JASA ARTICLE

Multi-modal cross-linguistic perception of fricatives in clear speech

Sylvia Cho,[1] Allard Jongman,[2] Yue Wang,[1,a)] and Joan A. Sereno[2]
[1]*Language and Brain Lab, Department of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada*
[2]*The University of Kansas Phonetics and Psycholinguistics Lab, Department of Linguistics, The University of Kansas, Lawrence, Kansas 66044, USA*

ABSTRACT:
Research shows that acoustic modifications in clearly enunciated fricative consonants (relative to the plain, conversational productions) facilitate auditory fricative perception, particularly for auditorily salient sibilant fricatives and for native perception. However, clear-speech effects on visual fricative perception have received less attention. A comparison of auditory and visual (facial) clear-fricative perception is particularly interesting since sibilant fricatives in English are more auditorily salient while non-sibilants are more visually salient. This study thus examines clear-speech effects on multi-modal perception of English sibilant and non-sibilant fricatives. Native English perceivers and non-native (Mandarin, Korean) perceivers with different fricative inventories in their native languages (L1s) identified clear and conversational fricative-vowel syllables in audio-only, visual-only, and audio-visual (AV) modes. The results reveal an overall positive clear-speech effect when visual information is involved. Considering the factor of AV saliency, clear speech benefits sibilants more in the auditory domain and non-sibilants more in the visual domain. With respect to language background, non-native (Mandarin and Korean) perceivers benefit from visual as well as auditory information, even for fricatives non-existent in their respective L1s, but the patterns of clear-speech gains are affected by the relative AV weighting and "nativeness" of the fricatives. These findings are discussed in terms of how saliency-enhancing and category-distinctive cues of speech sounds are adopted in AV perception to improve intelligibility.
 https://doi.org/10.1121/10.0001140

(Received 1 June 2019; revised 1 April 2020; accepted 6 April 2020; published online 27 April 2020)

[Editor: Benjamin V. Tucker] Pages: 2609–2624

I. INTRODUCTION

This study examines the auditory-visual (AV) perception of clearly and conversationally[1] produced English fricative consonants by native English, Mandarin, and Korean perceivers. The goal is to explore how clear speech enhances signal saliency and facilitates fricative category distinctions under the influence of auditory and visual salience of the input signal as well as the influence of native language (L1) phonetic systems.

Clear-speech research is typically framed under the hyper- and hypo-articulation (H & H) theory of speech production and perception, stating that intra-speaker phonetic variation that generates hyper- and hypo-speech is essentially a result of interacting talker-oriented output constraints that pertain to efforts in speech production and perceiver-oriented reception constraints that concern clarity of speech (Lindblom, 1990). In adverse conditions that make perception difficult (e.g., presence of background noise, perceivers with limited linguistic experience, hearing loss), speakers intentionally change the style of their speech in an attempt to facilitate perception (e.g., Summers *et al.*, 1988). Such efforts to aid intelligibility are often accompanied by a clear speaking style (relative to a plain, conversational style), which involves a more exaggerated manner of speaking that is characterized by properties such as greater movement of articulators, slower articulation, and more rapid vocal fold vibration, which is often associated with corresponding changes in the acoustics (e.g., Maniwa *et al.*, 2009; Moon and Lindblom, 1994).

These clear-speech features have been claimed to arise from two levels of modifications: language-independent, global-level enhancement of overall signal prominence (signal-based, e.g., increased overall amplitude of an utterance), and enhancement of language-specific cues used for sound category distinctions (code-based, e.g., increased F2 for front vowels) (Bradlow and Bent, 2002; Smiljanić and Bradlow, 2009). To facilitate perception, the extent of these modifications must retain the intrinsic characteristics of a segment and remain within-category, so that phonemic categorical distinctions can still be maintained (Moon and Lindblom, 1994; Ohala, 1995). Thus, one of the key issues to address in clear-speech research is to identify how signal- and code-based enhancement cues are utilized in perception to improve intelligibility of speech sound categories, and what factors influence the use of these cues.

a)Electronic mail: yuew@sfu.ca

A. Native clear speech perception as a function of AV weighting

Previous studies have found that clear speech significantly enhances intelligibility in a number of different native perceiver groups, including normal-hearing perceivers (Bradlow and Bent, 2002; Ferguson, 2004; Gagné *et al.*, 1995; Payton *et al.*, 1994), hearing-impaired perceivers (Payton *et al.*, 1994; Uchanski *et al.*, 1996), and children (Bradlow *et al.*, 2003). Such clear-speech perceptual gains are present at syllable (Gagné *et al.*, 2002), word (Gagné *et al.*, 1994; Uchanski *et al.*, 1996) and sentential levels (Bradlow and Alexander, 2007; Bradlow and Bent, 2002; Bradlow *et al.*, 2003; Schum, 1996).

Critically relevant to the present study, Maniwa *et al.* (2009) found that clear speech changes a number of acoustic features in English fricatives, including increases in frication duration, spectral measures (peak frequency, mean, skewness), frequency regions of energy concentrations, and fundamental frequency of the neighboring vowels, as well as lower energy in frequencies below 500 Hz (Maniwa *et al.*, 2009). Maniwa *et al.* (2009) further found an effect of sibilance (intensity of friction noise), as only clear non-sibilants (i.e., low-intensity fricatives: /f, v, θ, ð/) but not sibilants (i.e., high-intensity fricatives: /s, z, ʃ, ʒ/) decreased in spectral kurtosis and overall root-mean-square (rms) amplitude. There was also an effect of voicing as the inherently longer voiceless fricatives were lengthened more than the voiced fricatives in clear speech. Such changes in the acoustic signal were found to be relevant to perception in an accompanying study. Maniwa *et al.* (2008) investigated clear-speech effects on English fricative perception by embedding fricatives in multi-talker babble noise. They found that clear speech helps with both voicing and place distinctions of fricative pairs among native perceivers, as lower signal-to-noise (SNR) ratio thresholds of babble noise were required for fricative minimal pair distinctions in clear speech (compared to conversational speech). Maniwa *et al.* (2008) also found an effect of sibilance. Specifically, the perception of sibilant fricatives (which are auditorily more salient, with more defined spectral peaks and greater amplitude than non-sibilants; Jongman *et al.*, 2000; Maniwa *et al.*, 2009) benefited more from clear speech than the perception of non-sibilants. Indeed, some of the non-sibilant fricatives showed no clear-speech gains in perception (e.g., /v/-/ð/, /f/-/v/ distinctions).

In addition to the auditory domain, clear speech has also been found to affect intelligibility in visual (facial) speech perception (Gagné *et al.*, 1994; Gagné *et al.*, 2002; Helfer, 1997; Lander and Capek, 2013; Van Engen *et al.*, 2014). For example, Gagné *et al.* (1994) and Gagné *et al.* (2002) examined the perception of clear and conversational French consonant-vowel (CV) syllables (/b, d, g, v, z, ʒ/ + /i, y, a/) and found significant clear-speech gains in the intelligibility of AV, visual-only, as well as auditory-only presentations, demonstrating the existence of an overall clear-speech advantage across input modalities in the perception of consonants and vowels. No research has examined visual clear-speech effects for fricatives, although visual benefits were observed in auditory-visual perception of English non-sibilant fricatives (Jongman *et al.*, 2003; Wang *et al.*, 2008, 2009). While non-sibilants (e.g., labial, labiodental), which are more visible due to their more anterior articulation, benefit more from visual cues, the perception of acoustically more salient sibilants (e.g., alveolar, post-alveolar) mostly relies on auditory input (Hazan *et al.*, 2006; Iba, 2005; Iba *et al.*, 2004; Jongman *et al.*, 2003). These patterns of AV weighting, where perceivers utilize information from alternate modalities when one modality is degraded or less salient, have been observed in other research (Chen and Massaro, 2004; Gagné *et al.*, 2002; Hazan *et al.*, 2010; Robert-Ribes *et al.*, 1998; Traunmüller and Öhrström, 2007; Van Engen *et al.*, 2014), raise questions for further research regarding the role of clear speech in multi-modal perception of English fricatives. Given that auditory clear-speech effects are more prominent for the perception of sibilants, which are acoustically more salient than non-sibilants (Maniwa *et al.*, 2008, 2009), would the perception of the more visible non-sibilants enjoy greater visual clear-speech benefits? Or would visual clear-speech effects exist for the perception of sibilants and non-sibilants alike?

B. Non-native perception as a function of L1 background and AV weighting

Research on non-native (L2) perception of clear speech has indicated that clear speech may be helpful to L2 perceivers, but to a smaller degree compared to L1 perception due to influence by perceivers' L1 phonetic systems (Bradlow and Bent, 2002; Fenwick *et al.*, 2015; Granlund *et al.*, 2012; Rogers *et al.*, 2010; Smiljanić and Bradlow, 2011). For example, in the auditory perception of English fricatives, Kabak and Maniwa (2007) found that both native English and non-native German-L1 perceivers benefit more from clear speech for sibilants than non-sibilants. Although similar extents of clear-speech benefits were observed for non-native and native perceivers, non-natives did not reach the native level in the perception of clear non-sibilant fricatives, presumably stemming from the fact that German does not have interdental fricatives and therefore a phonological contrast does not exist between labiodental and interdental sounds.

Acoustic salience may also play a role in non-native auditory perception in clear speech. For example, Fenwick *et al.* (2015) examined Australian English perceivers and their patterns of perceptual assimilation in conversational and clear productions of Sindhi contrasts. The results showed that non-native perceivers benefited from clear speech when discriminating sounds with voicing contrasts (i.e., /f-v/ or /t̪-d̪/, which are acoustically quite salient), but not when discriminating sounds with place-of-articulation contrasts (i.e., /ɓ-ɗ/ or /d̪-ɖ/, which are not very salient acoustically).

These results raise the issue of whether non-native perception would benefit more from clear speech in the visual domain with visually more salient input than with visually

less salient input. However, little research has examined clear-speech effects on non-native multimodal perception. Studies on non-native visual speech perception have generally shown visual benefit as a function of L1 influence (de Gelder and Vroomen, 1992; Hazan *et al.*, 2006; Wang *et al.*, 2008, 2009). For example, Wang *et al.* (2009) showed an overall visual benefit in the perception of English non-sibilant fricatives by Korean and Mandarin perceivers, but these non-native perceivers could not reach native performance in the identification of fricatives absent in their L1s (i.e., labiodental fricatives and interdental fricatives for Korean, and interdental fricatives for Mandarin). Furthermore, perceivers are better at attuning to the visual cues of a non-native sound that has a native counterpart, even if that sound does not have phonemic status in a language. For example, Spanish perceivers benefit more from visual information than Japanese perceivers in the discrimination of English labial versus labiodental sounds as Spanish has an allophonic /f/ with the labiodental place of articulation, whereas labiodental sounds do not occur at all in Japanese (Hazan *et al.*, 2006).

