Pho

been shown with Spanish perceivers who can correctly perceive the nonnative /v/ due to the existence of /f/ in their L1 (Hazan et al., 2006), possibly because voicing is not as visually distinctive as other features such as place of articulation. Indeed, previous studies indicate that the difficulty in L2 visual speech perception may lie in the places of articulation that are not used in the L1 (Hardison, 1999; Hazan et al., 2006; Werker et al., 1992). However, while most research compares the visual perception of the place of articulation of L2 consonant contrasts, those contrasts (e.g., /b/ vs. /v/) are often spread across manners of articulation (e.g., Hardison, 2005b; Hazan et al., 2006; Sekiyama & Tohkura, 1993; Werker et al., 1992). Since manner of articulation, as well as place of articulation, may affect visual speech perception (Faulkner & Rosen, 1999; Green & Kuhl, 1989), an extension for further research is to control for these differences in order to further characterize the relationships of L1 and L2 visual categories and thus address L2 auditory–visual learning patterns.

1.2. The current study

This study examines nonnative auditory–visual perception of English fricatives differing in three visually distinct places of articulation: labiodental, interdental, and alveo-

English majors. They came to Canada after age 18 years and had been there for a relatively short length of time (1–4 years), but never resided in any other English speaking environments prior to their arrival in Canada. They reported using and being exposed to their native Korean or Mandarin as their more dominant language since their arrival. All three groups of participants reported having normal or corrected vision and no known history of speech and hearing impairments. They were compensated for their participation.

2.2. Stimuli

Stimuli, exemplified in

vowel. Thus 12 AVi stimuli were included (2 AV-input place [$A_{labiodental}-V_{alveolar}$, $A_{alveolar}-V_{labiodental}$] × 2 voicing conditions × 3 vowels).

A total of 66 stimuli were used across stimulus conditions (18 A, 18 V, 18 AVc, 12 AVi).

Audio and video recordings were made with an adult male speaker of Canadian English sitting against a white background in the recording studio in the Language and Brain Lab at Simon Fraser University. While the speaker produced six randomized repetitions of the 18 English syllables (6 fricatives × 3 vowels) at a normal speaking rate, recordings of the speaker's face were made using a digital camcorder (SONY DCR-HC30/40) positioned approximately 2 m away. In addition, separate audio recordings were simultaneously made with a Shure KSM 109 condenser microphone via an audio interface (M-audio MobilePre USB) to a PC at a 44.1 kHz sampling rate. These high quality audio recordings were used to replace the audio track from the camcorder recording.

A best example was selected from among the six repetitions for each syllable such that the durational difference among the 18 selected syllables was under 10%, the approximate just noticeable difference for duration (Lehiste, 1970). For these selected syllables, the audio-video recordings from the camcorder were aligned with the corresponding high quality audio recordings by synchronizing the two waveforms using SoundForge 8.0. The audio track from the camcorder was then deleted. The audio tracks were then normalized to attain the same unweighted RMS value for the resulting stimuli. The video tracks were edited to have a 1.2 s neutral face before and after the stimulus; that is, before the frame where mouth opening first occurred and after the frame where the mouth was fully closed. All stimuli had a frame length of .06 s, and a resolution of 640 × 480 pixels.

The resulting AV materials were used as AVc stimuli. They were also the basis for creating the A-only, V-only, and AVi stimuli. The A-only and V-only stimuli were created from the AVc stimuli by removing the video tracks or muting the audio tracks, respectively. The AVi stimuli were based on the same auditory and visual components as those used in the AVc, A-only and V-only conditions. To create the AVi stimuli, the audio and video components were aligned from syllables differing only in place of articulation (e.g., A[fa]−V[sa]). Starting with the AVc stimulus which had the target video component (e.g., AVc [sa]), the audio signal of the target audio component (e.g., [fa]) was aligned with the onset of the fricative audio signal from the AVc, and the original AVc audio (e.g., audio [sa]) was removed, so that the resulting AVi had the original AVc video component (e.g., [sa]) with a new audio signal (e.g., [fa]).

