deviation decreases more before high vowels from the onset to the offset and the curved trend is more downward concave before high vowels than before mid ones.

There is a significant main effect of $VOWEL_{y/i}$ ($\beta = 335.1$, $t = 7.68$, $p < 0.001$) and its interaction with the linear effect of $POSITION$ ($\beta = 35.35$, $t = 5.05$, $p < 0.001$). On the other hand, while the main effect of $VOWEL_{y/o}$ is not significant, its interactions with $POSITION$, both linearly ($\beta = -31.13$, $t = -4.42$, $t = -4.32$, $p < 0.001$) and quadratically ($\beta = 5.13$, $t = 5.33$, $p < 0.001$), are significant, suggesting that standard deviation has a steeper negative slope and a more upward concave curvature from sibilant onset to offset before /y/ than before /o/.

3. Skewness

The regression model shows that there is a gradual decrease in skewness (i.e., a gradual increase in energy) in the higher frequencies over the course of the experiment ($\beta = -0.03$, $t = -3.29$, $p < 0.001$) and across the frication interval ($\beta = -0.05$, $t = -3.25$, $p < 0.001$). There is a significant effect of $VOWEL_{Round}$ ($\beta = -0.35$, $t = -5.71$, $p < 0.001$), suggesting that skewness is more negative before rounded vowels than before unrounded ones. The negative slope of $POSITION$ is more negative in the round context than in the unrounded context ($\beta = -0.07$, $t = -8.98$, $p < 0.001$), suggesting a diverging pattern. While there is not a significant curved trend overall, skewness concaves upward before rounded vowels but concaves downward before unrounded ones ($\beta = 0.01$, $t = 13.69$, $p < 0.001$). The combined linear and quadratic effects of $POSITION$ in different rounding

contexts resulted in a pattern where there were diverging skewness trends across rounding contexts early on in the sibilant interval, but then a slight convergence of the two curved trends obtains toward the offset of the sibilant. Similar patterns are observed for the /y/ vs /i/ contrast. Skewness is more negative before /y/ than before /i/ in general ($\beta = -0.43$, $t = -6.21$, $p < 0.001$) and the slope of $POSITION$ more negative before /y/ than before /i/ ($\beta = -0.05$, $t = -4.32$, $p < 0.001$). Finally, skewness has an upward concave trend before /y/ and a downward concave trend before /i/ ($\beta = 0.01$, $t = 7.74$, $p < 0.001$).

Skewness also varies before vowels of different height; the frication noise has a more negatively skewed distribution before high vowels than mid ones ($\beta = -0.15$, $t = -5.28$, $p < 0.001$). This effect is replicated in the $VOWEL_{y/o}$ contrast as well ($\beta = 0.24$, $t = -5.62$, $p < 0.001$).

4. Kurtosis

As noted above, the kurtosis measure was log-transformed [i.e., $\log(x + 3)$, where x is the actual kurtosis value] before being subjected to modeling to avoid the heavily skewed distribution of this parameter. The average kurtosis at the sibilant onset is 0.14 [i.e., $\exp(1 - 3)$]. The longer the frication noise, the higher the kurtosis value ($\beta = 0.03$, $t = 3.25$, $p < 0.001$). There is a significant positive linear trend of $POSITION$ ($\beta = 0.04$, $t = 5.81$, $p < 0.001$), indicating that the spectrum gets increasingly peaky from the sibilant onset to its offset.

The middle right panel of Fig. 1 shows that the kurtosis measure varies depending on the vocalic context. In particular, before rounded vowels, kurtosis gets more positive

across measurement positions ($\beta = 0.04$, $t = 7.05$, $p < 0.001$) and the curved trend is more downward concave ($\beta = -0.01$, $t = -12.31$, $p < 0.001$). Likewise, before /y/, there is a more positive linear trend ($\beta = 0.03$, $t = 3.93$, $p < 0.001$) and a more downward concave curved trend ($\beta = -0.01$, $t = -7.29$, $p < 0.001$) than before /i/.

Kurtosis is also affected by vowel height; it is higher before high vowels than before mid vowels ($\beta = 0.04$, $t = 2.27$, $p < 0.05$) and this effect of vowel height gets stronger the closer it is to the following vowel ($\beta = 0.04$, $t = 6.83$, $p < 0.001$). Kurtosis also exhibits a more downward concave trend before high vowels than before mid vowels ($\beta = -0.01$, $t = -5.67$, $p < 0.001$). Kurtosis is higher in general ($\beta = 0.08$, $t = 2.81$, $p < 0.001$), the linear trend is more positive ($\beta = 0.04$, $t = 5.24$, $p < 0.001$), and the curved trend more downward concave ($\beta = -0.004$, $t = -3.88$, $p < 0.001$) before /y/ than before /o/.

5. Peak frequency

The Cantonese /s/ has an average peak frequency of 7333 Hz at the sibilant onset. There is a significant effect of TRIAL ($\beta = 51.91$, $t = 25.52$, $p < 0.001$), indicating that peak frequency increased as the experiment progressed. Peak frequency increases from the sibilant onset to the offset ($\beta = 384.43$, $t = 46.76$, $p < 0.001$) and has a significant negative quadratic (i.e., downward concave) trajectory ($\beta = -32.46$, $t = -6.86$, $p < 0.001$).

At the sibilant onset, peak frequency is higher before rounded vowels than before unrounded ones ($\beta = 463.23$, $t = 3.31$, $p < 0.001$). However, peak frequency has a steeper positive slope before unrounded vowels than before rounded ones ($\beta = -183.27$, $t = -6.65$, $p < 0.001$). Likewise, while peak frequency is higher before /y/ than before /i/ at the sibilant onset ($\beta = 333.58$, $t = 2.07$, $p < 0.05$), this difference reverses direction across measurement positions such that peak frequency is higher before /i/ than before /y/ at the sibilant offset ($\beta = -133.05$, $t = -3.14$, $p < 0.001$).

The peak frequency is 403 Hz higher before high vowels than mid vowels ($\beta = 403.33$, $t = 4.66$, $p < 0.001$). This difference is not temporally dependent. Peak frequency also differs significantly before /y/ and before /o/ ($\beta = 364.89$, $t = 2.59$, $p = 0.01$).

6. Amplitude ratio

The average amplitude ratio (ampRatio), i.e., the difference in amplitude between the F2 amplitude and the amplitude of the most prominent peak above the F2 region, is approximately 42 dB at the sibilant onset. AmpRatio increases across trials ($\beta = 0.20$, $t = 3.07$, $p < 0.001$) and across measurement positions ($\beta = 4.46$, $t = 38.71$, $p < 0.001$). It also exhibits a significant downward concave curvature ($\beta = -0.43$, $t = -35.05$, $p < 0.001$).

AmpRatio is significantly higher before high vowels ($\beta = 1.20$, $t = 4.53$, $p < 0.001$). The positive linear trend of POSITION is weaker ($\beta = -1.19$, $t = -14.88$, $p < 0.001$) and the curved trend is more upward concave ($\beta = 0.07$, $t = 6.73$, $p < 0.001$) before high vowels than before mid vowels. This height effect is replicated when comparing the effects of /y/

and /o/; ampRatio is significantly higher before /y/ than before /o/ at the sibilant onset ($\beta = 2.90$, $t = 7.08$, $p < 0.001$), but this difference reverses in direction across measurement positions such that, before /y/, ampRatio is lower ($\beta = -2.05$, $t = -18.15$, $p < 0.001$) and has a more upward concave trend ($\beta = 0.13$, $t = 8.55$, $p < 0.001$) than before /o/.

Relative to the unrounded vowel context, ampRatio is significantly lowered ($\beta = -2.46$, $t = -8.93$, $p < 0.001$) and the positive linear slope is more positive ($\beta = 0.79$, $t = 10.96$, $p < 0.001$) before rounded vowels. The concave curved trend of ampRatio is more downward concave before rounded vowels than before unrounded vowels ($\beta = -0.09$, $t = -9.05$, $p < 0.001$). Similar observations are found in the /y/ and /i/ comparison; ampRatio is lower before /y/ than before /i/ ($\beta = -1.26$, $t = -3.22$, $p < 0.001$); it also has a more downward concave curved trend ($\beta = -0.05$, $t = -3.13$, $p < 0.001$) before /y/.

7. Noise duration

The average noise duration is 180 ms. Duration decreased slightly as the task progressed, at approximately 2 ms per nine trials ($\beta = -1.58$, $t = -2.63$, $p = 0.01$). The participants might have sped up gradually due to an increased familiarity with the task and the carrier phrase. This duration reduction effect across trials might have reduced the extent of lip rounding, which in turn helps explain the significant effects of TRIAL on the majority of the acoustic measures discussed above.

Figure 2 illustrates the various vocalic effects on sibilant duration. The sibilant is significantly longer before high vowels than mid ones ($\beta = 30.93$, $t = 31.47$, $p < 0.001$). This difference is replicated between the /y/ and /o/ contexts ($\beta = 32.14$, $t = 21.18$, $p < 0.001$). This duration difference based on the height of the context vowel might be related to the tendency for high vowels to be shorter; speakers might be compensating for the shortness of the neighboring vowels to maintain duration equivalence across target syllables. There is also a 9 ms difference between round and unrounded vowels ($\beta = 9$, $t = 8.33$, $p < 0.001$), but this difference is likely due to the sibilant length before the /a/ context.

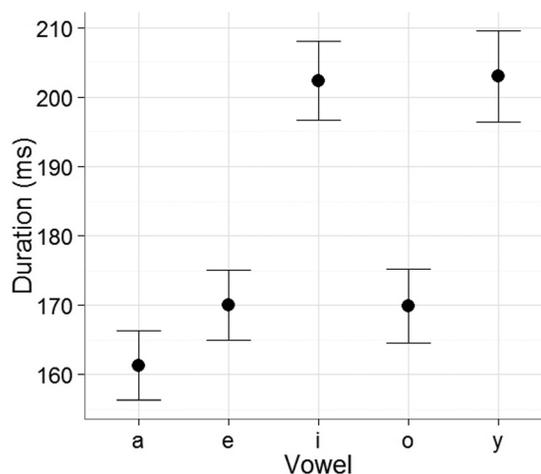


FIG. 2. Sibilant duration in different vocalic contexts.