Non-native perceivers also exhibit different auditory and visual weighting patterns than native perceivers. It has been found that visual reliance is greater in non-native relative to native perception (Chen and Hazan, 2007; Hannah *et al.*, 2017; Sekiyama and Tohkura, 1993; Wang *et al.*, 2008, 2009). However, further research reveals that such visual reliance interacts with the factors of AV saliency and L1 background. Specifically, visual effects were larger for contrasts that involve more visually salient sounds (e.g., labial, dental) than for contrasts that are visually less salient (e.g., alveolar, post-alveolar) (Hazan *et al.*, 2006; Iba, 2005; Kawase *et al.*, 2014; Wang *et al.*, 2008, 2009). Moreover, such reliance on visual information may be inhibitory and lead to poorer performance when processing unfamiliar L2 visual cues such as English interdental fricatives for Mandarin or alveolar approximants for Japanese natives (Kawase *et al.*, 2014; Wang *et al.*, 2009). These results raise the question of whether difficulty in the perception of visually salient L2 cues can be alleviated in clear speech. Limited research on clear speech shows an unexpected pattern of clear-speech effects as a function of AV saliency and L1 influence. In Fenwick *et al.* (2015), while English perceivers showed a visual benefit across speaking styles in the perception of Sindhi consonant place of articulation contrasts (i.e., clear speech did not further enhance the visual gain already exhibited in conversational speech), visual benefits were exhibited only in clear speech (but not in conversational speech) for the perception of the less visually salient voicing contrasts. The results indicate that clear speech with an enhanced signal may be beneficial for the perception of less visually salient (and thus more challenging) non-native contrasts.

C. The present study

1. Rationale

As discussed in the above review, clear speech may enhance auditory speech intelligibility, particularly for acoustically prominent and native sounds. Factors concerning signal saliency and nativeness are central to the issue of how clear-speech modifications that increase overall signal saliency and those that enhance specific phonemic category distinctions are utilized to aid intelligibility. However, in the multi-modal perception domain, previous research has not been able to determine how the existence and degree of clear-speech benefits are affected by the relative auditory-acoustic and visual-articulatory saliency of a sound, nor whether these auditory and visual cues are familiar in perceivers' L1.

English fricatives present a particularly good test for research concerning clear-speech effects on multi-modal perception given the conflict in auditory and visual salience found with different fricative sounds. That is, the non-sibilants are relatively salient in the visual domain while being less salient in the auditory domain, whereas the opposite is true for the sibilant fricatives. The aforementioned studies suggest that clear speech helps sibilant identification in the auditory domain (Maniwa *et al.*, 2009), but it is unclear whether the same effects will be found in modalities that involve visual information as English sibilants are visually less marked. The post-alveolar fricatives can involve lip protrusion (Flemming, 2002; Ladefoged and Maddieson, 1996) and therefore perception may benefit from the added visual information to a certain degree, but the same assumptions cannot be applied to the alveolar fricatives as they are not as visually salient during production. On the other hand, identification of the challenging non-sibilant contrasts in the auditory domain has been found to benefit from added visual information (Jongman *et al.*, 2003; Wang *et al.*, 2008, 2009). This suggests that visual information is especially important for the accurate perception of the non-sibilants, and, therefore, greater clear-speech benefits may be observed in non-sibilant fricatives in domains that include visual information. Additionally, it is worthwhile to examine perception in the AV modality in which both auditory salience and visual salience are at play since it is largely unknown how these factors interact with sibilance and place of articulation to influence auditory and visual cue weightings. Language background is also a key factor to unravel these interactive effects on multi-modal fricative perception. Previous studies find that non-native (relative to native) perceivers have greater visual reliance, but they may not benefit from visual input if the visual cues associated with certain fricatives are not utilized in their L1 (Kawase *et al.*, 2014; Wang *et al.*, 2008, 2009). What is unclear is whether clear speech can enhance perception of such non-native visual cues, and how the effects of visual as well as auditory enhancements in clear speech interact with L1 influence and signal saliency.

2. Questions and predictions

This study thus examines native and non-native multimodal perception of English fricatives that differ in sibilance and place of articulation (i.e., labiodental, interdental,

alveolar, post-alveolar) in conversational and clear speech. By doing so, this study aims to address the following questions.

First, are clear fricatives more intelligible than conversational fricatives when visual information is available? And if so, are greater clear-speech benefits found with certain fricatives as a function of auditory and visual saliency? Based on previous findings, we predict that enhancements will be observed in fricatives when clearly-articulated fricatives provide more exaggerated visual cues. Furthermore, we hypothesize that the perception of non-sibilant fricatives will benefit more from the visual cues of clear speech than the perception of sibilants in the visual modality. In contrast, we expect a greater degree of auditory clear-speech gain for sibilants, as the English sibilants are acoustically more salient than the non-sibilants. We therefore predict that clear speech may benefit all fricatives in a complementary manner, where the clearly-articulated visual cues facilitate perception of visually salient non-sibilants, and the clearly enunciated auditory information helps sibilant fricative identification.

Second, how does L1 background affect multi-modal clear fricative perception? Based on the previous findings, we expect clear-speech benefits to vary as a function of the interaction between saliency and "nativeness" of the fricatives. If perception is driven by overall enhanced saliency of the signal in clear speech, we should expect similar clear-speech effects across fricatives and L1s. However, if perception is affected by category-distinctive clear-speech cues, the extent of clear-speech facilitation may differ among fricatives and L1s, depending on whether the clearly-articulated auditory or visual cues exist in their native fricative inventories.

3. Selection of non-native groups

To address the issue of L1 influence and determine whether clear-speech modifications enhance AV identification of non-native fricatives, the current study included perceivers with different L1 backgrounds. Two groups of non-native perceivers, namely, Korean and Mandarin Chinese, are included in this study because of the differences in their L1 fricative inventories. In contrast with English, the Mandarin fricative inventory does not contain interdental fricatives (Ladefoged and Wu, 1984; Svantesson, 1986), while Korean only contains alveolar fricatives (lax /s/ and tense /s*/) (Cheon and Anderson, 2008; Kim, 1972; Kong *et al.*, 2014; Lee and Jongman, 2016; Schmidt, 1996). Additionally, neither Mandarin nor Korean has any voiced fricatives. A comparison of the fricative inventories in English, Mandarin, and Korean is given in Table I. These differences allow us to address the questions regarding L1 effects on clear fricative perception.

For non-sibilants, the auditory and visual cues of interdental fricatives are non-native to both Mandarin and Korean groups, while the labiodental cues are non-native only to Korean perceivers. We expect Mandarin perceivers to outperform Korean perceivers with labiodental but not interdental perception (cf. Wang *et al.*, 2009). However, differences between the two groups in the extent of clear-speech gain, particularly in the visual domain, should reveal whether clear-speech enhancement in perception is affected by L1 background. With sibilant perception, both Mandarin and Korean groups may benefit from clear speech in the auditory domain. If clear-speech perception is L1-driven, Korean perceivers are expected to be worse than Mandarin perceivers at using the enhanced auditory and visual cues for the post-alveolar fricatives that do not exist in Korean.

II. METHODS

A. Perceivers

Twenty-four (14 female, 10 male) native perceivers of Western Canadian English [$M = 22$ years, standard deviation (SD) = 2], eighteen (11 female, 7 male) non-native perceivers who spoke Mandarin Chinese as their L1 ($M = 22$ years, $SD = 3$ years), and thirty (15 female, 15 male) non-native perceivers whose L1 was Korean ($M = 26$ years, $SD = 4$) were recruited in the greater Vancouver area, Canada. None of the participants reported issues with hearing, vision, or language disorders.

The Mandarin perceivers were intermediate-level, late learners of English that were recruited from the student population at Simon Fraser University (SFU). They had studied English in an L2 classroom setting for about 11 years on average ($SD = 4$), starting at the age of 11 ($SD = 4$). They had arrived in Canada at a mean age of 16 ($SD = 4$). Their mean length of residence (LOR) was approximately 5 years ($SD = 4$). The Mandarin perceivers reported that their daily use of English was 55% on average ($SD = 19$). Their reported International English Language Testing System (IELTS, which has a range from 1 to 9) scores ranged from 5.5 to 7.0.

TABLE I. Fricative inventory of English, Mandarin, and Korean (x, present; blank cell, absent).

	Non-sibilant				Sibilant			
	Labiodental		Interdental		Alveolar		Post-alveolar	
	Voiceless	Voiced	Voiceless	Voiced	Voiceless	Voiced	Voiceless	Voiced
English	x	x	x	x	x	x	x	x
Mandarin	x				x		x	
Korean					x			

https://doi.org/10.1121/10.0001140

The Korean perceivers were also intermediate-level, late learners of English and they had studied English in an L2 classroom setting for about 12 years on average ($SD = 4$), starting at the age of 10 ($SD = 4$). Most of the participants were in Vancouver on a working holiday visa, and they had arrived in Canada at a mean age of 22 ($SD = 2$). Their mean LOR was approximately one year ($SD = 2$). The Korean perceivers reported that their daily use of English was 48% on average ($SD = 23$). Their reported Test of English for International Communication (TOIEC, which ranges from 10 to 990) scores or IELTS scores ranged from 500 to 860 and from 6.0 to 7.0, respectively.

B. Materials

1. Stimuli

The stimuli included conversationally and clearly spoken English CV syllables that contained eight English fricatives differing in sibilance, place of articulation, and voicing, followed by /a/. The stimuli are provided in Table II.

2. Talkers

Four native speakers of Western Canadian English (2 male, 2 female; aged 17–30) were recruited from SFU as talkers. None of the talkers exhibited strong regional dialects, and all talkers reported normal hearing without a history of speech or language disorders.

3. Recording of stimuli

Audio and video recordings of the talkers were made in a sound-attenuated booth in the Language and Brain Lab at SFU. Video recordings were made using a Canon Vixia HF S30 camera at a recording rate of 29 frames per second (fps) and separate but concurrent audio recordings were made with a Shure KSM 109 condenser microphone, using SoundForge 6.4 at a sampling rate of 48 kHz. The elicitation of clear and conversational tokens followed the procedures from previous studies (Maniwa *et al.*, 2009; Tang *et al.*, 2015). A simulated interactive computer software program was developed using MATLAB, which seemingly attempted to perceive and recognize the tokens that were produced by talkers. One of the eight target syllables was randomly presented on a computer screen, and talkers were instructed to pronounce the presented syllable naturally as if in a conversation. After each production, the screen displayed the program's identification of the token. The software was programmed to make systematic errors in its guesses in three different response types in terms of place of articulation (e.g., /sa/ for /ʃa/), voicing (e.g., /sa/ for /za/), and both (e.g., /sa/ for /ʒa/). When a target syllable had been misperceived by the software program, talkers were instructed to reiterate the token more clearly, to help the software correctly identify the production. The talkers' first attempt at pronouncing a target syllable was considered the conversational style production, and the reiteration in response to feedback comprised the clear style production. In total, 48 tokens were acquired for each speaker (eight syllables × two styles × three response types), and the talkers produced these tokens three times in three successive recording blocks. Each of the audio and video tokens was then evaluated by two phonetically trained native speakers of Canadian English to ensure the accuracy of pronunciation and quality of recordings. For each token, one best production from the three repetitions was selected. The tokens that were included in the experiment were rated to be correct and most intelligible by both raters (above 4 on a scale from 1 to 5, with 1 being least intelligible and 5 being most intelligible).

4. Stimulus preparation

Three sets of stimuli were developed for the different input modalities: auditory-only (AO), visual-only (VO), and AV. The AO stimuli were obtained by excising two-second audio clips from the microphone recordings. The VO stimuli were created by removing the on-camera audio track from the video recordings. The AV stimuli were obtained by replacing the on-camera audio tracks with the high-quality audio recordings taken from the microphone. Both the VO and AV stimuli were longer than the AO stimuli, namely, four seconds long, to ensure that each video clip included the opening and closing of the talkers' mouths. The overall intensity of the audio stimuli for the AO and AV trials was normalized at 60 dB sound pressure level (SPL). A rms normalization was performed in Praat (which normalizes overall intensity of the stimulus words to a specified dB SPL).