F. Discussion

The investigation above revealed that the Cantonese sibilant is acoustically more in line with the acoustics of /s/ in other languages than that of other previously proposed transcriptions, such as /ʃ/ (Wang, 1937) or /ç/ (Chao, 1947). Consider, for example, the spectral mean. The Cantonese /s/ has a spectral mean of approximately 8200 Hz, ranging from a high of 8539 Hz (averaged across measurement points; see Table I) before /i/ to a low of 7924 Hz before /o/. In contrast, Li (2012) reported that English /s/ has a spectral mean above 6000 Hz, while /ʃ/ has a spectral mean below 6000 Hz. She also found that Japanese /s/ has a mean around 7500 Hz, while /ʃ/ in Japanese has a spectral mean around 6000 Hz. Likewise, the centroid norms (in Bark) for post-alveolar fricatives in Mandarin is 16.80 for the retroflex /ʃ/ and the 19.12 for /ç/ (Chang *et al.*, 2011) and is 17.79 (4229 Hz) for the English /ʃ/ and /ʒ/ and 19.71 (or 6133 Hz) for /s/ and /z/ (calculated based on averages reported in Jongman *et al.*, 2000), compared to 21.41 (8200 Hz) for the Cantonese /s/.

The investigation above also documented significant vocalic influences on the realization of Cantonese /s/. Both vowel height and vocalic rounding exert significant effects on the acoustic realization of the preceding sibilant. The Cantonese /s/ does not appear to approximate closely the spectral properties of /ʃ/ in English and Japanese or that of /ç/ in Mandarin, even in the rounded vowel contexts. To illustrate this, Fig. 3 provides sample spectra of sibilants

(averaged across the entire frication interval for each sibilant) from Cantonese, English, and Mandarin.¹ The top two panels of Fig. 3 shows the spectra of /s/ in the contexts of /a/ (in black) and /y/ (in grey) from two speakers of Cantonese, one female (a) and one male (b). Figure 3(c) shows the spectra of /s/ (black) and /ʃ/ (grey) produced by a female speaker of American English. Figure 3(d) shows the spectra of [s] (solid black), [ʃ] (solid grey), and [ç] (dotted black) produced by a female Beijing Mandarin speaker. The spectral profile of the Cantonese /s/, which has the greatest noise from around 4000 Hz to 10 000 Hz, most closely approximates that of Mandarin /s/ [solid grey line in Fig. 3(d)]. Cantonese [s] before /y/ shows a spectral peak at a lower frequency [around 3500 Hz in Fig. 3(a) and around 4000 Hz in Fig. 3(b)] than /s/ before /a/ [around 5200 Hz in Fig. 3(a) and around 4400 Hz in Fig. 3(b)]. However, unlike English /ʃ/ [grey line in Fig. 3(c)], which shows an intensity decrease in the higher frequencies (above 8000 Hz), spectral intensity remains high even in higher frequencies in Cantonese /s/ before /y/ [grey lines in Figs. 3(a) and 3(b)], as well as in other contexts for that matter. The Cantonese /s/ before /y/ is similar to the Mandarin apical post-alveolar (as explained in Lee and Zee, 2003, this particular Beijing Mandarin speaker does not have a retroflex sibilant), where higher frequencies maintain a relatively stable high intensity. However, the Cantonese /s/ before /y/, which has one peak centered around 3500 Hz and another at around 7000 Hz, does not have frequency peaks in the same frequencies as the Mandarin apical

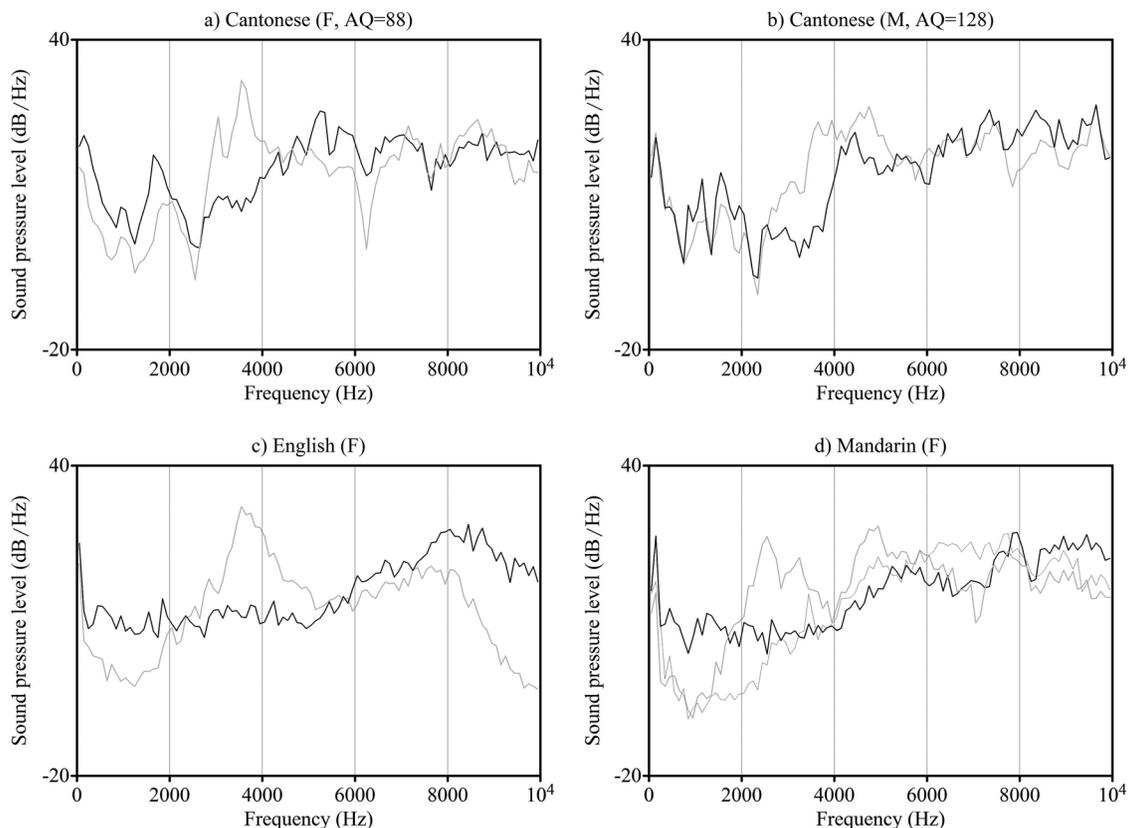


FIG. 3. Long-term average spectra of sibilants across three languages: /s/ in [sa] “sand” (black) and [sy] “book” (grey) produced by (a) a female Cantonese speaker with an AQ of 88 and (b) a male Cantonese speaker with an AQ of 128; (c) [s] (black) and [ʃ] (grey) in “sigh” and “shy,” respectively, produced by a female American English speaker; (d) [s] (solid black line), [ʃ] (solid grey line), and [ç] (dotted black line) in [sa] “to cast,” [ʃa] “sand,” and [çia] “shrimp,” respectively, produced by a female Beijing Mandarin speaker.

post-alveolar, which has one peak at around 2500 Hz and another at around 4800 Hz. The Cantonese /s/ before /y/ also does not resemble the Mandarin [ç], which has a shallower rising intensity profile starting from 920 Hz up to the frequency peak at around 7000 Hz.

Recall also that a larger amplitude ratio has been claimed to indicate a more palatal tongue posture (Li *et al.*, 2007). Our data, however, suggest that ampRatio is higher when the following vowel is back and is lower when the following vowel is rounded. Such findings are not consistent with the idea of /s/ being more [ç]-like before front or rounded vowels, contra the description of Chao (1947) and Bauer and Benedict (1997).

The fact that the vocalic effects on /s/, particularly the effects of vowel height and rounding, interact with POSITION suggests that the influence of vowel quality is not uniform throughout the sibilant interval. However, this dynamic tendency is not consistently observed across spectral parameters. Generally speaking, vocalic influence gets stronger toward the following vowel with respect to spectral peak and mean, as well as ampRatio. Indeed, the main effects of rounding on spectral mean and kurtosis are not significant, only its interactions with the linear and quadratic terms of POSITION are. On the other hand, the main effects of vowel height on spectral mean, spectral peak, and skewness are significant but not its interaction with POSITION. These observations are important as they show that vocalic influences are not necessarily an all-or-nothing phenomenon. Different aspects of the following vowels lead to different coarticulatory trends; the vowel height effects seem to be less time-dependent than the rounding effects. While it remains unclear why the different acoustic parameters do not behave uniformly with respect to vocalic influences (a topic to be explored further in Sec. III), the findings nonetheless suggest that a wholesale allophonic shift from /s/ to [ʃ] in particular, vocalic contexts, as

previous scholars have suggested, is too simplistic. The gradient nature of the vocalic effects, both qualitatively and temporally, and the findings of Lee and Zee (2010) that the lingual articulation of Cantonese /s/ does not differ significantly across vowel contexts strongly support the conclusion that variability in the acoustic realization of /s/ is coarticulatory in nature. Studies that do not examine the spectral temporal dynamic during the frication interval in its entirety would miss this important aspect of Cantonese sibilant behavior.