To ensure that sufficient errors were induced in these conditions (in order to compare conversational vs clear speech), the normalized audio stimuli were embedded in 12 talkers (six females, six males) babble noise, which was created for a previous study (Maniwa *et al.*, 2008). Four stretches of noise were selected from a random location within a 60-s noise sample.

The SNRs for sibilant and non-sibilant stimuli were established empirically through a pilot study. The sibilants were presented at SNRs of −3, −6, −9, −12, −15, and −18 dB, while the non-sibilants were presented at SNRs of +6,

TABLE II. Stimuli used in the current study.

Sibilance	Non-sibilant				Sibilant			
Place	Labiodental		Interdental		Alveolar		Post-alveolar	
Voicing	Voiceless	Voiced	Voiceless	Voiced	Voiceless	Voiced	Voiceless	Voiced
Stimuli	/**fa**/	/**va**/	/**θa**/	/**ð**a/	/**sa**/	/**za**/	/**ʃa**/	/**ʒa**/

+3, 0, −3, −6, and −9 dB. Based on previous research (Gagné *et al.*, 2002; Wang *et al.*, 2008), an SNR that yielded a 20%–30% error rate was selected as the target SNR level. The pilot included 192 randomly selected audio stimuli (two talkers × two styles × eight words × six SNRs) that were not included in the main perception experiment. Nine native English listeners (8 females, 1 male) and five native Korean listeners (3 females, 2 males) participated in the pilot study. After being presented with an audio stimulus embedded in noise, the participants were instructed to indicate which syllable they had heard among eight possible alternatives that were displayed on a computer screen (i.e., eight fricatives in onset position followed by /a/). The pilot study revealed that the target error rate was achieved at SNRs of −9 dB for sibilants and 0 dB for non-sibilants.

C. Procedures

Perceivers were presented with either a conversational or a clear stimulus syllable in AV, AO, or VO input modalities using Paradigm software (Perception Research Systems, 2007). A total of 576 stimuli (four talkers × two styles × three modalities × eight syllables × three response types) comprised the main perception experiment, and the stimuli were presented over two testing sessions that lasted approximately 60 min each, with a 10–15 min inter-session break. Each testing session consisted of three blocks, one for each input modality (AV, AO, VO). In each block, conversational and clear fricative tokens were presented randomly, and the order of the input modalities was counter-balanced across participants. Each participant completed a total of six blocks (i.e., two testing sessions; three blocks in each testing session).

The experiment was administered in a sound-treated booth in the Language and Brain Lab. The participants were seated in front of a flat screen monitor. The visual stimuli were presented on the computer screen and the auditory stimuli were presented binaurally over AKG Studio 141 headphones at 50 dBA, measured at the headphone using a Galaxy Audio checkmate CM-130 SPL meter. Following the presentation of a stimulus, participants were required to identify the fricative sound that they perceived in an eight-alternative forced choice identification task. All eight response options were shown on the computer screen, after the presentation of each stimulus. Participants were instructed to make their decision as soon as possible using a computer mouse; they had up to four seconds to enter their response. Response choices were given using English orthography (i.e., *fa* for /fa/, *va* for /va/, *tha* for /θa/, *dha* for /ða/, *sa* for /sa/, *za* for /za/, *sha* for /ʃa/, *zha* for /ʒa/,). All participants were given real-word examples to illustrate phoneme-to-grapheme mappings. For example, participants were told that *tha* has the same first sound as the word *think* and that *dha* has the same first sound as *than*. The participants were then instructed to read out loud the eight response choices to confirm their awareness before they were seated in the testing booth.

A familiarization session was administered at the beginning of each of the two testing sessions to make sure that perceivers were familiar with the eight-alternative forced choice task and the syllables they would be perceiving throughout the experiment. The participants were first presented with the eight stimulus syllables auditorily without any noise, allowing perceivers to hear the unmasked target stimuli. Afterward, the participants were presented with two example trials of different input modalities in noise (AV and AO stimuli were embedded in noise; VO stimuli do not contain auditory stimuli; therefore, they are not embedded in noise) so that they would know what to expect in each condition during the actual experiment and could practice the task.

III. RESULTS

A. Identification accuracy

The accuracy of the responses was calculated according to speech style, sibilance, input modality, and L1 group. The percent correct identification of the target fricatives was analyzed using a four-way mixed analysis of variance (ANOVA) with percent accuracy as the dependent variable, Style (Conversational, Clear), Sibilance (Non-sibilant, Sibilant), and Modality (AV, AO, VO) as repeated measures; and L1 (English, Mandarin, Korean) as the between-subjects factor. Percent accuracy was calculated by averaging the scores obtained by each participant across the four talkers, three response types, and individual tokens for each style, sibilance, and modality.

The mixed ANOVA analysis found significant main effects of Style [$F(1, 69) = 60.527$, $p < 0.001$, $\eta^2 = 0.467$], Sibilance [$F(1, 69) = 53.603$, $p < 0.001$, $\eta^2 = 0.437$] and Modality [$F(2, 111.934) = 522.117$, $p < 0.001$, $\eta^2 = 0.883$]. In general, clearly produced tokens ($M = 64\%$, $SD = 19\%$) were more intelligible than conversational ones ($M = 60\%$, $SD = 20\%$; $p < 0.001$), and sibilant fricatives ($M = 65\%$, $SD = 20\%$) were identified with higher accuracy than non-sibilant fricatives ($M = 59\%$, $SD = 20\%$; $p < 0.001$). For modality, *post hoc* comparisons with Bonferroni adjustments reveal that fricatives were most intelligible in the AV modality ($M = 76\%$, $SD = 14\%$) and most difficult to perceive in the VO modality ($M = 44\%$, $SD = 11\%$), with AO modality in the middle ($M = 66\%$, $SD = 16\%$) (all $p < 0.05$). The analysis also found a significant main effect of L1 [$F(2, 69) = 14.464$, $p < 0.001$, $\eta^2 = 0.295$]. Pairwise comparisons with Tukey's HSD (honestly significant difference) test adjustments found that English perceivers ($M = 69\%$, $SD = 19\%$) performed better than Mandarin ($M = 58\%$, $SD = 19\%$) and Korean ($M = 59\%$, $SD = 18\%$) perceivers (all $p < 0.05$), while the difference between Mandarin and Korean perceivers was not significant ($p = 0.974$).

There was no significant four-way interaction between Style, Sibilance, Modality, and L1 [$F(3.582, 123.569) = 0.106$, $p = 0.972$, $\eta_p^2 = 0.003$], but significant three-way interactions were found between Style × Sibilance × L1 [$F(2, 69) = 6.304$, $p = 0.003$, $\eta_p^2 = 0.154$], Style × Modality

https://doi.org/10.1121/10.0001140

× L1 [$F(3.409, 117.611) = 5.049$, $p = 0.002$, $\eta_p^2 = 0.128$], and Style × Sibilance × Modality [$F(1.791, 123.569) = 16.995$, $p < 0.001$, $\eta_p^2 = 0.198$]. A two-way interaction was found between Style × Modality [$F(1.705, 117.611) = 18.702$, $p < 0.001$, $\eta_p^2 = 0.213$]. For brevity, significant interactions that do not involve Style are not reported here since the present study focuses on effects of speech style. Further analyses were carried out to explore the specific nature of these significant interactions involving Style.

1. Style comparisons

The first set of additional analyses compared speech style effects for individual input modalities, sibilance categories, and L1 groups. Figure 1 presents style comparisons for the mean percent of correct fricative identification in each modality and sibilance category for the English, Mandarin, and Korean groups.

Separate one-way ANOVAs with Style as a factor for each Modality, Sibilance category, and each L1 group show different patterns of clear-speech effects. In the AO modality, while the English group showed a significant positive clear-speech effect for both non-sibilants [$F(1, 23) = 14.265$, $p = 0.001$, $\eta^2 = 0.383$] and sibilants [$F(1, 23) = 67.634$, $p < 0.001$, $\eta^2 = 0.746$], both non-native groups (Mandarin, Korean) revealed a significant positive clear-speech effect for sibilants {Mandarin: [$F(1, 17) = 14.111$, $p = 0.002$, $\eta^2 = 0.454$]; Korean: [$F(1, 29) = 34.557$, $p < 0.001$, $\eta^2 = 0.544$]} but not non-sibilants {Mandarin: [$F(1, 17) = 0.006$, $p = 0.937$, $\eta^2 < 0.001$]; Korean: [$F(1, 29)$

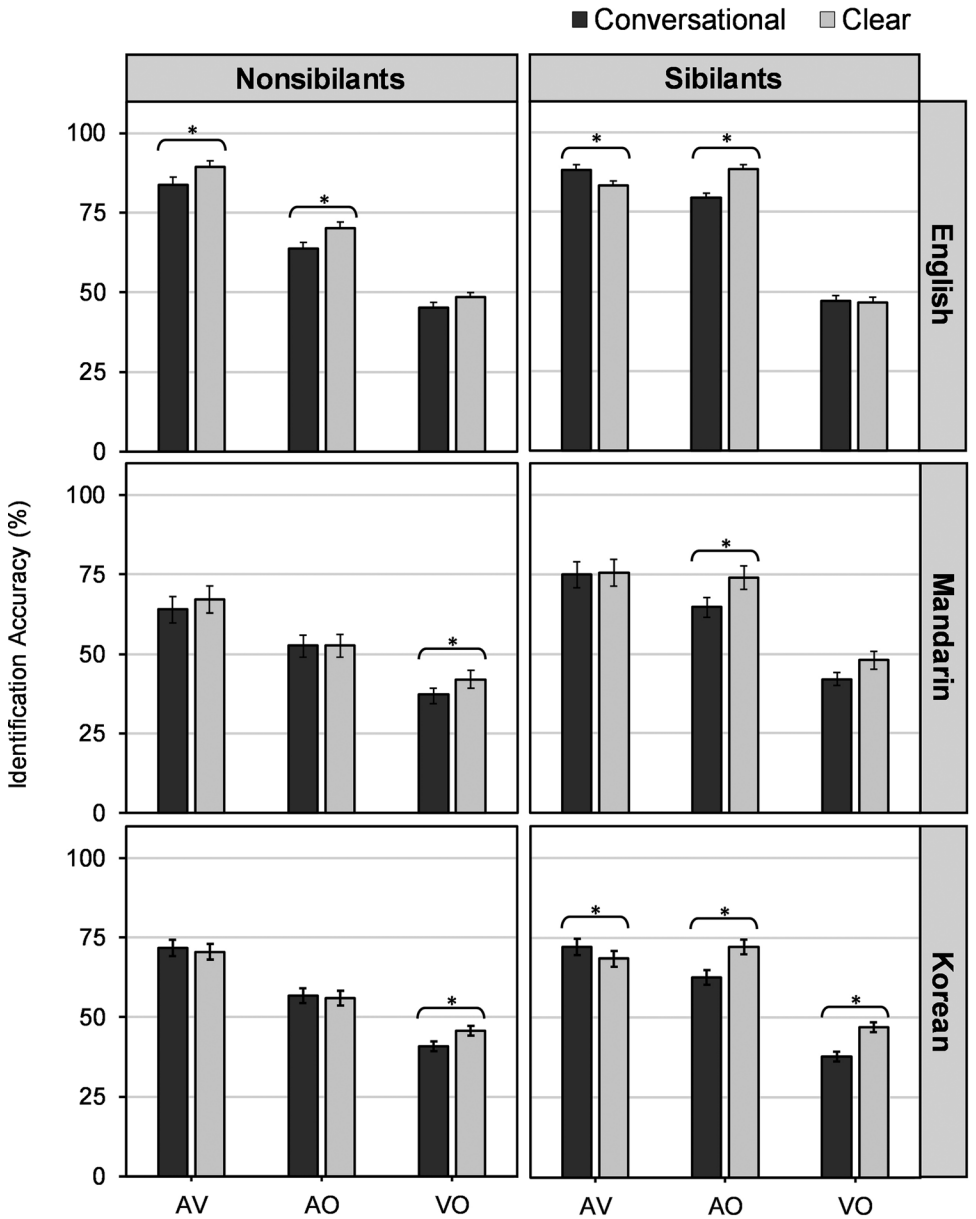

FIG. 1. Mean percent accuracy for different speech styles (conversational, clear), as a function of Modality (AV, AO, VO) and Sibilance by native perceivers of English (top panel), Mandarin (middle panel), and Korean (bottom panel). "*" indicates statistically significant differences ($p < 0.05$). Error bars indicate standard error.