Another noteworthy aspect of the investigation thus far is the fact that the inclusion of by-subject random intercepts, and particularly the by-subject random slopes for VOWEL, significantly improves model likelihood. This suggests that there exists a significant amount of inter-subject variability in the way the vocalic context affects the acoustic realization of /s/ in Cantonese. As noted in the Introduction, individual variability in sibilant realization has been discussed in previous literature (Newman, 1997; Newman *et al.*, 2001). What is novel here is the finding that individual variation extends to the degree of vocalic influence on sibilant realization. By way of illustration, consider the model predictions (conditional modes) of the by-subject random effects for spectral mean. Figure 4 displays the model predictions for the 105 subjects, sorted by the size of the average spectral mean. While the actual regression model includes by-subject random slopes for the linear and quadratic terms for POSITION, as well as a four-way contrast of VOWEL, only the by-subject random intercepts and the by-subject random slopes for VOWEL_{Round} and VOWEL_{High} are shown here for simplicity-sake. The figure illustrates clearly the extent of the individual differences in the overall spectral mean and the effects of VOWEL_{Round} and VOWEL_{High}, as evidenced by the fact that many participants' prediction intervals are on the opposite sides of the zero line (the zero line represents the corresponding fixed-effect estimate). Similar observations can be found with the other

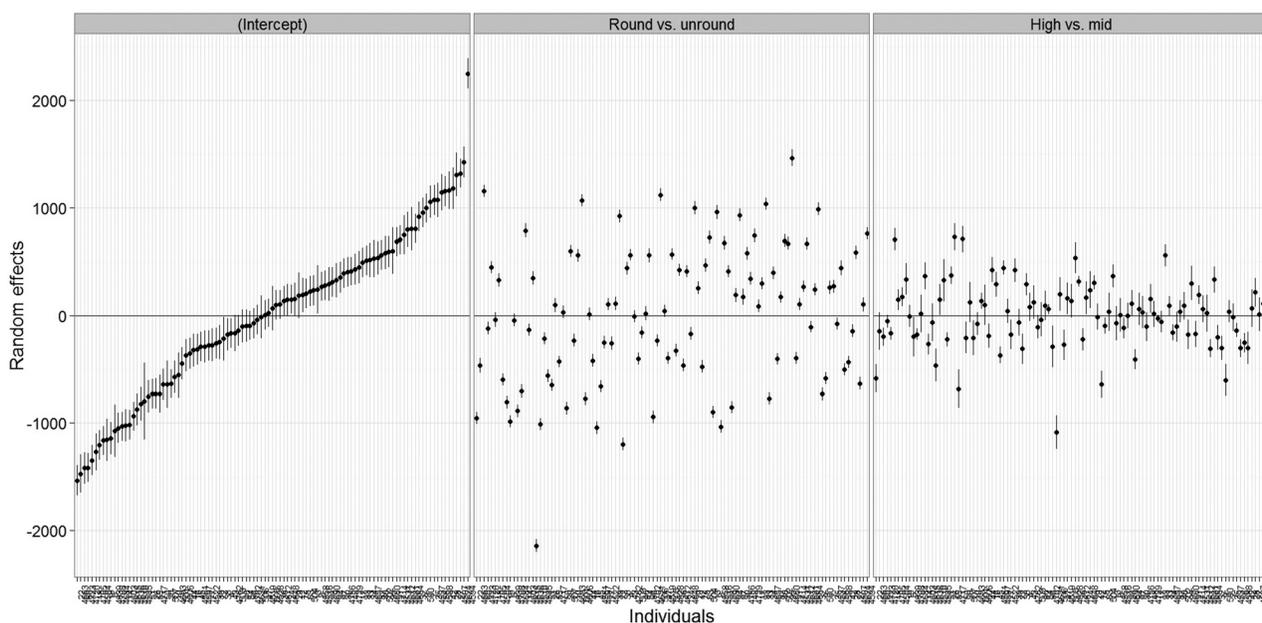


FIG. 4. “Caterpillar plots” for the conditional modes of the by-subject random intercept and the by-subject random slopes for VOWEL_{Round} and VOWEL_{High} for spectral mean across 105 participants. The participants are ordered by the conditional modes of the by-subject intercepts.

acoustic parameters. What accounts for the inter-individual variability in the vocalic effects on sibilant realization? This question is explored in detail in Sec. III.

III. INDIVIDUAL-DIFFERENCE ANALYSIS

This section explores the nature of inter-individual variability in the vocalic effects on /s/ realization in Cantonese by focusing on two individual-difference dimensions: the participant's sex, and self-reported autistic-like traits.

To begin with, the realization of /s/ is known to vary across men and women in general (Nittrouer, 1995; Stuart-Smith, 2007). Specifically, /s/ is often more /ʃ/-like in men than in women. While this difference can be partly explained by physiological differences between male and female vocal tracts, there is evidence for the role of learned sociolinguistic differentiation in the varied realization of /s/ (Stuart-Smith, 2007). In the context of Cantonese, Hashimoto (1972), who acknowledges potential stylistic variation in /s/ realization, notes that male speakers tend to palatalize more than female speakers (p. 120). This section explores how the sex of the participant might modulate the way the quality of the following vowel influences the realization of /s/ in Cantonese.

Another individual-difference dimension explored here is self-reported autistic-like traits. As noted in the Introduction, individual variability in perceptual compensation for coarticulation is linked to variability in cognitive processing style, as indexed by the AQ (Baron-Cohen *et al.*, 2001b). For example, low AQ English-speaking females are found to be less likely to perceptually compensate for vowel-dependent effects on sibilant (Yu, 2010; see also Turnbull, 2015). In terms of speech production, Turnbull (2015) recently reported that individual theory of mind ability, an index of autistic-like traits, predicts the extent of phonetic reduction in speech; individuals with poorer theory of mind ability (i.e., more autistic-like) are more likely to exhibit predictability-based phonetic reduction. These, and other studies reporting similar effects of autistic-like traits mediating the processing of speech and language within neurotypical populations (Stewart and Ota, 2008; Yu, 2010; Nieuwland *et al.*, 2010; Yu *et al.*, 2011; Xiang *et al.*, 2013; Jun and Bishop, 2015), suggest an important link between individual variability in speech and language perception and production on the one hand and fundamental differences cognitive processing styles as indexed by the AQ on the other. Thus, to further explore the effects of sub-threshold autistic traits on speech production, the effect of AQ on the nature of vocalic influence on /s/ is considered.

A. Predictions

To the extent that the magnitude of coarticulation mirrors the magnitude of perceptual compensation, as predicted by gestural accounts of perceptual compensation (Fowler, 2006) that assume the objects of perception and production are one and the same (i.e., phonetic gestures of the vocal tract), high AQ females and males in general are expected to exhibit stronger coarticulation since they have been shown to exhibit stronger perceptual compensation. However, if perceptual compensation involves expectation adjustments

and if production targets are calculated at this expectation-adjusted level, then high AQ females and males in general are expected to exhibit weaker coarticulation. Finally, if coarticulation is for the benefit of the listeners (Pycha, 2015; Scarborough, 2004, 2013; Wright, 2004), individuals with more autistic-like traits, and hence poorer theory-of-mind abilities, should exhibit less coarticulation in speech given their poorer ability to assign complex mental states to others, including the interlocutor's likelihood of comprehending the message produced (cf. Turnbull, 2015).

B. Principal component analysis

Rather than examining the correlation between individual-difference dimensions and individual acoustic cue one by one, the dimensionality of the acoustic space was reduced using principal component analysis (PCA) to obtain linear combinations of weights that would capture the maximum amount of variation in the data. This approach should also provide additional clarity in terms of the nature of the vocalic influence and minimizing the disparate results observed above when examining the acoustic parameters individually, since closely related subsets of acoustic parameters are analyzed together.

Before submitting to the PCA, spectral peak frequency, spectral mean, and spectral standard deviation, which are all in Hz, amplitude ratio (in dB), and sibilant duration (in ms) were log-transformed (natural log). The kurtosis variable, while unitless, was nonetheless log-transformed as it was extremely positively skewed, as noted above. (Recall also that, since the kurtosis can be negative, the value 3 was added to all kurtosis values to ensure that they are positive prior to log-transformation.) The spectral measures, averaged across the eight measurement points, and the duration measurement were analyzed using the `prcomp()` function in R. 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 III. The optimal linear combination (PC1) accounts for about 47% of the variance, the second component accounted for 21% of the variance, while the third accounted for 14%. PC1 has strong loadings for peak frequency, spectral mean, kurtosis, and skewness, which are spectral measures that characterize the way spectral energy is concentrated. PC2, on the other hand, is dominated by standard deviation and `ampRatio`, which pertain to the relative energy levels across frequency ranges of the spectrum.

Given that PC1 and PC2 both have eigenvalues above 1 and they collectively account for about 70% of the variance, these two linear combinations were taken as the independent variables for further analysis. The results of the linear transformed components, obtained using the `predict()` function in R, shall be referred to from here on as sibilant index 1 (SI 1) and sibilant index 2 (SI 2), respectively.

The two SIs were modeled separately using linear mixed-effects regression. Both indexes were tested for the effects of VOWEL and the linear and quadratic effects of POSITION, as well as the interaction between the two predictors. Also

TABLE III. The cumulative proportion of variance accounted for and loadings from the PCA of sibilant from spectral measures. All spectral measures were z-scored.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Variance (eigenvalue)	3.282	1.440	0.963	0.731	0.371	0.127	0.084
Proportion of variance	0.469	0.206	0.138	0.105	0.053	0.018	0.012
Cumulative proportion	0.469	0.675	0.812	0.917	0.970	0.988	1.000
Peak frequency	0.473	-0.303	0.103	-0.153	-0.309	0.733	0.130
Spectral mean	0.481	-0.268	0.125	-0.215	-0.287	-0.508	-0.543
Standard deviation	-0.348	-0.551	-0.051	-0.245	0.504	0.244	-0.445
Skewness	-0.440	0.387	0.001	-0.110	-0.536	0.288	-0.524
log(kurtosis + 3)	0.404	0.325	-0.080	0.560	0.372	0.241	-0.464
AmpRatio	0.214	0.523	0.216	-0.700	0.375	0.060	0.003
Duration	0.146	0.053	-0.958	-0.236	-0.045	-0.007	0.006

included is sibilant DURATION, which serves as a proxy for speaking rate. The VOWEL variable was contrast-coded the same way as described above. The model also includes by-subject random intercepts as well as by-subject random slopes for all main task factors. The model also includes by-item random intercepts. The effects of SEX and AQ were tested by comparing a model without any individual-level factor with a model with an individual-level factor and its interaction with the task predictors using likelihood ratio tests. SEX was sum-coded (female = 1; male = -1) while AQ was entered as a centered and z-scored continuous variable.

As noted above, the AQ (Baron-Cohen *et al.*, 2001b) is a short, self-administered scale for identifying the degree to which any individual adult of normal IQ may have traits associated with ASD. The test consists of 50 items, made up of 10 questions assessing five subscales: social skills, communication, attention to detail, attention-switching, and imagination. The AQ items were scored on a Likert scale (1-4), following Stewart and Ota (2008) and Yu (2010), as this method retains more information about the participants' responses and also increases item-item correlations, scale reliability, and validity coefficients (Austin, 2005; Muñiz *et al.*, 2005). A total AQ score was calculated by summing all the scores for each of the items, with a maximum score of 200 and a minimum score of 50. The AQ scale was scored in such a way that a higher score is more autistic-like, i.e., lower social skills, difficulty in attention switching/strong focus of attention, high attention to detail and patterns, lower ability to communicate, and low imagination. Participants were given the option of filling out the Chinese or English version of the AQ. The cross-cultural psychometric validity of the AQ has been examined, including in Chinese (Lau *et al.*, 2013). The mean and median AQ score of the present cohort of 105 participants are 114.9 and 115, respectively (SD = 11.55, range = 88-149), compared to a mean of 102 (N = 55, SD = 14.5, range = 71-150) in Stewart and Ota (2008) and 110.05 (N = 60, SD = 18, range = 78-155) in Yu (2010). Hoekstra *et al.* (2008), in a study of the AQ in Dutch population and patient groups, reported that ASD patients have AQ higher than 145. Only one participant (AQ = 149) in the present cohort has an AQ higher than 145.

The inclusion of SEX and AQ both significantly and independently improves model likelihood using a likelihood ratio test. The model formula in lme4 style was SI (1 or 2)

$\sim (\text{SEX} + \text{AQ}) * (\text{DURATION} + \text{VOWEL} * (\text{POSITION} + \text{POSITION}^2)) + (1 + \text{DURATION} + \text{VOWEL} + \text{POSITION} + \text{POSITION}^2) | \text{SUBJECT} + (1 | \text{Word})$. A summary of the models for SI1 and SI2 are presented in Table IV.

Recall that Yu (2010) reported that, for the American English speakers he tested, there is an interaction between SEX and AQ such that it is the women with lower AQ that are not perceptually compensating for sibilant-vowel coarticulation. To the extent that behavior in perception finds analog in production, and to the extent that findings with respect to American English speakers have relevance for native speakers of Cantonese, one might expect similar interaction between SEX and AQ. Thus, in order to further examine potential interactions between SEX and AQ, rather than testing for potential four-way interactions, separate regression models were also tested for male and female participants. The model formula tested were SI (1 or 2) $\sim \text{AQ} * (\text{DURATION} + \text{VOWEL} * (\text{POSITION} + \text{POSITION}^2)) + (1 + \text{DURATION} + \text{VOWEL} + \text{POSITION} + \text{POSITION}^2) | \text{SUBJECT} + (1 | \text{Word})$. A summary of these sex-specific models are presented in Table V.

C. SI 1

The average onset SI 1 is -2.42. Recall that SI1 has strong loadings for peak frequency, spectral mean, kurtosis, and skewness. An increase in SI1 corresponds to an increase in peak frequency, spectral mean, and kurtosis, and a lowering of skewness. The longer the sibilant, the higher the SI1 ($\beta = 0.22$, $t = 6.24$, $p < 0.001$). SI1 increases across measurement points ($\beta = 0.74$, $t = 17.35$, $p < 0.001$); it also has a downward concave trend ($\beta = -0.05$, $t = -14.78$, $p < 0.001$). SI1 is higher in the high vowel context than in the mid vowel context ($\beta = 0.29$, $t = 3.04$, $p < 0.001$). The curved trend is more downward concave ($\beta = -0.01$, $t = -2.2$, $p < 0.05$) before high vowels than before mid ones. While there is not a main effect of rounding, SI1 has a more positive linear slope and more downward concave trend before rounded vowels than before unrounded ones. There is also a significant $\text{VOWEL}_{y/i} \times \text{POSITION}^2$ interaction, suggesting that the curved trend is more downward concave before /y/ and before /i/.

The top panel of Fig. 5 illustrates the overall effects of participants' sex on SI1 as a function of vocalic context and measurement positions.² To begin with, there is a significant difference between males and females in terms of SI1;

TABLE IV. Summary of regression models for SI1 and SI2.

Predictor	SI1	SI2
	Coef(SE)p	Coef(SE)p
Intercept	-2.415(0.160) ^a	-1.839(0.115) ^a
DURATION	0.219(0.035) ^a	0.133(0.023) ^a
VOWEL _{Round}	0.046(0.149)	-0.671(0.080) ^a
VOWEL _{High}	0.289(0.095) ^a	-0.009(0.064)
VOWEL _{yi}	0.219(0.181)	-0.576(0.110) ^a
VOWEL _{yo}	0.516(0.138) ^a	0.128(0.095)
POSITION	0.739(0.043) ^a	0.593(0.032) ^a
POSITION ²	-0.046(0.003) ^a	-0.040(0.002) ^a
VOWEL _{Round} : POSITION	0.073(0.028) ^b	0.035(0.020)
VOWEL _{High} : POSITION	0.038(0.030)	-0.007(0.022)
VOWEL _{yi} : POSITION	0.005(0.043)	-0.003(0.031)
VOWEL _{yo} : POSITION	-0.017(0.043)	-0.048(0.031)
VOWEL _{Round} : POSITION ²	-0.020(0.002) ^a	-0.004(0.002) ^b
VOWEL _{High} : POSITION ²	-0.006(0.003) ^b	-0.003(0.002)
VOWEL _{yi} : POSITION ²	-0.015(0.004) ^a	-0.002(0.003)
VOWEL _{yo} : POSITION ²	-0.002(0.004)	-0.002(0.003)
SEX	-0.451(0.164) ^b	0.173(0.118)
SEX: DURATION	0.025(0.036)	0.043(0.024)
SEX: VOWEL _{Round}	-0.400(0.151) ^b	0.098(0.082)
SEX: VOWEL _{High}	0.086(0.094)	-0.098(0.066)
SEX: VOWEL _{yi}	-0.521(0.182) ^a	0.249(0.112) ^b
SEX: VOWEL _{yo}	-0.020(0.137)	0.022(0.097)
SEX: POSITION	0.184(0.044) ^a	-0.072(0.033) ^b
SEX: POSITION ²	-0.013(0.003) ^a	0.005(0.002) ^b
SEX: VOWEL _{Round} : POSITION	-0.038(0.029)	-0.104(0.021) ^a
SEX: VOWEL _{High} : POSITION	0.001(0.031)	-0.009(0.023)
SEX: VOWEL _{yi} : POSITION	-0.005(0.044)	-0.138(0.032) ^a
SEX: VOWEL _{yo} : POSITION	0.039(0.045)	-0.037(0.032)
SEX: VOWEL _{Round} : POSITION ²	-0.002(0.003)	0.011(0.002) ^a
SEX: VOWEL _{High} : POSITION ²	-0.001(0.003)	0.001(0.002)
SEX: VOWEL _{yi} : POSITION ²	-0.004(0.004)	0.012(0.003) ^a
SEX: VOWEL _{yo} : POSITION ²	-0.004(0.004)	0.003(0.003)
AQ	-0.243(0.160)	0.106(0.115)
AQ: DURATION	0.003(0.036)	-0.006(0.024)
AQ: VOWEL _{Round}	-0.203(0.148)	0.252(0.081) ^a
AQ: VOWEL _{High}	0.109(0.093)	-0.026(0.065)
AQ: VOWEL _{yi}	-0.099(0.179)	0.353(0.110) ^a
AQ: VOWEL _{yo}	0.211(0.135)	0.072(0.096)
AQ: POSITION	0.043(0.043)	-0.023(0.033)
AQ: POSITION ²	-0.003(0.003)	0.001(0.002)
AQ: VOWEL _{Round} : POSITION	0.073(0.028) ^b	-0.082(0.020) ^a
AQ: VOWEL _{High} : POSITION	-0.072(0.031) ^b	0.009(0.022)
AQ: VOWEL _{yi} : POSITION	0.006(0.043)	-0.107(0.031) ^a
AQ: VOWEL _{yo} : POSITION	-0.147(0.044) ^a	-0.020(0.032)
AQ: VOWEL _{Round} : POSITION ²	-0.004(0.003)	0.008(0.002) ^a
AQ: VOWEL _{High} : POSITION ²	0.006(0.003) ^b	0.000(0.002)
AQ: VOWEL _{yi} : POSITION ²	0.000(0.004)	0.008(0.003) ^a
AQ: VOWEL _{yo} : POSITION ²	0.011(0.004) ^a	0.001(0.003)

^a $p < 0.001$.^b $p < 0.05$.

females have lower SI1 than males ($\beta = -0.45$, $t = -2.76$, $p < 0.05$). There is a significant interaction between SEX and VOWEL_{Round}, suggesting that the male participants actually exhibit the opposite rounding effect on SI1 compared to the female participants ($\beta = -0.40$, $t = -2.65$, $p < 0.05$). That is, relative to the unrounded vowel context, while females show a reduction of SI1 before rounded vowels, males show a relative increase of SI1 before rounded vowels. As illustrated by

the top panel of Fig. 5, this reversal effect in males is likely driven by the relatively stable SI1 across measurement points before rounded vowels, but a rising trend before unrounded vowels. This sex difference in rounding effect on SI1 is reflected in the contrast between /y/ and /i/ as well. There are also significant interactions between sex and the linear and quadratic terms of POSITION, indicating that female participants have a stronger positive linear trend ($\beta = 0.184$, $t = 4.22$, $p < 0.001$) and a more downward concave trend ($\beta = -0.013$, $t = -3.93$, $p < 0.001$). It is worth noting that the effect of DURATION was not differentiated by the sex of the participant, suggesting that whatever the differences in vocalic effects are, they are not likely to be the results of sex-based differences in segmental duration. Also noteworthy is the lack of a sex-based difference in the VOWEL \times POSITION interaction, regardless whether POSITION is linear or quadratic.