$= 0.114$, $p = 0.738$, $\eta^2 = 0.111$]}. In the VO modality, the English group exhibited no significant clear-speech effect for either non-sibilants [$F(1, 23) = 2.878$, $p = 0.103$, $\eta^2 = 0.111$] or sibilants [$F(1, 23) = 0.054$, $p = 0.818$, $\eta^2 = 0.002$], whereas significant positive clear-speech effects were found for non-sibilants in both non-native groups {Mandarin: [$F(1, 17) = 6.916$, $p = 0.018$, $\eta^2 = 0.289$]; Korean: [$F(1, 29) = 8.907$, $p = 0.006$, $\eta^2 = 0.235$]} and for sibilants in the Korean group [$F(1, 29) = 15.457$, $p < 0.001$, $\eta^2 = 0.348$]. In the AV modality, the English group showed a significant positive clear-speech effect for non-sibilants [$F(1, 23) = 8.625$, $p = 0.007$, $\eta^2 = 0.273$] and an unexpected negative clear-speech effect (i.e., conversational speech more accurately perceived than clear speech) for sibilants [$F(1, 23) = 36.937$, $p < 0.001$, $\eta^2 = 0.616$]. For non-native perceivers, no clear-speech effects were observed for non-sibilants in either the Mandarin [$F(1, 17) = 4.218$, $p = 0.056$, $\eta^2 = 0.199$] or the Korean groups [$F(1, 29) = 1.802$, $p = 0.190$, $\eta^2 = 0.058$], or for sibilants in the Mandarin group [$F(1, 17) = 0.115$, $p = 0.739$, $\eta^2 = 0.007$]. However, a negative clear-speech effect was found for Korean sibilant perception [$F(1, 29) = 8.010$, $p = 0.008$, $\eta^2 = 0.216$], just as for the English group.

Analyses were also conducted using generalized linear mixed effects modeling (GLMM), built with the function "glmer" from R package "lme4." The overall GLMM involved four fixed effects (Style, Modality, Sibilance, and L1), four random effects (Perceiver, Talker, Response Type, and Fricative), and the response variable Accuracy is treated as binary (correct, incorrect). The optimal overall model was determined by forward selection to find the most complex random slope structure necessary to fit the full data: Accuracy ~ Modality * Style * L1 * Sibilance + (1 + Modality + Sibilance | Perceiver) + (1 + Modality + Style + L1 + Sibilance | Talker) + (1| Response.Type) + (1 + Modality +Style + L1| Fricative). The main results were comparable to those performed with the ANOVAs. Since the current study involves a complex design with four main factors, the ANOVA results are presented in this article for clarity and conciseness of reporting.

During the stepwise model selection procedure in the GLMM analysis, the only difference between GLMM and ANOVA was that adding the random slope of "Style|Talker" resulted in changes in the significance of the main effect of "Style," suggesting potential interactions of Style × Talker. To investigate such potential interactions, a separate set of ANOVAs were conducted involving Talker (two males and two females: 1M, 2M, 1F, 2F) as a factor along with Style, Modality, and Sibilance. The talker analyses are based on the native English perceivers' data given that non-native perceiver patterns may involve confounding factors rooted in influences other than the talker (e.g., a reduced ability to discriminate the fricatives). The results reveal a significant main effect of Talker [$F(2.765, 182.504) = 37.036$, $p < 0.001$, $\eta^2 = 0.359$] in addition to significant main effects of the other factors shown previously. Significant interactions were found between Style, Modality, Sibilance, and Talker [$F(5.272, 347.970) = 8.339$, $p < 0.001$, $\eta_p^2 = 0.112$], and particularly between Style x Talker [$F(2.825, 186.438) = 40.130$, $p < 0.001$, $\eta_p^2 = 0.378$]. Subsequent one-way ANOVAs were conducted with Style as a factor for each Talker, Modality and Sibilance category.

In the AV perception of sibilant fricatives, one talker (1M) yielded a significant negative clear-speech effect [$F(1, 20) = 48.215$, $p < 0.001$, $\eta^2 = 0.707$], which was consistent with the overall results reported previously, with the other talkers showing a similar trend. For non-sibilant fricatives, the positive clear-speech effect revealed in the overall results was shown in three talkers' data {two males: [$F(1, 20) = 8.532$, $p = 0.008$, $\eta^2 = 0.299$]; 1F: [$F(1, 20) = 6.636$, $p = 0.018$, $\eta^2 = 0.249$]; two females: [$F(1, 20) = 8.438$, $p = 0.009$, $\eta^2 = 0.297$]}, with talker 1M showing a null effect of clear speech.

In the AO perception of sibilant fricatives, talker 1F showed a significant positive clear speech effect [$F(1, 20) = 86.822$, $p < 0.001$, $\eta^2 = 0.813$], aligned with the group effect and the other talkers yielded the same trend, although non-signficant, presumably in part due to a ceiling effect in the conversational tokens that were perceived with a high accuracy (above 85%). For non-sibilants, two talkers' data showed a significant positive clear-speech effect {2M: [$F(1, 20) = 9.369$, $p = 0.006$, $\eta^2 = 0.319$]; 1F: [$F(1, 20) = 26.405$, $p < 0.001$, $\eta^2 = 0.569$]} and one talker showed a similar trend, consistent with the overall results, with the exception again of talker 1M who showed a trend of a negative clear-speech effect.

In the VO perception of sibilant fricatives, no significant talker effects were found, consistent with the overall results. For non-sibilants, while a significant clear-speech effect was not found in the group, talker 2M did produce a positive clear-speech effect [$F(1, 20) = 13.277$, $p = 0.002$, $\eta^2 = 0.399$]. Together, the individual talker results are consistent with the overall patterns. A notable exception is talker 1M, who consistently weakened the statistical power of the positive clear-speech effects shown in the group data.

In sum, the style comparisons show that, across groups, clear speech benefited sibilants more in the auditory domain and non-sibilants more in the visual domain. Across fricatives, native English perceivers benefited more from clear speech in the auditory domain while non-native perceivers (particularly Koreans) benefited more in the visual domain.

2. Modality comparisons

To explore the three-way interactions that involved Modality, one-way ANOVAs were conducted with Modality as a within-subject factor for each Style, Sibilance category, and L1 group. Pairwise comparisons were conducted as *post hoc* analyses with Bonferroni adjustments to compare the modality pairs.

Results revealed a significant effect of Modality for all conditions (all $F > 52.731$, all $p < 0.001$, all $\eta^2 > 0.645$). For all L1 groups, perceptual accuracy of both conversational and clear non-sibilants and conversational sibilants

https://doi.org/10.1121/10.0001140

followed the hierarchy of higher accuracy in the AV condition than in the AO condition, and in turn than in the VO condition (AV > AO > VO), with Bonferroni-adjusted pairwise comparisons revealing that mean accuracy scores for all pairs were significantly different (all $p < 0.001$). However, for clear sibilants, the English perception showed highest accuracy in AO followed by AV and then VO (AO > AV > VO, all $p < 0.001$), while for both non-native groups, there was no significant difference between AO and AV (all $p > 0.05$), both of which had higher accuracy than VO (all $p < 0.001$) (AO, AV > VO).

Overall, the modality comparisons revealed that the AV modality was the most intelligible for all L1 groups in most conditions, with the exception of sibilants in the clear condition, where AO was more intelligible than AV for the natives or just as intelligible as AV for the non-natives.

3. Language comparisons

The data in this section are presented to compare the effects of L1 background on fricative perception, as the overall results show a significant main effect of L1 and significant interactions involving L1. One-way ANOVAs were conducted with L1 as a between-subject factor for each speech style, modality, and sibilance category. The performance of the L1 groups was further compared using pairwise comparisons as *post hoc* analyses with Tukey HSD adjustments. The statistically significant results on L1 group comparisons are presented in Table III.

The analyses revealed a significant effect of L1 in all AV and AO conditions (all $F > 4.504$, all $p < 0.015$, all $\eta^2 > 0.116$). *Post hoc* comparisons showed higher mean identification scores for English than either Mandarin and/or Korean perceivers in all of the above conditions (all $p < 0.05$). While the Mandarin and Korean groups performed on par with each other in most conditions (all $p > 0.05$), Mandarin perceivers outperformed Korean perceivers in the perception of clear sibilant AV ($p = 0.040$). In the VO condition, an L1 effect was only found with conversational (but not clear) fricative perception {Conversational, Nonsibilant: [$F(2, 69) = 4.216$, $p = 0.019$, $\eta^2 = 0.108$]; Conversational, Sibilant: [$F(2, 69) = 3.433$, $p = 0.038$, $\eta^2 = 0.090$]; Clear, Nonsibilant: [$F(2, 69) = 2.535$, $p = 0.087$, $\eta^2 = 0.068$]; Clear, Sibilant: [$F(2, 69) = 0.093$, $p = 0.911$, $\eta^2 = 0.002$]}, where English perceivers were better than Korean ($p = 0.029$), but not Mandarin ($p = 0.411$), perceivers at identifying sibilants; and they were better than Mandarin ($p = 0.019$) but not Korean ($p = 0.222$) perceivers at perceiving non-sibilants. Mandarin and Korean perceivers' performance did not differ ($p > 0.05$).

These results suggest better native than non-native perception in the AV and AO modalities across speech styles, and in the VO modality in conversational but not in clear speech. The two non-native groups performed similarly in most conditions, with only a few differences across non-native groups for sibilants and non-sibilants. It is not evident whether any differences across non-native groups can be attributed to any specific L1 effects since sibilant and non-sibilant data are pooled across place of articulation with differences in fricative inventories across the L1 groups.