There is a significant three-way interaction between AQ, VOWEL_{Round}, and POSITION ($\beta = 0.07$, $t = 2.58$, $p < 0.05$); the difference in slope in different rounding contexts is greater in lower AQ individuals than in higher AQ individuals. Lower AQ individuals also show a stronger difference in SI1 before high and mid vowel contexts toward the sibilant offset ($\beta = -0.07$, $t = -2.36$, $p < 0.05$). The difference in concavity before high and mid vowels is less pronounced in higher AQ individuals than in lower AQ individuals ($\beta = 0.01$, $t = 2.07$, $p < 0.05$). The difference in SI1 between /y/ and /o/ contexts across measurement points is also modulated by AQ; individuals with lower AQ show a less positive linear trend ($\beta = -0.15$, $t = -3.35$, $p < 0.001$) and a less downward concave trend ($\beta = 0.01$, $t = 2.90$, $p < 0.001$) before /y/ than before /o/. Taken together, the vocalic influence on SI1, which is temporally dependent, is mainly observed in individuals with lower AQ.

The effect of AQ on SI1 is modulated by the sex of the speaker. For male speakers, there is a significant three-way interaction between AQ, VOWEL_{Round}, and POSITION. Figure 6 illustrates the sex-dependent effects of AQ on SI1 as a function of vocalic contexts and measurement positions. While the AQ was entered in the regression model as a continuous variable, for ease of presentation, Fig. 6 shows only the behaviors of participants in three AQ quartile ranges: top (AQ > 123, approximately 24 on the original non-Likert AQ scale), bottom (AQ \leq 106, approximately 15 on the original AQ scale), and the middle range in between. As shown in the lower panel of Fig. 6, the slope of SI1 across measurement points is more positive ($\beta = 0.14$, $t = 3.64$, $p < 0.001$) and has a stronger downward concavity ($\beta = -0.01$, $t = -2.52$, $p < 0.05$) in the rounded vowel context for the higher AQ speakers than for the lower AQ speakers. On the other hand, higher AQ females (i.e., upper panel of Fig. 6) show a shallower slope ($\beta = -0.09$, $t = -2.14$, $p < 0.05$) and less downward concavity ($\beta = 0.01$, $t = 2.02$, $p < 0.05$) before high vowels than lower AQ females.

D. SI 2

The average SI 2 at the sibilant onset is -1.84. In general, an increase in SI2 corresponds to a decrease in standard deviation and an increase in ampRatio. The longer the sibilant, the higher the SI2 ($\beta = 0.13$, $t = 5.74$, $p < 0.001$). SI2

TABLE V. Summary of sex-specific regression models for SI1 and SI2.

Predictor	SI1		SI2	
	Male N = 43 Coef (SE) p	Female N = 62 Coef (SE) p	Male N = 43 Coef (SE) p	Female N = 62 Coef (SE) p
Intercept	-1.956 (0.257) ^a	-2.877 (0.208) ^a	-1.981 (0.162) ^a	-1.660 (0.156) ^a
DURATION	0.185 (0.045) ^a	0.239 (0.052) ^a	0.090 (0.032) ^a	0.174 (0.034) ^a
VOWEL _{Round}	0.474 (0.224) ^b	-0.392 (0.200)	-0.704 (0.132) ^a	-0.481 (0.106) ^a
VOWEL _{High}	0.179 (0.153)	0.370 (0.130) ^a	0.130 (0.091)	-0.114 (0.093)
VOWEL _{/y/}	0.780 (0.279) ^b	-0.348 (0.244)	-0.762 (0.176) ^a	-0.266 (0.147)
VOWEL _{/o/}	0.514 (0.217) ^b	0.478 (0.191) ^b	0.121 (0.135)	0.111 (0.137)
POSITION	0.542 (0.057) ^a	0.921 (0.061) ^a	0.657 (0.049) ^a	0.522 (0.043) ^a
POSITION ²	-0.033 (0.004) ^a	-0.059 (0.005) ^a	-0.045 (0.004) ^a	-0.035 (0.003) ^a
VOWEL _{Round} : POSITION	0.076 (0.042)	0.036 (0.038)	0.117 (0.029) ^a	-0.101 (0.028) ^a
VOWEL _{High} : POSITION	0.038 (0.046)	0.039 (0.041)	-0.017 (0.032)	-0.014 (0.030)
VOWEL _{/y/} : POSITION	-0.039 (0.065)	0.004 (0.058)	0.115 (0.045) ^b	-0.160 (0.043) ^a
VOWEL _{/o/} : POSITION	-0.067 (0.065)	0.027 (0.059)	-0.022 (0.045)	-0.072 (0.043)
VOWEL _{Round} : POSITION ²	-0.016 (0.004) ^a	-0.022 (0.003) ^a	-0.013 (0.003) ^a	0.009 (0.002) ^a
VOWEL _{High} : POSITION ²	-0.005 (0.004)	-0.007 (0.004)	-0.003 (0.003)	-0.002 (0.003)
VOWEL _{/y/} : POSITION ²	-0.007 (0.006)	-0.020 (0.005) ^a	-0.012 (0.004) ^a	0.012 (0.004) ^a
VOWEL _{/o/} : POSITION ²	0.004 (0.006)	-0.006 (0.005)	-0.003 (0.004)	0.000 (0.004)
AQ	-0.210 (0.240)	-0.267 (0.216)	0.006 (0.152)	0.191 (0.163)
AQ: DURATION	0.042 (0.042)	-0.037 (0.055)	-0.018 (0.030)	0.000 (0.036)
AQ: VOWEL _{Round}	-0.159 (0.202)	-0.272 (0.206)	-0.098 (0.122)	0.644 (0.110) ^a
AQ: VOWEL _{High}	0.029 (0.128)	0.110 (0.132)	-0.037 (0.083)	-0.022 (0.096)
AQ: VOWEL _{/y/}	-0.080 (0.244)	-0.177 (0.251)	-0.006 (0.163)	0.690 (0.151) ^a
AQ: VOWEL _{/o/}	0.128 (0.182)	0.183 (0.195)	0.081 (0.124)	-0.007 (0.141)
AQ: POSITION	0.071 (0.054)	0.020 (0.064)	-0.005 (0.046)	-0.039 (0.045)
AQ: POSITION ²	-0.006 (0.003)	-0.001 (0.005)	0.001 (0.003)	0.002 (0.003)
AQ: VOWEL _{Round} : POSITION	0.143 (0.039) ^a	0.025 (0.040)	0.047 (0.027)	-0.225 (0.029) ^a
AQ: VOWEL _{High} : POSITION	-0.017 (0.043)	-0.093 (0.043) ^b	0.023 (0.030)	0.004 (0.032)
AQ: VOWEL _{/y/} : POSITION	0.102 (0.061)	-0.050 (0.061)	0.017 (0.043)	-0.224 (0.045) ^a
AQ: VOWEL _{/o/} : POSITION	-0.071 (0.062)	-0.171 (0.062) ^b	-0.012 (0.042)	0.002 (0.046)
AQ: VOWEL _{Round} : POSITION ²	-0.009 (0.004) ^b	-0.002 (0.004)	-0.002 (0.002)	0.018 (0.003) ^a
AQ: VOWEL _{High} : POSITION ²	0.001 (0.004)	0.008 (0.004) ^b	-0.002 (0.003)	0.000 (0.003)
AQ: VOWEL _{/y/} : POSITION ²	-0.006 (0.005)	0.004 (0.005)	-0.001 (0.004)	0.017 (0.004) ^a
AQ: VOWEL _{/o/} : POSITION ²	0.004 (0.005)	0.014 (0.006) ^b	0.000 (0.004)	-0.001 (0.004)

^a $p < 0.001$.

^b $p < 0.05$.

increases across measurement points ($\beta = 0.59$, $t = 18.24$, $p < 0.001$); it also has a downward concave trend ($\beta = -0.04$, $t = -16.69$, $p < 0.001$). SI2 is lower generally ($\beta = -0.67$, $t = -8.35$, $p < 0.001$) and its curved trend is more downward concave ($\beta = -0.004$, $t = -2.13$, $p < 0.05$) before rounded vowels than before unrounded ones. There is also a significant VOWEL_{/y/} effect ($\beta = -0.58$, $t = -5.25$, $p < 0.001$), mirroring the general effect of lip rounding.

The VOWEL_{/y/} is mediated by the sex of the participant ($\beta = 0.25$, $t = 2.22$, $p < 0.05$); females show a larger difference in SI2 before /y/ and /i/ than males. The linear and quadratic effects of POSITION is modulated by SEX as well. Females show a less positive slope ($\beta = -0.07$, $t = -2.15$, $p < 0.05$) and a less downward concave trend ($\beta = 0.01$, $t = 2.17$, $p < 0.05$) than males. More intriguing are the sex-based differences in the interaction between vocalic rounding and POSITION. While there is not a significant interaction between VOWEL_{Round} and POSITION, suggesting that the linear rise in SI2 is at the same rate regardless of vocalic contexts,

the male participants nonetheless show a steeper positive slope in the rounded vowel contexts than in the unrounded ones ($\beta = -0.10$, $t = -5.07$, $p < 0.001$), suggesting that the male participants have a larger rounding difference toward the sibilant offset. This sex-based effect is mirrored in the interaction between SEX, VOWEL_{/y/}, and POSITION ($\beta = -0.14$, $t = -4.34$, $p < 0.001$). To be sure, this does not mean that males have larger rounding effects than females as the female participants exhibit a larger rounding influence on SI2 in general. Rather, the SEX \times VOWEL_{/y/} \times POSITION interaction indicates that the rounding effect is more temporally dynamic in males than females. The curved trend of SI2 before rounded vowels are also less downward concave in females than in males ($\beta = 0.01$, $t = 5.83$, $p < 0.001$). A similar effect is observed with respect to the contrast between the /y/ and /i/ contexts ($\beta = 0.01$, $t = 4.29$, $p < 0.001$).