B. Place accuracy

To further observe the effects of L1 and visual information on clear speech perception, the mean percent accuracy of place identification regardless of voicing was calculated for each place of articulation. Since facial information does not contain critical cues to voicing (Fisher, 1968; Jongman *et al.*, 2003), this analysis is particularly relevant for the VO modality. A four-way mixed ANOVA was conducted with place accuracy (percent correct identification across voicing) as the dependent variable, Style, Modality, and Place (labiodental, interdental, alveolar, post-alveolar) as repeated measures; and L1 as the between-subject factor. Percent correction identification of place was calculated by averaging the place identification accuracy scores obtained by each participant. That is, every participant's scores were averaged across the four talkers, three response types, and individual tokens for each style, modality, and place of articulation.

A significant main effect was found for all four factors: Style [$F(1, 69) = 49.458$, $p < 0.001$, $\eta^2 = 0.418$], Modality [$F(1, 69) = 60.527$, $p < 0.001$, $\eta^2 = 0.467$], Place [$F(1, 169.781) = 8.746$, $p < 0.001$, $\eta^2 = 0.112$], and L1 [$F(2, 111.934) = 522.117$, $p < 0.001$, $\eta^2 = 0.883$]. Consistent with the overall results, clear speech ($M = 85\%$, $SD = 7\%$) was more intelligible than conversational speech ($M = 83\%$, $SD = 6\%$). For Modality, *post hoc* pairwise comparisons with Bonferroni adjustments showed that place accuracy was highest in the AV modality ($M = 91\%$, $SD = 7\%$), followed by the VO modality ($M = 85\%$, $SD = 7\%$) and in turn by the AO modality ($M = 76\%$, $SD = 8\%$) (all $p < 0.001$). The greater place accuracy in the VO than the AO modality (which is the reverse of the overall results in Sec. III A 2) indicates that visual information was more salient than auditory information when voicing was not considered. For Place, *post hoc* analyses showed greater accuracy for

TABLE III. Summary of L1 comparisons in different speech styles (conversational, clear), Modality (AV, AO, VO) and Sibilance by native perceivers of English (E), Mandarin (M), and Korean (K). ">"represents significantly more accurate identification ($p < 0.05$).

	Non-sibilant			Sibilant		
	AV	**AO**	**VO**	**AV**	**AO**	**VO**
Conversational	E > M, K	E > M	E > M	E > M, K	E > M, K	E > K
Clear	E > M, K	E > M, K		E > M > K	E > M, K	

labiodental fricatives ($M = 87\%$, $SD = 9\%$) than both interdental fricatives ($M = 80\%$, $SD = 8\%$) ($p = 0.002$) and post-alveolar fricatives ($M = 83\%$, $SD = 11\%$) ($p = 0.016$), as well as greater accuracy for alveolar fricatives ($M = 85\%$, $SD = 10\%$) than interdental fricatives ($p = 0.005$). For L1, English perceivers ($M = 90\%$, $SD = 9\%$) performed better than both Mandarin ($M = 81\%$, $SD = 10\%$) ($p = 0.005$) and Korean perceivers ($M = 80\%$, $SD = 7\%$) ($p < 0.001$).

1. Style comparisons as a function of place of articulation

The mixed ANOVA did not find a significant four-way interaction between Style × Modality × Place × L1 [$F(9.283, 320.275) = 1.708$, $p = 0.084$, $\eta_p^2 = 0.047$], but it yielded a significant three-way interaction between Style × Modality × Place [$F(4.642, 320.275) = 9.461$, $p < 0.001$, $\eta_p^2 = 0.121$], Style × Place × L1 [$F(4.551, 157.017) = 14.042$, $p < 0.001$, $\eta_p^2 = 0.154$], and Modality × Place × L1 [$F(8.538, 294.565) = 2.334$, $p = 0.017$, $\eta_p^2 = 0.063$]. Two-way interactions were also found between Style × Place [$F(2.276, 157.017) = 29.653$, $p < 0.001$, $\eta_p^2 = 0.301$] as well as Style × Modality [$F(1.678, 115.769) = 11.369$, $p < 0.001$, $\eta_p^2 = 0.141$]. Given these key interactions, subsequent one-way ANOVAs were further conducted to examine the effects of Style with place accuracy as the dependent variable for each Modality, Place, and L1. Figure 2 presents the style comparisons for English, Mandarin, and Korean perceivers.

For English perceivers, clear speech significantly improved place identification in the VO modality for labiodental [$F(1, 23) = 52.808$, $p < 0.001$, $\eta^2 = 0.697$], interdental [$F(1, 23) = 30.029$, $p < 0.001$, $\eta^2 = 0.566$], and post-alveolar fricatives [$F(1, 23) = 10.338$, $p = 0.004$, $\eta^2 = 0.310$]. For Mandarin perceivers, positive clear-speech effects were found in the VO modality for interdental fricatives [$F(1, 17) = 13.240$, $p = 0.002$, $\eta^2 = 0.438$] and post-alveolar fricatives [$F(1, 17) = 6.666$, $p = 0.019$, $\eta^2 = 0.282$], and in the AO modality for post-alveolar fricatives [$F(1, 17) = 12.072$, $p = 0.003$, $\eta^2 = 0.415$]. For Korean perceivers, clear speech had a positive effect in the VO modality for labiodental fricatives [$F(1, 29) = 54.016$, $p < 0.001$, $\eta^2 = 0.651$], interdental fricatives [$F(1, 29) = 111.520$, $p < 0.001$, $\eta^2 = 0.794$], and post-alveolar fricatives [$F(1, 29) = 48.409$, $p < 0.001$, $\eta^2 = 0.625$]; in the AV modality for labiodental fricatives [$F(1, 29) = 9.850$, $p = 0.004$, $\eta^2 = 0.254$] and post-alveolar fricatives [$F(1, 29) = 14.588$, $p = 0.001$, $\eta^2 = 0.335$], and in the AO modality for post-alveolar fricatives [$F(1, 29) = 56.386$, $p < 0.001$, $\eta^2 = 0.660$]. However, Korean perceivers showed a negative clear-speech effect for alveolar fricatives across all three modalities {AV: [$F(1, 29) = 15.474$, $p < 0.001$, $\eta^2 = 0.348$]; AO: [$F(1,$

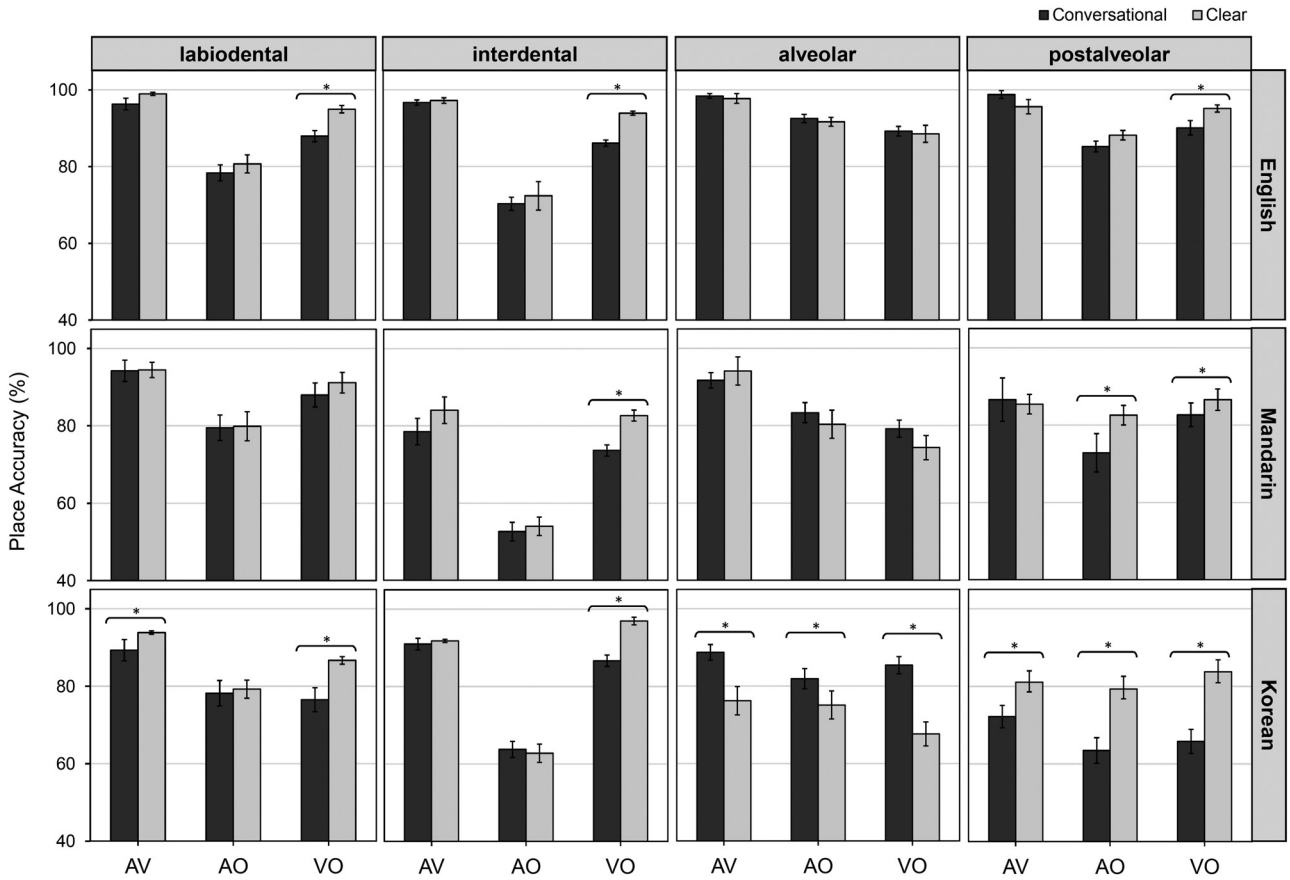

FIG. 2. Style (conversational, clear) comparisons of mean percent place accuracy for each Place of articulation and Modality Modality (AV, AO, VO) by native perceivers of English (top panel), Mandarin (middle panel), and Korean (bottom panel). "*" indicates statistically significant differences ($p < 0.05$). Error bars indicate standard error.

29) = 5.035, $p = 0.033$, $\eta^2 = 0.148$]; VO: [$F(1, 29) = 29.890$, $p < 0.001$, $\eta^2 = 0.508$]}.

In sum, the current analysis of perception of place of articulation showed that in the VO modality, all places of articulation except alveolar benefited from clear speech when voicing is not considered. This suggests that the lack of positive clear-speech effects found with non-sibilant fricatives in VO conditions for English perceivers in the previous section was largely due to misperception of voicing, rather than place. Furthermore, these place results may also indicate that the negative clear-speech effects found with sibilant fricatives in the AV modality for English and Korean perceivers in Sec. III A 1 were most likely due to voicing errors, and additionally for Korean perceivers, due to place errors of alveolar fricatives as the place analysis showed that clear speech significantly impeded Koreans' perception of alveolar fricatives while it seemed to benefit post-alveolar fricatives. It is also possible that conversational speech may be as intelligible as clear speech, given that only one talker (1M) produced a negative clear-speech effect. The visual benefits exhibited by post-alveolar fricatives could be due to the lip-rounding and protrusion gestures involved in their production, which make post-alveolar fricatives more visually salient than alveolar fricatives (Ladefoged and Maddieson, 1996). Together, the results indicate that the visual cues associated with a clear-speech style are beneficial for all places of articulation that are visually salient.