Strong effects of AQ on rounding coarticulation are observed with respect to SI2. To begin with, the main effects of lip rounding are significantly modulated by the

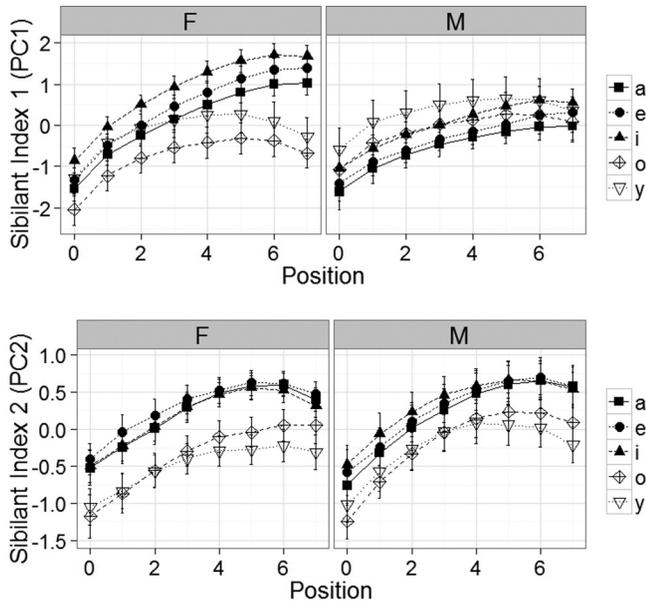


FIG. 5. SIs 1 and 2 by each sex in different vocalic contexts plotted from the onset to the offset of frication noise at 11% increments. The SI values are subject-normalized by centering relative to each subject's mean SI during all /s/ tokens.

participant's AQ; the higher the AQ, the weaker the effects of $VOWEL_{Round}$ ($\beta = 0.25$, $t = 3.11$, $p < 0.001$) and $VOWEL_{y/i}$ ($\beta = 0.35$, $t = 3.20$, $p < 0.001$). AQ also modulates the position-dependent effects of lip rounding. Higher AQ participants show a smaller difference in SI2 before round and unrounded vowels in the sibilant offset than in the onset ($\beta = -0.08$, $t = -4.03$, $p < 0.001$), but this reduction of the rounding effect is likely driven by a weaker SI2 reduction before /o/ than before /y/, even though neither the $VOWEL_{y/o}$ nor its interaction with POSITION is significant.

Like SI1, the AQ effect is also modulated by the sex of the speaker. Figure 7 illustrates the sex-dependent effects of AQ on SI2 as a function of vocalic contexts and

measurement positions. Female speakers show significant interactions between AQ, $VOWEL_{Round}$ (also $VOWEL_{y/i}$), and both the linear and quadratic terms of POSITION. No AQ effect is observed among the male speakers, however. As shown in the upper panel of Fig. 7, lower AQ female speakers show a strong general effect of lip rounding (i.e., the rounding difference pervades the entire sibilant interval), while higher AQ females exhibit a weaker rounding effect toward the onset of the sibilant; the slope of SI2 before rounded vowels is steeper ($\beta = -0.23$, $t = -7.67$, $p < 0.001$) and the curve trend shows more downward concavity ($\beta = 0.02$, $t = 6.94$, $p < 0.001$) among the lower AQ females compared to the higher AQ ones.

IV. DISCUSSION

This study establishes the acoustic properties of Hong Kong Cantonese /s/ in different vocalic contexts. Significant inter-individual variability, particularly concerning the nature of the vocalic influence, was uncovered. SI1 exhibits primarily height-based influence on /s/, while SI2 highlights mainly roundness-based influence on sibilant realization. While the main effects of $VOWEL_{High}$ (in the case of SI1) and $VOWEL_{Round}$ (in the case of SI2) are significant, both exhibit temporal sensitivity, suggesting that, while the vocalic influence on /s/, in the population as a whole, extends across the entire noise interval, the magnitude of the influence can nonetheless vary the closer it is to the vocalic neighbor.

This investigation also establishes that the participant's sex and the degree of self-reported autistic-like traits both contribute significantly toward explaining the variance in the data. As noted earlier, the fact that the participant's sex interacts with the influence of vocalic rounding on the acoustics of /s/ points to the existence of a potential sociolinguistic difference. This difference between the sexes involves an effect of rounding that is temporally dynamic, however, suggesting that it is not a simple categorical alternation between

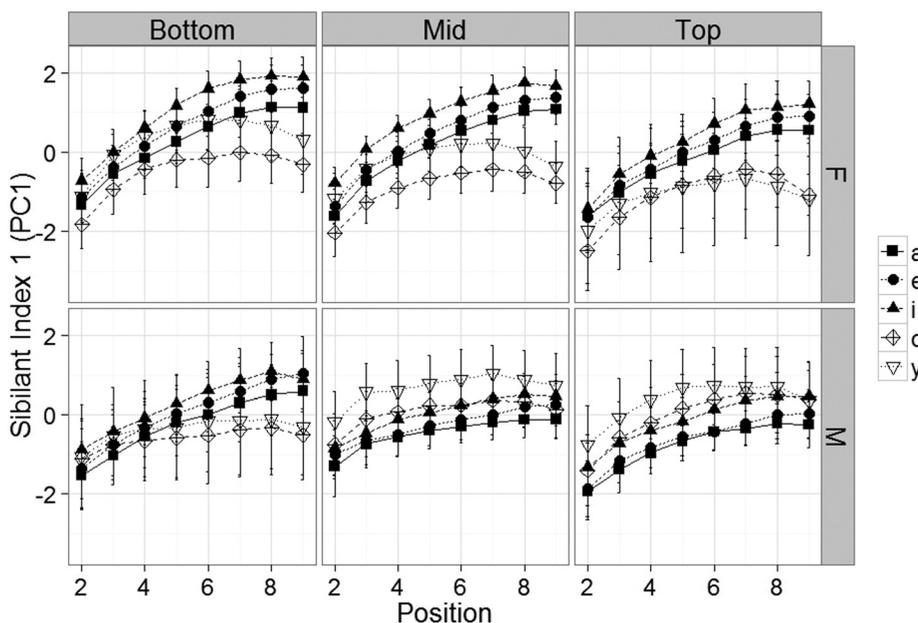


FIG. 6. SIs 1 by sex and AQ quartiles (the mid two quartiles are combined) in different vocalic contexts plotted from the onset to the offset of frication noise at 11% increments. The SI values are subject-normalized by centering relative to each subject's mean SI during all /s/ tokens. The top AQ quartile is defined as $AQ > 123$ while the bottom quartile equals $AQ \leq 106$.

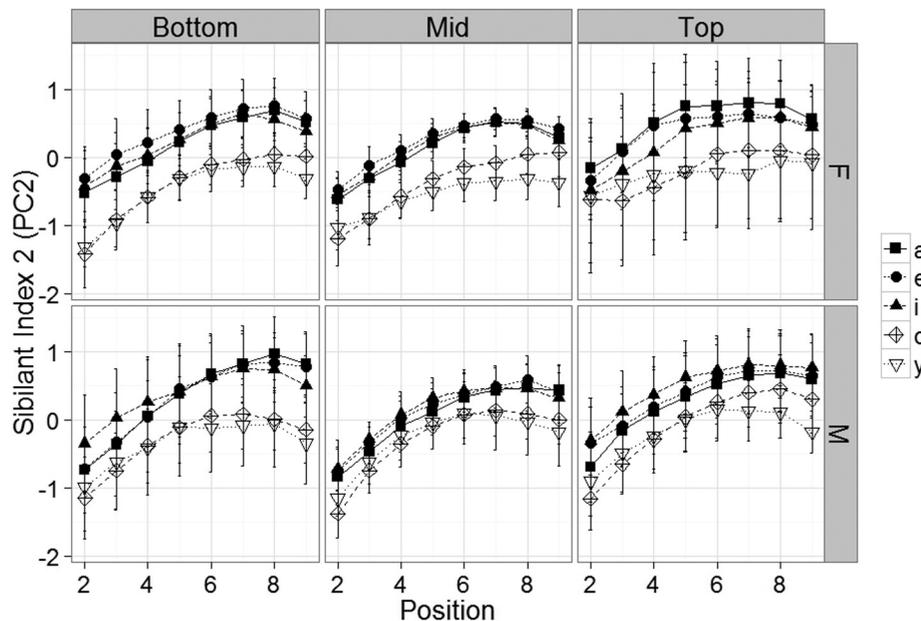


FIG. 7. SIs 2 by sex and AQ quartiles (the mid two quartiles are combined) in different vocalic contexts plotted from the onset to the offset of frication noise at 11% increments. The SI values are subject-normalized by centering relative to each subject's mean SI during all /s/ tokens. The top AQ quartile is defined as $AQ > 123$ while the bottom quartile equals $AQ \leq 106$.

allophones as assumed by previous researchers, but it appears to involve differences in coarticulatory routines.

Furthermore, there also exists inter-individual variability, such that individuals with less autistic-like traits exhibit more vocalic influence on /s/ production than those with more autistic-like traits. As illustrated in the top panel of Fig. 3, /s/ is produced by speakers with different AQ with markedly different spectral profiles in different vocalic contexts. Figure 3(a) shows the spectra of /s/ produced by a speaker with low AQ; the spectrum for /s/ before /y/ has a stronger positive skew (left-leaning tilt) than the spectrum for /s/ before /a/. On the other hand, as shown in Fig. 3(b), a high AQ speaker exhibits very similar spectra for /s/ regardless of vocalic contexts. Further research is needed to ascertain the mechanisms behind the effects of autistic-like traits on vocalic coarticulation on sibilant production. As noted earlier, if coarticulation in production mirrors the magnitude of perceptual compensation, as suggested by a gesturalist approach to speech perception, a positive relationship between AQ and the extent of the vocalic influence would be expected. That is, individuals with higher AQ should exhibit larger vocalic influence in /s/ production than lower AQ individuals since earlier studies found individuals with higher AQ exhibit stronger perceptual compensation for vocalic influence on /s/ than lower AQ individuals. Yet, the opposite pattern is observed here. In particular, lower AQ females, the segment of the American English-speaking population that exhibits the least compensation for coarticulation in Yu (2010), turned out to exhibit the strongest vocalic influence in production in the present study. To be sure, the mismatch between the production of coarticulation in Cantonese and the expected direction of perceptual compensation might reflect a cross-linguistic difference as the AQ effect on perceptual compensation in Yu (2010) was based on an American English-speaking population. Further research is needed to ascertain the language-specific influence on the effect of AQ on speech production.

The present findings are consistent with an expectation-adjustment approach to perceptual compensation for coarticulation where individuals who engage in robust expectation adjustments (i.e., males and high AQ females) are less likely to coarticulate. The present findings are also consistent with the predictions of the listener-oriented approach to coarticulation. That is, to the extent that lower AQ individuals (particularly lower AQ females) can be said to have stronger theory-of-mind abilities, a listener-driven approach to coarticulation predicts that individuals with lower AQ would show stronger vocalic influence on /s/ than higher AQ individuals, assuming that coarticulation can benefit the listeners in predicting contextual information and thus aiding lexical recovery.

To be sure, the AQ difference in vocalic effects on /s/ production does not appear to be related to differences in segmental duration, and by extension, speaking rate; the interaction between AQ and the DURATION does not have a significant effect on either SII or SI2.

V. CONCLUSION

This study elucidated the acoustic properties of Cantonese /s/ in different vocalic contexts, establishing significant inter-individual variability in how the qualities of the neighboring vowel influence the acoustic realization of /s/. Given earlier findings of stable lingual articulation of /s/ in different vocalic contexts and the gradient and temporally dynamic nature of the vocalic influence, a coarticulatory interpretation of the vocalic effects, rather than a strictly categorical, all-or-nothing, allophonic interpretation, seems warranted. While this study is limited by the number of productions analyzed per subject, the large number of participants analyzed allows us to examine the nature of the interspeaker variation in far greater detail than has been afforded in previous studies.

From the perspective of linguistic theory, the discovery of systematic variation in coarticulation at the level of the

individual lends, in part, support to the idea that coarticulation is phonological in nature (Flemming, 2001; Gafos, 2002; Flemming, 2016). That is, individual speakers must have their own production grammars with different parameters or weights set for the magnitude and temporal extent of the vocalic influence. While the question of what grammatical models are suitable for handling inter-individual variability in language is beyond the scope of this study, it seems clear that individual learners must be allowed to come up with different temporally sensitive phonological “rules” or constraints for particular coarticulation types.

The findings of this study not only further our understanding of the factors that govern coarticulation in speech production, they also have significant implications for research on sound change and language variation and change in general. As noted above, recent studies have called for situating the explanation for sound change actuation in the context of individual variability in speech perception and production (Beddor, 2009; Yu, 2010; Dimov *et al.*, 2012; Yu, 2013; Garrett and Johnson, 2013; Stevens and Harrington, 2014). To this end, it is interesting to note that the participants who exhibit the most robust vocalic effects on /s/ realization are females that have lower AQ scores. Such findings are consistent with research on language variation and change, which found females to often lead in early stages of sound change and the agents of propagation of change tend to have wider social network (Labov, 2001). Individuals with lower AQ would serve the role of agents of propagation of change well since they tend to be the individuals with better social and communication skills and have larger friendship bases (Yu, 2013). Nonetheless, it is important to note that the sound change that gives rise to the allophonic, albeit temporally dependent, variation of /s/ in Cantonese is well under way. That is, from the individual-difference perspective on sound change adopted here, even if speakers differ in the nature of the vocalic “coarticulation” on /s/ both qualitative and temporally, a sound change is actuated as soon as the speakers are producing the context-dependent variability in a controlled and systematic fashion (i.e., a difference in the production grammar). Thus, even if the female participants and those with lower AQ scores might seem like they are leading the way, in actuality, others are not far behind.

ACKNOWLEDGMENTS

Sincere thanks go to the anonymous reviewers and the associate editor for their comments and suggestions. Many thanks to Peggy Mok at Chinese University of Hong Kong for her assistance in subject recruitment and recording. Special thanks to the audiences at the linguistics department colloquia at University of California, Los Angeles, Hong Kong University, New York University, and at the Workshop on Sound Change in Interacting Human Systems at University of California, Berkeley. This work was partially supported by National Science Foundation Grant No. BCS-0949754. Any errors in this work are my own.

¹The English and Mandarin spectra were made from recordings of speakers in Ladefoged (1989) and Lee and Zee (2003), respectively.

²One reviewer commented that the striking difference in SII might stem from SII being better at capturing differences in sibilants across vowel contexts produced by females (cf. more than 40% of the data comes from male speakers). Subsequent analyses using PCA results performed on the male and female acoustic measures separately consistently confirmed a large difference in vocalic effects on SI between males and females, suggesting that the sex difference is not introduced by potential bias in the PCA procedure.

- Allen, J. S., Miller, J. L., and DeSteno, D. (2003). “Individual talker differences in voice-onset-time,” *J. Acoust. Soc. Am.* **113**(1), 544–552.
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. (American Psychiatric Association, Washington, DC).
- Austin, E. J. (2005). “Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ),” *Person. Ind. Diff.* **38**, 451–460.
- Baker, A., Archangeli, D., and Mielke, J. (2011). “Variability in American English s-retraction suggests a solution to the actuation problem,” *Lang. Var. Change* **23**(3), 347–374.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001a). “The ‘Reading the Mind in the Eyes’ Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism,” *J. Child Psychol. Psych.* **42**, 241–251.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001b). “The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males, females, scientists and mathematicians,” *J. Autism Develop. Disorders* **31**, 5–17.
- Bates, D., Maechler, M., and Bolker, B. (2011). *lme4*. R package version 0.999375-38.
- Bauer, R. S., and Benedict, P. K. (1997). *Modern Cantonese Phonology, Trends in Linguistics: Studies and Monographs 102* (Mouton de Gruyter, Berlin).
- Beddor, P. S. (2009). “A coarticulatory path to sound change,” *Language* **85**(4), 785–832.
- Beddor, P. S., Harnsberger, J., and Lindemann, S. (2002). “Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates,” *J. Phon.* **30**, 591–627.
- Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., and Bonnel, A.-M. (2003). “Enhanced pitch sensitivity in individuals with autism: A signal detection analysis,” *J. Cogn. Neurosci.* **15**, 226–235.
- Byrd, D. (1992). “Preliminary results on speaker-dependent variation in the TIMIT database,” *J. Acoust. Soc. Am.* **92**, 593–596.
- Chang, C., Yao, Y., Haynes, E. F., and Rhodes, R. (2011). “Production of phonetic and phonological contrast by heritage speakers of Mandarin,” *J. Acoust. Soc. Am.* **129**(6), 3964–3980.
- Chao, Y.-R. (1947). *Cantonese Primer* (Harvard University Press, Cambridge, MA).
- Cheung, K.-H. (1986). “The phonology of present-day Cantonese,” Ph.D. thesis, University College, London.
- Choi, J. D., and Keating, P. (1991). “Vowel-to-vowel coarticulation in three Slavic languages,” *UCLA Work. Pap. Phon.* **78**, 78–86.
- Cole, J., Lindebaugh, G., Munson, C. M., and McMurray, B. (2010). “Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach,” *J. Phon.* **38**, 167–184.
- Constantino, J. N., and Todd, R. D. (2003). “Autistic traits in the general population,” *Arch. Gen. Psych.* **60**, 524–530.
- Dimov, S., Katseff, S., and Johnson, K. (2012). “Social and personality variables in compensation for altered auditory feedback,” in *The Initiation of Sound Change: Perception, Production, and Social Factors*, edited by M. J. S. Sabater and D. Recasens (John Benjamins, Philadelphia), pp. 185–210.
- Flemming, E. (2001). “Scalar and categorical phenomena in a unified model of phonetics and phonology,” *Phonology* **18**(1), 7–44.
- Flemming, E. (2016). “The grammar of coarticulation,” in *La Coarticulation: Indices, Direction et Représentation (Coarticulation: Indices, Management and Representation)*, edited by M. Embarki and C. Dodane (L’Harmattan, Paris, in press).
- Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (1988). “Statistical analysis of word-initial voiceless obstruents: Preliminary data,” *J. Acoust. Soc. Am.* **84**, 115–123.
- Fowler, C. (2006). “Compensation for coarticulation reflects gesture perception, not spectral contrast,” *Percept. Psychophys.* **68**(2), 161–177.