2. L1 group comparisons as a function of place of articulation

The place accuracy data revealed differences between L1 groups in the perception of conversational and clear fricatives as a function of place of articulation. Subsequent one-way ANOVAs were conducted with L1 as a between-subject factor and place accuracy as the dependent variable for each Style, Modality, and Place, each followed by *post hoc* analyses with Tukey HSD adjustments. We focused on L1 comparisons in the VO modality, since collapsing across voicing is only appropriate for this condition. Table IV displays statistically significant results on the L1 comparisons of place perception accuracy for conversational and clear styles in the VO modality.

For labiodental fricatives in the VO modality, a significant effect of L1 was observed in the conversational style [$F(2, 69) = 5.894$, $p = 0.004$, $\eta^2 = 0.146$] but not in the clear style [$F(2, 69) = 2.509$, $p = 0.089$, $\eta^2 = 0.068$]. *Post hoc* comparisons show that Mandarin perceivers performed on par with English perceivers ($p = 0.989$), both being better than Korean perceivers (all $p < 0.022$). For interdental fricatives in the VO modality, there was a significant effect of L1 on the perception of both conversational [$F(2, 69) = 5.236$, $p = 0.008$, $\eta^2 = 0.132$] and clear [$F(2, 69) = 5.626$, $p = 0.005$, $\eta^2 = 0.140$] tokens, with *post hoc* comparisons showing English ($p = 0.023$) and Korean ($p = 0.011$) perceivers outperforming Mandarin perceivers. For alveolar fricatives in the VO modality, an L1 effect was observed for clear speech [$F(2, 69) = 11.549$, $p < 0.001$, $\eta^2 = 0.251$] with *post hoc* comparisons showing English perceivers outperforming both Mandarin ($p = 0.017$) and Korean ($p < 0.001$) groups, but no L1 effect was observed for conversational speech [$F(2, 69) = 3.118$, $p = 0.051$, $\eta^2 = 0.083$]. For post-alveolar fricatives in the VO modality, an L1 effect was observed in the conversational style [$F(2, 69) = 16.368$, $p < 0.001$, $\eta^2 = 0.322$] with *post hoc* comparisons showing English ($p < 0.001$) and Mandarin ($p = 0.002$) perceivers outperforming Korean perceivers, while in clear speech, the significant L1 effect [$F(2, 69) = 4.911$, $p = 0.010$, $\eta^2 = 0.125$] with *post hoc* comparisons showing English ($p = 0.009$) but not Mandarin ($p = 0.961$) perceivers outperforming Korean perceivers.

Overall, L1 group comparisons of the place-only results show that, in conversational speech presented in the VO modality, perception of labiodental fricatives and post-alveolar fricatives was poorer for the Korean group, whose L1 does not contain these fricatives, compared to the Mandarin perceivers familiar with these fricatives in their L1. Conversely, in the clear condition, the Korean group's performance was on par with the Mandarin group. On the other hand, for interdental fricatives, Korean perceivers outperformed the Mandarin perceivers in both the conversational and clear conditions, although neither language has interdental fricatives.

TABLE IV. Summary of L1 group comparisons in different speech styles (conversational, clear), Modality (AV, AO, VO) and Place (labiodental, interdental, alveolar, postalveolar) by native perceivers of English (E), Mandarin (M), and Korean (K). ">" represents significantly more accurate identification ($p < 0.05$).

	Labiodental	Interdental	Alveolar	Postalveolar
Conversational Speech VO	E, M > K	E, K > M		E, M > K
Clear speech VO		E, K > M	E > M, K	E > K

C. Summary of results

Taken as a whole, the results demonstrated significant clear-speech effects in the multi-modal identification of English fricatives as a function of AV saliency and L1 background. Overall, perception of the auditorily salient sibilants benefited more from clear speech in the auditory domain, while perception of the visually salient non-sibilant and post-alveolar fricatives benefited more from clear speech in the visual domain. Native English perceivers demonstrated clear-speech benefits across modality and sibilance conditions, except for alveolar perception in the visual conditions. Mandarin and Korean perceivers showed clear-speech gain in the auditory perception of sibilants and the visual perception of all but the alveolar fricatives. Both Mandarin and Korean perceivers' visual perception of their respective non-L1 fricatives improved in clear speech, with the Korean perceivers showing a greater clear-speech gain than the Mandarin perceivers.

IV. DISCUSSION

Extending the results of previous auditory-based studies (Kabak and Maniwa, 2007; Maniwa *et al.*, 2008), the present study made a novel contribution by exploring the effects of clear speech on fricative perception in domains that include visual information, thereby acknowledging the inherent salience differences of auditory and visual cues associated with English fricatives, and their possible roles in clear speech to enhance native and non-native perception.

A. Style effects on AV saliency in native English perception

The first research question was whether clearly-produced fricatives are more intelligible when visual information is available, and if so, how auditory and visual input in clear speech are weighted to enhance intelligibility.

The current results relating to visual effects on the clear-speech intelligibility advantage in native English fricative perception agree with previous findings of clear-speech benefits for segmental perception that involve visual input (e.g., Gagné *et al.*, 1994; Gagné *et al.*, 2002). In particular, English perceivers benefit from clear speech in the AV modality when presented with non-sibilant but not sibilant fricatives. Moreover, although the overall results do not show significant clear-speech effects in the VO modality, the place-only analysis reveals that clear (relative to conversational) speech aids the perception of labiodental, interdental, and post-alveolar fricatives. This suggests that, when disregarding errors due to voicing (which is difficult to determine from facial-only information, e.g., Fisher, 1968; Jongman *et al.*, 2003), clear speech benefits the non-sibilant fricatives with more anterior (and thus more visible) places of articulation, as well as post-alveolar fricatives which involve lip protrusion (Flemming, 2002; Ladefoged and Maddieson, 1996) and are thus visually marked (Tang *et al.*, 2015; Traunmüller and Öhrström, 2007). Together, these results demonstrate the role of visual salience since clear speech benefits the fricatives with visually distinct cues regardless of place of articulation and sibilance.

Clear-speech benefits are also found in the AO modality for both sibilant and non-sibilant perception, corroborating the previous auditory-based findings of a clear-speech advantage for native English fricative perception (Maniwa *et al.*, 2008). Furthermore, the modality comparisons also reveal effects of sibilance, in that clear speech enhanced the intelligibility of sibilant fricatives in the AO compared to the AV modality, but not of non-sibilant fricatives, indicating the saliency of auditory information in clear sibilant perception. This clear-speech benefit for the perception of sibilant fricatives is in line with the finding that sibilant fricatives that are auditorily more salient benefit more from clear speech than non-sibilants (Maniwa *et al.*, 2008). The positive clear-speech effects with the non-sibilants may have been due to the improved perception of voicing distinctions which have presumably been enhanced in clear speech as the result of greater lengthening for voiceless than voiced fricatives (Maniwa *et al.*, 2009).[2]

Bringing together the results from perception in the auditory and visual modalities, the findings support our hypothesis of improved intelligibility in clear speech as a function of AV weighting. In particular, in the clear relative to the conversational speech condition, perceivers showed enhanced attunement to the auditory information in perceiving sibilants (which are acoustically salient), whereas they showed enhanced attunement to the visual information in perceiving non-sibilant fricatives (which are visually salient). It should be noted that perceivers already follow these patterns in conversational speech, as shown by more accurate perception in the AO than the VO modality for the auditorily salient sibilants (from the overall results), but more accurate perception in the VO than the AO modality for the visually salient non-sibilant and post-alveolar fricatives (from the place-only results). These patterns are consistent with the previous findings on fricative perception where articulatorily more visible non-sibilants (e.g., labial, labiodental) benefit more from visual cues, while the perception of acoustically more salient sibilants (e.g., alveolar, post-alveolar) mostly relies on auditory input (Hazan *et al.*, 2006; Iba *et al.*, 2004; Jongman *et al.*, 2003). The findings are consistent with the claim that the degree of visual reliance is inversely linked to the auditory prominence of a sound (Chen and Massaro, 2004).

Thus, the current results extend these previous findings of AV weighting patterns to clear speech, indicating that clear-speech modifications that enhance the inherently prominent features of a speech signal are more likely to be utilized to further improve intelligibility. Acoustic studies have consistently shown that clear-speech modifications enhance phoneme-intrinsic properties. For example, in clear relative to conversational speech, duration of the inherently longer voiceless fricatives (Jongman *et al.*, 2000) is increased more than that of the inherently shorter voiced fricatives (Maniwa *et al.*, 2009). Similarly, for vowels, it has also been found that clear-speech modifications are aligned with vowel-intrinsic characteristics, where the inherently long tense vowels are further lengthened in clear speech and the spectrally more variable lax vowels become more dynamic in clear speech (Leung *et al.*, 2016). Therefore, it appears that clear-speech modifications strengthen the intrinsic cues that are characteristic of a sound category to make the sound more distinctive, and such modifications can be adopted in perception to enhance identification. This further implies that clear speech involves modifications that enhance categorical distinctions in both auditory and visual streams of speech in a complementary manner to achieve optimal perceptual benefits.

However, it is possible that enhancement of overall signal saliency (beyond sound-specific, category-distinctive modifications) in clear speech also contributes to perceptual gains. For example, exaggerated articulation of visually salient fricatives may draw more attention to the visual input and therefore further increase intelligibility gains. As

discussed, previous studies have shown a compensatory channel weighting effect where perceivers rely more on information from an alternate modality when the other modality was degraded or less salient (Chen and Massaro, 2004; Gagné *et al.*, 2002; Hazan *et al.*, 2010; Traunmüller and Öhrström, 2007; Van Engen *et al.*, 2014). Thus, the current saliency-based perceptual patterns may also have resulted from greater attention to the (acoustically or visually) more prominent channel.

The issue regarding relative contributions of sound-specific, category-defining versus generic, signal-enhancing clear-speech modifications to intelligibility may be further elucidated by comparing the native results with non-native patterns because such comparisons help unravel the extent to which clear-speech gains in perception are attributable to language-specific aspects as a function of clear-speech features.

B. Style effects on L1 background

Comparing native English patterns with those from the Mandarin and Korean perceivers enables us to address the second question of the current study, which concerns how language-specific aspects of the L1 fricative systems affect multi-modal fricative perception. Similar to native English perceivers, both Mandarin and Korean perceivers demonstrated clear-speech gains in the VO perception of the non-sibilant and post-alveolar fricatives, and in the AO perception of sibilants; although, in contrast to the native English perceivers, their non-sibilant perception in the AO modality did not benefit from clear speech. In particular, the non-native perceivers' visual perception of their respective non-L1 fricatives (interdental fricatives for Mandarin, and labiodental fricatives, interdental fricatives, and post-alveolar fricatives for Korean) improved in clear speech, with the Korean perceivers' identification of place distinctions between the non-sibilants being on par with that of the native English perceivers.