- Gafos, A. I. (2002). "A grammar of gestural coordination," *Nat. Lang. Ling. Theory* 20, 269–337.
- Garrett, A., and Johnson, K. (2013). "Phonetic biases in sound change," in *Origins of Sound Change: Approaches to Phonologization*, edited by A. C. L. Yu (Oxford University Press, Oxford), pp. 51–97.
- Gordon, M., Barthmaier, P., and Sands, K. (2002). "A cross-linguistic acoustic study of fricatives," *J. Int. Phon. Assoc.* 32, 141–174.
- Grosvald, M. (2009). "Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation," *J. Phon.* 37(2), 173–188.
- Grosvald, M., and Corina, D. (2012). "Perception of long-distance coarticulation: An event-related potential and behavioral study," *Appl. Psycholing.* 33, 55–82.
- Happé, F. (1999). "Autism: Cognitive deficit or cognitive style?," *Trends Cogn. Sci.* 3, 216–222.
- Happé, F., and Frith, U. (2006). "The weak coherence account: Detail-focused cognitive style in autism spectrum disorders," *J. Autism Develop. Disorders* 36, 5–25.
- Harrington, J., Kleber, F., and Reubold, U. (2008). "Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study," *J. Acoust. Soc. Am.* 123(5), 2825–2835.
- Hashimoto, O. Y. (1972). *Phonology of Cantonese: Studies in Yue Dialects I* (Cambridge University Press, Cambridge).
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* 97, 3099–3111.
- Hoekstra, R. A., Bartels, M., Cath, D. C., and Boomsma, D. I. (2008). "Factor structure, reliability and criterion validity of the Autism-Spectrum Quotient (AQ): A study in Dutch population and patient groups," *J. Autism Develop. Disorders* 38(8), 1555–1566.
- Hughes, G. W., and Halle, M. (1956). "Spectral properties of fricative consonants," *J. Acoust. Soc. Am.* 28, 303–310.
- ICD-10 (1994). *International Classification of Diseases*, 10th ed. (World Health Organisation, Geneva, Switzerland).
- Iskarous, K., Shadle, C. H., and Proctor, M. I. (2011). "Articulatory-acoustic kinematics: The production of American English /s/," *J. Acoust. Soc. Am.* 129(2), 944–954.
- Jones, D., and Woo, K. T. (1912). *A Cantonese Phonetic Reader* (University of London Press, London).
- Jongman, A., Wayland, R., and Wong, S. (2000). "Acoustic characteristics of English fricatives," *J. Acoust. Soc. Am.* 108(3), 1252–1263.
- Jun, S.-A., and Bishop, J. (2015). "Priming implicit prosody: Prosodic boundaries and individual differences," *Lang. Speech* 58(4), 459–473.
- Kao, D. (1971). *Structure of the Syllable in Cantonese* (Mouton, The Hague).
- Kataoka, R. (2011). "Phonetic and Cognitive Bases of Sound Change," Ph.D. thesis, University of California, Berkeley, CA.
- Klatt, D. H. (1986). "The problem of variability in speech recognition and in models of speech perception," in *Invariance and Variability in Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 300–319.
- Labov, W. (2001). *Principles of Linguistic Change: Social Factors* (Blackwell, Oxford), Vol. 2.
- Ladefoged, P. (1989). "Report on the 1989 Kiel Convention," *J. Int. Phon. Assoc.* 19(2), 77–80.
- Lau, W. Y.-P., Gau, S. S.-F., Chiu, Y.-N., Wu, Y.-Y., Chou, W.-J., Liu, S.-K., and Chou, M.-C. (2013). "Psychometric properties of the Chinese version of the Autism Spectrum Quotient (AQ)," *Res. Dev. Disab.* 34(1), 294–305.
- Lee, W.-S., and Zee, E. (2003). "Standard Chinese (Beijing)," *J. Int. Phon. Assoc.* 33(1), 109–112, <https://www.internationalphoneticassociation.org/member/audio-files-illustrations-ipa> (Last viewed 10/21/15).
- Lee, W.-S., and Zee, E. (2010). "Articulatory characteristics of the coronal stop, affricate, and fricative in Cantonese," *J. Chin. Ling.* 38(2), 336–372, available at <http://www.jstor.org/stable/23754137>.
- Li, F. (2012). "Language-specific developmental differences in speech production: A cross-language acoustic study," *Child Develop.* 83(4), 1303–1315.
- Li, F., Edwards, J., and Beckman, M. (2007). "Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese," in *Proceedings of the XVIth International Congress of Phonetic Sciences, ICPhS*, Vol. 4, pp. 917–920.
- Li, F., Edwards, J., and Beckman, M. E. (2009). "Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers," *J. Phon.* 37(1), 111–124.
- Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H & H Theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht, the Netherlands), pp. 403–439.
- Lindblom, B., Guion, S., Hura, S., Moon, S.-J., and Willerman R. (1995). "Is sound change adaptive?," *Riv. Ling.* 7, 5–36, available at <http://www.italian-journal-linguistics.com/wp-content/uploads/LindblomC.pdf>.
- Lundström, S., Chang, Z., Råstam, M., Gillberg, C., Larsson, H., Anckarsäter, H., and Lichtenstein, P. (2012). "Autism spectrum disorders and autistic like traits: Similar etiology in the extreme end and the normal variation," *Arch. Gen. Psych.* 69(1), 46–52.
- Mann, V. A. (1980). "Influence of preceding liquid on stop-consonant perception," *Percept. Psychophys.* 28, 407–412.
- Manuel, S. Y. (1990). "The role of contrast in limit vowel-to-vowel coarticulation in different languages," *J. Acoust. Soc. Am.* 88, 1286–1298.
- Matingly, I. G. (1981). "Phonetic representation and speech synthesis by rule," in *The Cognitive Representation of Speech*, edited by T. Meyers, J. Laver, and J. Anderson (North-Holland, Amsterdam, the Netherlands), pp. 415–420.
- McMurray, B., and Jongman, A. (2011). "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations," *Psychol. Rev.* 118, 219–246.
- Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R* (Chapman and Hall, Boca Raton, FL).
- Mottron, L., Dawson, M., Soulières, I., Hubert, B., and Burack, J. (2006). "Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception," *J. Autism Develop. Disorders* 36, 27–43.
- Muñiz, J., García-Cueto, E., and Lozano, L. M. (2005). "Item format and the psychometric properties of the Eysenck personality questionnaire," *Person. Ind. Diff.* 38, 61–69.
- Newman, R. S. (1997). "Individual differences and the link between speech perception and speech production," Ph.D. thesis, State University of New York, Buffalo, NY.
- Newman, R. S., Clouse, S. A., and Burnham, J. L. (2001). "The perceptual consequences of within-talker variability in fricative production," *J. Acoust. Soc. Am.* 109, 1181–1196.
- Nieuwland, M. S., Ditman, T., and Kuperberg, G. R. (2010). "On the incrementality of pragmatic processing: An erp investigation of informativeness and pragmatic abilities," *J. Mem. Lang.* 63, 324–346.
- Nittrouer, S. (1995). "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," *J. Acoust. Soc. Am.* 97, 520–530.
- Ohala, J. J. (1994). "The frequency codes underlies the sound symbolic use of voice pitch," in *Sound Symbolism*, edited by L. Hinton, J. Nichols, and J. J. Ohala (Cambridge University Press, Cambridge), pp. 325–347.
- Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* 24, 175–184.
- Pulleyblank, E. G. (1996). "The Cantonese vowel system in historical perspective," in *Studies in Chinese Phonology*, edited by J. Wang and N. Smith (Walter De Gruyter, Berlin), pp. 185–218.
- Pycha, A. (2015). "Co-articulatory cues for communication: An investigation of five environments," *Lang. Speech* (in press), <http://las.sagepub.com/content/early/2015/09/07/0023830915603878.abstract> (Last viewed 10/21/15).
- Repp, B. H. (1981). "Two strategies in fricative discrimination," *Percept. Psychophys.* 30(3), 217–227.
- Robinson, E. B., Koenen, K. C., McCormick, M. C., Munir, K., Hallett, V., Happé, F., Plomin, R., and Ronald, A. (2011). "Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%)," *Arch. Gen. Psych.* 68(11), 1113–1121.
- Sachs, J., Lieberman, P., and Erickson, D. (1973). "Anatomical and cultural determinants of male and female speech," in *Language Attitudes: Current Trends and Prospects*, edited by R. W. Shuy and R. W. Fasold (Georgetown University Press, Washington, DC), pp. 74–84.
- Scarborough, R. A. (2004). "Coarticulation and the structure of the lexicon," Ph.D. thesis, University of California, Los Angeles, Los Angeles, CA.
- Scarborough, R. A. (2013). "Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation," *J. Phon.* 41, 491–508.
- Shadle, C. H., and Mair, S. (1996). "Quantifying spectral characteristics of fricatives," in *ICSLP 96. Proceedings of the Fourth International Conference on Spoken Language Processing*, IEEE, pp. 1521–1524.
- Singer, J. D., and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* (Oxford University Press, New York).

- Solé, M.-J. (2007). "Controlled and mechanical properties in speech: A review of the literature," in *Experimental Approaches to Phonology*, edited by M.-J. Solé, P. Beddor, and M. Ohala (Oxford University Press, Oxford), pp. 302–321.
- Soli, S. D. (1981). "Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation," *J. Acoust. Soc. Am.* **70**, 976–984.
- Stevens, K. N. (1998). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
- Stevens, M., and Harrington, J. (2014). "The individual and the actuation of sound change," *Loquens* **1**(1), e003.
- Stewart, M. E., and Ota, M. (2008). "Lexical effects on speech perception in individuals with 'autistic' traits," *Cognition* **109**, 157–162.
- Stuart-Smith, J. (2007). "Empirical evidence for gendered speech production: /s/ in Glaswegian," in *Laboratory Phonology 9*, edited by J. Cole and J. I. Hualde (Mouton de Gruyter, New York), pp. 65–86.
- Theodore, R. M., Miller, J. L., and DeSteno, D. (2009). "Individual talker differences in voice-onset-time: Contextual influences," *J. Acoust. Soc. Am.* **125**(6), 3974–3982.
- Tomiak, G. R. (1990). "An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents," Ph.D. thesis, SUNY Buffalo, Buffalo, NY.
- Turnbull, R. J. (2015). "Assessing the listener-oriented account of predictability-based phonetic reduction," Ph.D. thesis, The Ohio State University, Columbus, OH.
- Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., and Chung, M. K. (2011). "Development sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study," *J. Speech Lang. Hear. Res.* **54**, 995–1010.
- Wang, L. (1937). *Zhongguo yinyun-xue (Chinese Phonology)* (Shangwu, Shanghai), Vol. 2.
- Whalen, D. H. (1990). "Coarticulation is largely planned," *J. Phon.* **18**, 3–35.
- Wright, R. (2004). "A review of perceptual cues and cue robustness," in *Phonetically-Based Phonology*, edited by B. Hayes, R. Kircher, and D. Steriade (Cambridge University Press, Cambridge), pp. 34–57.
- Xiang, M., Grove, J., and Giannakidou, A. (2013). "Dependency dependent interference: NPI interference, agreement attraction, and global pragmatic inferences," *Front. Psychol.* **4**, 708.
- Yu, A. C. L. (2010). "Perceptual compensation is correlated with individuals' 'autistic' traits: Implications for models of sound change," *PLoS One* **5**(8), e11950.
- Yu, A. C. L. (2013). "Individual differences in socio-cognitive processing and the actuation of sound change," in *Origins of Sound Change: Approaches to Phonologization*, edited by A. C. L. Yu (Oxford University Press, Oxford, UK), pp. 201–227.
- Yu, A. C. L., Grove, J., Martinović, M., and Sonderegger, M. (2011). "Effects of working memory capacity and 'autistic' traits on phonotactic effects in speech perception," in *Proceedings of the International Congress of the Phonetic Sciences XVII*, edited by E. Zee (International Congress of the Phonetic Sciences, Hong Kong), pp. 2236–2239.
- Yu, A. C. L., and Lee, H. (2014). "The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study," *J. Acoust. Soc. Am.* **136**(1), 382–388.
- Zellou, G., and Tamminga, M. (2014). "Nasal coarticulation changes over time in Philadelphia English," *J. Phon.* **47**(1), 18–35.