In the auditory domain, the results of clear-speech facilitation in the perception of English sibilants confirmed our prediction, agreeing with the previous finding of greater clear-speech benefits for the (auditorily salient) sibilants than non-sibilants for native and non-native perceivers alike (Kabak and Maniwa, 2007; Maniwa *et al.*, 2008). However, unlike the native perceivers, non-native perceivers were unable to benefit from clear speech in the perception of non-sibilant fricatives, some of which include non-L1 fricative sounds. These patterns may find some support from Fenwick *et al.* (2015), where clear speech only benefited non-native perception of acoustically more salient (voicing) contrasts but not the less salient place-of-articulation distinctions. Since non-sibilants are not as acoustically salient as sibilant fricatives (Behrens and Blumstein, 1988; Jongman *et al.*, 2000; Strevens, 1960), the lack of clear-speech benefits found in the non-native perception of non-sibilants may well be a result of their relatively less robust acoustic properties. However, the German perceivers in Kabak and Maniwa (2007) as well as the native English perceivers in the current study did obtain a clear-speech advantage with non-sibilants. Given the absence of voiced fricatives in both Mandarin and Korean phonetic inventories (Table I), it could be the case that the Mandarin and Korean perceivers in this study could not take advantage of the more enhanced voicing distinctions in clear speech as English and German perceivers did. Overall, the AO results suggest that clear speech primarily aided the perception of the acoustically salient sibilants, but not the acoustically non-salient non-sibilants, possibly attributable to L1 influence. The effects of L1 influence are more evident from the results in the visual domain, given that the English non-sibilants as well as post-alveolar fricatives are visually salient but involve visual cues (places of articulation) absent in Mandarin and Korean.

In the visual domain, our prediction for visual clear-speech effects as a function of visual saliency was supported in that, similar to the native English perceivers, both non-native groups benefited from clearly-articulated visual information in the perception of the visually salient non-sibilant and post-alveolar fricatives, but not the less visually salient alveolar fricatives. Thus, the current results extend the previous findings on clear speech in the auditory domain (e.g., Bradlow and Bent, 2002; Kabak and Maniwa, 2007; Rogers *et al.*, 2010), showing that visual clear-speech information can benefit native and non-native perception alike. Language-independent factors may contribute to this process. For example, it has been claimed that the ability to use visual cues may be language-universal (Burnham *et al.*, 2015), and non-native perceivers tend to rely more heavily on visual input when both auditory and visual inputs are available (Chen and Hazan, 2007; Hannah *et al.*, 2017; Sekiyama and Tohkura, 1993; Wang *et al.*, 2008, 2009). Presumably, if more attention is focused on the visual domain, exaggerated visual articulatory movements in clear speech may more likely enhance perception. However, this does not necessarily imply that perception is driven by an overall enhanced saliency of the signal in clear speech. Given the current finding of greater visual clear-speech effects for the fricatives with inherently salient visual cues, it is conceivable that clear speech that further strengthens these cues makes speech categories more distinct, thereby facilitating perception. Hence, effective clear-speech cues correspond to the phoneme-intrinsic cues that characterize sound categories in a language.

Indeed, language-specific factors are at play in the current results, as shown by how clear-speech perception patterns differ depending on L1 fricative inventories. Specifically, in the visual perception of conversational fricatives, Mandarin perceivers, whose L1 contains labiodental and post-alveolar fricatives, perform better than Korean perceivers, who lack these places of articulation in their L1 (Table IV). These results are aligned with the previous claims that non-native perceivers benefit more from visual information in identifying the fricatives existent in their L1 than those unfamiliar to them (e.g., Hazan *et al.*, 2006;

Wang *et al.*, 2009). However, in clear speech, visual cue enhancements are beneficial for the Korean perceivers to the extent that they perform on par with the Mandarin group in the perception of these fricatives. Similarly, the Korean perceivers show improved AV integration in clear speech, as evidenced by their better performance in the perception of labiodental fricatives and post-alveolar fricatives in clear compared to conversational speech. Language-specific factors are also observed in the perception patterns of interdental fricatives not available in Mandarin and Korean perceivers' L1s. In this case, the Korean group is on par with the native English group, outperforming the Mandarin group. These results replicate previously observed visual non-sibilant perception patterns by Mandarin and Korean perceivers (Wang *et al.*, 2009). As discussed previously, perception of the interdental fricatives can be more difficult for the Mandarin perceivers than for the Koreans, since the existence of labiodental fricatives makes the Mandarin perceptual space more crowded compared to Korean. Despite the differences, Mandarin perceivers benefit from clear speech to the same degree as the Korean and native English perceivers. Taken together, perception of visually salient fricatives may be more susceptible to L1 influence but may also more likely gain from clear speech, as visual information intrinsic to these fricatives can be more informative to categorization when enhanced.

Patterns of alveolar perception consistently suggest how saliency and L1 factors influence clear-speech effects, given the lack of a clear-speech gain for the English and Mandarin perceivers and the negative clear-speech effect for the Korean perceivers. It is likely that perceivers do not benefit from clear speech for the visual perception of alveolar fricatives because they are more acoustically robust and less visually marked (Jongman *et al.*, 2000; Jongman *et al.*, 2003; Maniwa *et al.*, 2009; Wang *et al.*, 2008), and their auditory prominence may reduce reliance on visual cues (Chen and Massaro, 2004; Fenwick *et al.*, 2015; Wang *et al.*, 2009). These patterns support our prediction that visual saliency determines clear-speech benefits in the visual domain. Finally, the Korean perceivers' decreased sensitivity to clear (relative to conversational) alveolar fricatives may have been due to interference from their L1. A closer inspection of the perception confusion patterns shows that Korean perceivers confused alveolar fricatives with post-alveolar fricatives (and vice versa). Although Korean does not have a separate phonemic category for post-alveolar fricatives, its alveolar sibilant /s/ may be realized as a palatalized allophone, which involves a more posterior articulation similar to post-alveolar in English (Kong *et al.*, 2014; Schmidt, 1996). It has been claimed that English alveolar and post-alveolar fricatives are taken to be allophones by Korean speakers (Eckman and Iverson, 1997). We speculate that the phonetic variation of English alveolar fricatives caused by speaking clearly may have made them more confusable, as the variation resembles the allophonic differences in Korean.

V. GENERAL DISCUSSION

Bridging the findings across input modalities and native groups provides a comprehensive picture in support of our predictions that clear-speech benefits in multi-modal fricative perception are affected by the relative auditory and visual saliency of a fricative sound, as well as by the "nativeness" of these auditory and visual cues. Indeed, the factors of signal saliency and nativeness are central to the theoretical accounts of clear-speech functions regarding the extent to which clear-speech effects are governed by global saliency-enhancing (signal-based) or language-specific category-enhancing (code-based) strategies (Bradlow and Bent, 2002; Smiljanić and Bradlow, 2009).

The current results indicate that both types of strategies are at play in multi-modal perception of English fricatives. At first glance, results demonstrate overall AV weighting patterns as a function of AV saliency, in that both native and non-native perceivers enjoy greater clear-speech gain for sibilants in the auditory domain and for non-sibilants in the visual domain. This is conceivably due to the fact that English sibilants are more acoustically prominent while non-sibilants are more prominent visually. When the overall signal saliency is enhanced in clear speech, perceivers pay more attention to the acoustically or visually more prominent channel, thus achieving greater clear-speech gain on the basis of the saliency of the physical sound input. However, if clear-speech strategies are only attributable to global factors in terms of the saliency of the input, we should expect similar clear-speech effects for all the fricatives of the same "saliency level" across native groups. This is apparently not always the case as the present results show different clear-speech effects due to differences in L1 background. Particularly, the present results reveal that the perception of visually salient fricatives tends to be more susceptible to L1 influence but can also benefit more from clear speech (e.g., Korean perceivers' greater visual benefits from conversational to clear speech for the labiodental fricatives and post-alveolar fricatives non-existent in Korean; and Mandarin perceivers' greater clear-speech gain for the non-L1 interdental fricatives than the L1-like labiodental fricatives). These patterns indicate that non-native perceivers are able to transcend the effects of signal saliency on visual clear-speech gains by effectively adopting code-based cues specific to category distinctions.

Together, these findings broaden the scope of research on clear speech by drawing evidence from the visual as well as auditory domains, demonstrating interactive effects of signal saliency and language specificity underlying clear-speech mechanisms. Such findings have significant implications for the auditory-based theoretical accounts of clear speech (Lindblom, 1990; Ohala, 1995; Smiljanić and Bradlow, 2009). As discussed previously, to facilitate perception, the extent of clear-speech modifications, whether signal- or code-based, must retain phoneme-intrinsic characteristics and remain within-category in order to maintain categorical phonemic distinctions (Moon and Lindblom, 1994; Ohala, 1995). Indeed,

previous speech production research has revealed that clear-speech modifications strengthen the intrinsic cues that define a sound category to make the sound more distinctive (e.g., greater lengthening of the intrinsically long English tense vowels and voiceless fricatives in clear speech, relative to their intrinsically short lax and voiced counterparts, respectively; Leung *et al.*, 2016; Maniwa *et al.*, 2009). The current results demonstrate these patterns from the perceptual perspective, with more prominent visual clear-speech effects on the (non-sibilant) fricatives bearing inherently salient articulatory cues, and more prominent auditory clear-speech effects on the (sibilant) fricatives bearing inherently salient acoustic cues. Thus, these findings suggest that clear-speech gains in perception involve interrelated coordination of enhanced perceptual attributes that are congruent with the phoneme-intrinsic cues characterizing the corresponding sound categories.

VI. CONCLUDING REMARKS

Taken as a whole, findings of this study imply that multi-modal clear-speech perception strategies should involve effective adaptation to multiple factors in a complementary manner to achieve optimal perceptual benefits, including balancing saliency-enhancing and category-enhancing modifications and weighting auditory and visual streams of input. These strategies have practical implications for face-to-face communication, suggesting that efforts to disambiguate speech sounds in adverse perceptual conditions should take into consideration such factors as language background and auditory/visual salience of the speech signal. Additionally, the current results also show evidence of individual talker influence. Indeed, previous research has shown that individual talker characteristics may contribute to different levels of intelligibility in clear speech (Bradlow, 2002; Smiljanić and Bradlow, 2009) as well as to different degrees of visual weighting in AV speech perception (Hazan *et al.*, 2010). Likewise, intelligibility patterns may also depend on individual perceiver strategies and aptitudes (Chandrasekaran *et al.*, 2010; Hazan *et al.*, 2010; Ingvalson *et al.*, 2013; Smiljanić and Bradlow, 2009). Future research needs to take into account effects of individual talker and perceiver strategies on clear-speech perception in auditory and visual domains.

Ultimately, this line of research contributes to a better understanding of how multi-modal speech enhancement principles can be applied in different communicative contexts to optimally serve different needs.

ACKNOWLEDGMENTS

This study was funded by a research grant from the Social Sciences and Humanities Research Council of Canada (SSHRC Insight Grant No. 435-2012-1641). We thank Zoe Buekers, Yujin Han, Beverly Hannah, Eleanor Hendriks, Michelle Kim Le, Quince Sholberg, and Jennifer Williams from the Language and Brain Lab at SFU for their assistance with audio-video recording, stimulus development, data collection, and analysis. We also thank Haoyao Ruan from SFU Department of Statistics and Actuarial Science for assistance with statistical analysis. Portions of this work were presented at the 174th Meeting of the Acoustical Society of America, New Orleans, LA, USA, December 2017.

[1]In the context of the present study, the terms "conversational speech" and "clear speech" are used following previous clear-speech studies (e.g., Ferguson and Kewley-Port, 2002; Helfer, 1997; Maniwa *et al.*, 2008). These two terms refer to the contrasting speech styles resulting from instructions to talkers to speak "normally" (i.e., in the manner used in a normal conversation) and "clearly," respectively. Thus, "conversational speech" here may also be described as "plain speech," as has also been used previously (e.g., Leung *et al.*, 2016; Tang *et al.*, 2015).

[2]Indeed, the fricative productions in the current study show longer durations for voiceless ($M = 290$ ms, 43% of the syllable duration) than voiced ($M = 250$ ms, 38% of the syllable duration) non-sibilants in clear speech, whereas their durations in conversational speech do not differ ($M = 150$ ms for both voiceless and voiced, 34% and 32% of the syllable duration, respectively). This is presumably why the identification based on voicing accuracy was better in clear ($M = 87\%$) than in conversational ($M = 82\%$) conditions [$F(1,23) = 10.070$, $p = 0.004$, $\eta_p^2 = 0.305$], particularly in identifying the voicing of /f/, with 94% accuracy in clear compared to 87% in conversational speech [$F(1,23) = 17.955$, $p < 0.001$, $\eta_p^2 = 0.438$].

Behrens, S. J., and Blumstein, S. E. (**1988**). "Acoustic characteristics of English voiceless fricatives—A descriptive analysis," J. Phon. **16**, 295–298.
Bradlow, A. R. (**2002**). "Confluent talker- and listener-oriented forces in clear speech production," in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin).
Bradlow, A. R., and Alexander, J. A. (**2007**). "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners," J. Acoust. Soc. Am. **121**, 2339–2349.
Bradlow, A. R., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.
Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for children with learning disabilities," J. Speech Lang. Hear. Res. **46**, 80–97.
Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., Rattanasone, N. X., Schwarz, I. C., and Webster, D. (**2015**). "Universality and language-specific experience in the perception of lexical tone and pitch," Appl. Psycholinguist. **36**, 1459–1491.
Chandrasekaran, B., Sampath, P., and Wong, P. C. (**2010**). "Individual variability in cue-weighting and lexical tone learning," J. Acoust. Soc. Am. **128**, 456–465.
Chen, T., and Massaro, D. (**2004**). "Mandarin speech perception by ear and eye follows a universal principle," Atten. Percept. Psychophys. **66**, 820–836.
Chen, Y., and Hazan, V. (**2007**). "Language effects on the degree of visual influence in audiovisual speech perception," in *Proceedings of the 16th International Congress of Phonetic Sciences*, August 6–10, Saarbrueken, Germany, pp. 6–10.
Cheon, S. Y., and Anderson, V. B. (**2008**). "Acoustic and perceptual similarities between English and Korean sibilants: Implications for second language acquisition," Korean Ling. **14**, 41–64.
de Gelder, B., and Vroomen, J. (**1992**). "Auditory and visual speech perception in alphabetic and non-alphabetic Chinese-Dutch bilinguals," in *Advances in Psychology*, edited by R. J. Harris (Elsevier. Amsterdam, The Netherlands), Vol. 83, pp. 413-426.
Eckman, F. R., and Iverson, G. K. (**1997**). "Structure preservation in interlanguage phonology," in *Focus on Phonological Acquisition*, edited by S. J. Hannahs and M. Young-Scholten (John Benjamins, Amsterdam, The Netherlands), pp. 183–208.
Fenwick, S., Davis, C., Best, C. T., and Tyler, M. D. (**2015**). "The effect of modality and speaking style on the discrimination of non-native phonological and phonetic contrasts in noise," in *Proceedings of the 1st Joint Conference on Facial Analysis, Animation, and Auditory-Visual Speech Processing*, September 11–13, Vienna, Austria, pp. 67–72.

Ferguson, S. H. (**2004**). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," J. Acoust. Soc. Am. **116**, 2365–2373.
Ferguson, S. H., and Kewley-Port, D. (**2002**). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**, 259–271.
Fisher, C. G. (**1968**). "Confusions among visually perceived consonants," J. Speech Hear. Res. **11**, 796–804.
Flemming, E. S. (**2002**). *Auditory Representations in Phonology* (Routledge, New York).
Gagné, J. P., Masterson, V., Munhall, K. G., Bilida, N., and Querengesser, C. (**1994**). "Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech," J. Acad. Rehabil. Audiol. **27**, 135–158.
Gagné, J. P., Querengesser, C., Folkeard, P., Munhall, K. G., and Masterson, V. M. (**1995**). "Auditory, visual, and audiovisual speech-intelligibility for sentence-length stimuli—An investigation of conversational and clear speech," Volta Rev. **97**, 33–51.
Gagné, J. P., Rochette, A. J., and Charest, M. (**2002**). "Auditory, visual and audiovisual clear speech," Speech Commun. **37**, 213–230.
Granlund, S., Hazan, V., and Baker, R. (**2012**). "An acoustic–phonetic comparison of the clear speaking styles of Finnish–English late bilinguals," J. Phon. **40**, 509–520.
Hannah, B., Wang, Y., Jongman, A., Sereno, J. A., Cao, J., and Nie, Y. (**2017**). "Cross-modal association between auditory and visuospatial information in Mandarin tone perception in noise by native and non-native perceivers," Front. Psychol. **8**, 1–15.
Hazan, V., Kim, J., and Chen, Y. (**2010**). "Audiovisual perception in adverse conditions: Language, speaker and listener effects," Speech Commun. **52**, 996–1009.
Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., and Chunge, H. (**2006**). "The use of visual cues in the perception of non-native consonant contrasts," J. Acoust. Soc. Am. **119**, 1740–1751.
Helfer, K. S. (**1997**). "Auditory and auditory-visual Perception of clear and conversational Speech," J. Speech Lang. Hear. Res. **40**, 432–443.
Iba, M. (**2005**). "Perceptual training and the production of English consonants by Japanese learners," Speech Commun. **47**, 29–40.
Iba, M., Sennema, A., Hazan, V., and Faulkner, A. (**2004**). "Use of visual cues in the perception of a labial/labiodental contrast by Spanish-L1 and Japanese-L1 learners of English," in *Proceedings of Interspeech 2004*, October 4–8, Jeju Island, Korea.
Ingvalson, E. M., Barr, A. M., and Wong, P. C. (**2013**). "Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning," J. Speech Lang. Hear. Res. **56**, 1092–4388.
Jongman, A., Wang, Y., and Kim, B. H. (**2003**). "Contributions of semantic and facial information to perception of nonsibilant fricatives," J. Speech Lang. Hear. Res. **46**, 1367–1377.
Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives: I. Static cues," J. Acoust. Soc. Am. **108**, 1252–1263.
Kabak, B., and Maniwa, K. (**2007**). "L2 perception of English fricatives in clear and conversational speech: The role of phonemic, phonetic, and acoustic factors," in *Proceedings of the 16th International Congress of Phonetic Sciences*, August 6–10, Saarbrucken, Germany, pp. 781-784.
Kawase, S., Hannah, B., and Wang, Y. (**2014**). "The influence of visual speech information onthe intelligibility of English consonants produced by non-native speakers," J. Acoust. Soc. Am. **136**, 1352–1362.
Kim, D. J. (**1972**). "A contrastive study of English and Korean phonology," Lang. Teach. **5**, 1–36.
Kong, E. J., Kang, S., and Seo, M. (**2014**). "Gender difference in the affricate productions of young Seoul Korean speakers," J. Acoust. Soc. Am. **136**, EL329–EL335.
Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Oxford, UK).
Ladefoged, P., and Wu, Z. (**1984**). "Places of articulation: An investigation of Pekingese fricatives and affricates," J. Phon. **12**, 267–278.
Lander, K., and Capek, C. (**2013**). "Investigating the impact of lip visibility and talking style on speechreading performance," Speech Commun. **55**, 600–605.
Lee, H., and Jongman, A. (**2016**). "A diachronic investigation of the vowels and fricatives in Korean: An acoustic comparison of the Seoul and South Kyungsang dialects," J. Int. Phon. Assoc. **46**, 157–184.
Leung, K. W., Jongman, A., Wang, Y., and Sereno, J. A. (**2016**). "Acoustic characteristics of clearly spoken English tense and lax vowels," J. Acoust. Soc. Am. **140**, 45–58.
Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht, the Netherlands), pp. 403–439.
Maniwa, K., Jongman, A., and Wade, T. (**2008**). "Perception of clear fricatives by normal-hearing and simulated hearing-impaired listeners," J. Acoust. Soc. Am. **123**, 1114–1125.
Maniwa, K., Jongman, A., and Wade, T. (**2009**). "Acoustic characteristics of clearly spoken English fricatives," J. Acoust. Soc. Am. **125**, 3962–3973.
Moon, S., and Lindblom, B. (**1994**). "Interaction between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**, 40–55.
Ohala, J. J. (**1995**). "Clear speech does not exaggerate phonemic contrast," in *Proceedings of the Fourth European Conference on Speech Communication and Technology*, September 18–21, Madrid, Spain, pp. 18–21.
Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581–1592.
Perception Research Systems (**2007**). "Paradigm stimulus presentation," http://www.paradigmexperiments.com (Last viewed 4/17/2020).
Robert-Ribes, J., Schwartz, J. L., Lallouache, T., and Escudier, P. (**1998**). "Complementarity and synergy in bimodal speech: Auditory, visual, and audio-1025 visual identification of French oral vowels in noise," J. Acoust. Soc. Am. **103**, 3677–3689.
Rogers, C. L., DeMasi, T. M., and Krause, J. C. (**2010**). "Conversational and clear speech intelligibility of /bVd/ syllables produced by native and non-native English speakers," J. Acoust. Soc. Am. **128**, 410–423.
Schmidt, A. M. (**1996**). "Cross-language identification of consonants. Part 1. Korean perception of English," J. Acoust. Soc. Am. **99**, 3201–3211.
Schum, D. J. (**1996**). "Intelligibility of clear and conversational speech of young and elderly talkers," J. Am. Acad. Audiol. **7**, 212–218.
Sekiyama, K., and Tohkura, Y. (**1993**). "Inter-language differences in the influence of visual cues in speech perception," J. Phon. **21**, 427–444.
Smiljanić, R., and Bradlow, A. R. (**2009**). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," Lang. Ling. Compass **3**, 236–264.
Smiljanić, R., and Bradlow, A. R. (**2011**). "Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness," J. Acoust. Soc. Am. **130**, 4020–4031.
Strevens, P. (**1960**). "Spectra of fricative noise in human speech," Lang. Speech **3**, 32–49.
Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**, 917–928.
Svantesson, J. O. (**1986**). "Acoustic analysis of Chinese fricatives and affricates," J. Chin. Ling. **14**, 53–70.
Tang, L. Y. W., Hannah, B., Jongman, A., Sereno, J., Wang, Y., and Hamarneh, G. (**2015**). "Examining visible articulatory features in clear and plain speech," Speech Commun. **75**, 1–13.
Traunmüller, H., and Öhrström, N. (**2007**). "Audiovisual perception of openness and lip rounding in front vowels," J. Phon. **35**, 244–258.
Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," J. Speech Hear. Res. **39**, 494–509.
Van Engen, K. J., Phelps, J. E. B., Smiljanić, R., and Chandrasekaran, B. (**2014**). "Enhancing speech intelligibility: Interactions among context, modality, speech style, and masker," J. Speech Lang. Hear. Res. **57**, 1908–1918.
Wang, Y., Behne, D. M., and Jiang, H. (**2008**). "Linguistic experience and audio-visual perception of non-native fricatives," J. Acoust. Soc. Am. **124**, 1716–1726.
Wang, Y., Behne, D. M., and Jiang, H. (**2009**). "Influence of native language phonetic system on audio-visual speech perception," J. Phon. **37**, 344–356.