To test the intelligibility of the audio signals and the naturalness of the AV signals, the final AVc and AVi stimuli were evaluated by two phonetically trained native speakers of English. An identification task testing intelligibility of the audio signals showed 95% correct responses (errors were 1.5% labiodental, 2% interdental, .5% alveolar, and 1% voicing and other errors) for the audio signals in the AVc stimuli and 100% correct responses for the audio signals in the AVi stimuli. These scores are comparable (and exceed) those in previous AV studies of auditory perception of English fricatives by native English listeners (e.g., Jongman et al., 2003; Werker et al., 1992). The naturalness of the AV stimuli was tested in a 5-point goodness rating task (5 being the most natural) where the same two evaluators judged whether the audio and video signals were naturally synchronized, regardless of what they heard or saw. AVc stimuli were rated 4.4 and the AVi stimuli rated 4.6.

2.3. Procedure

A perception experiment was generated using E-prime 1.0 (Psychology Software Tools, integrating video clips imported from Microsoft Powerpoint files) to present stimuli and log participant responses. Participants were tested in an identification task using the full set of stimuli blocked by modality: A, V, AV (with AVi and AVc stimuli in the same block). Two randomized repetitions of each stimulus were included in each block. The presentation order of the blocks was counter-balanced across participants. The test began with instructions for the task, familiarization with the stimuli (e.g., matching the symbols and the sounds they represent), and five practice trials for each modality (A, V, AVc/AVi). The practice trials were presented with the same speaker as used in the test, but no feedback was provided to avoid any learning effect. The test was followed by a debriefing session, which included a post-test questionnaire. The full experiment, including a short break after three test blocks, lasted about one hour for a participant.

Stimuli were presented auditorily over loudspeakers, visually on a computer monitor, or both. Participants were tested individually, sitting approximately 1 m from a 20 in LCD flat panel computer monitor and two loudspeakers (Altec Lansing) positioned on each side of the monitor, so that the audio and video sources were approximately the same distance from the perceiver. Loudspeakers were used instead of headsets to avoid any bias for the audio component. The audio signal had a comfortable level of approximately 70 dB which has previously been shown to be an appropriate level for speech perception experiments (Nábelek & Robinson, 1982; Takata & Nábelek, 1990). For each trial, a visual fixation point was displayed in the middle of the monitor for one second, followed by the target stimulus.

Six response alternatives [f, v, θ, ð, s, z] were then shown on the monitor, together with an ''other'' option to allow participants to type in an alternative response. ''Th'' and ''dh'' were used to represent [θ] and [ð], respectively (Jongman et al., 2003). The participants were given the ''other'' option since previous research has shown that participants' responses are not limited to the given type of

As shown in Fig. 1, significant group differences were observed in the following conditions. For the labiodentals, in the V condition, only the Korean perceivers showed lower identification accuracy than did the English natives [$F(2,47) = 5.1$, $p < .010$]. For the interdentals in the A condition, the Mandarin perceivers exhibited a significantly lower identification accuracy than the Korean perceivers, who in turn had a significantly lower identification accuracy than the native English perceivers [$F(2,47) = 18.9$, $p < .0001$], and in the AVc condition, only the Mandarin perceivers' identification was lower than that of the native English [$F(2,47) = 10.2$, $p < .0001$]. No other conditions revealed significant group differences.

To summarize, (1) for the labiodentals, both the Korean and Mandarin perceivers showed an increase in identifica-t group dif2-1.208denti5cy c than modality.3319(How8(seealr]TJ0-1.201

the following conditions: labiodental responses in the V condition [$F(4,45) = 3.5$, $p < .038$], interdental responses in the A [$F(4,45) = 11.4$, $p < .0001$], V [$F(4,45) = 3.9$, $p < .026$], and AVc [$F(4,45) = 8.0$, $p < .0001$] conditions, and alveolar responses in the A [$F(4,45) = 7.5$, $p < .001$], V [$F(4,45) = 3.1$, $p < .020$], and AVc [$F(4,45) = 22.7$, $p < .0001$] conditions.

Post hoc analyses show that, for the labiodentals in the V condition, while all groups to some degree misperceived the labiodentals as interdentals, the Koreans (for whom the labiodentals were nonnative) gave a greater percentage of alveolar responses (8%) than the Mandarin group (5%) which in turn showed more alveolar responses than did the English group (2%) ($p < .040$), indicating that for the nonnatives, particularly for the Koreans, the labiodentals and alveolars were less visually distinguishable. The three groups did not differ significantly in the A and AVc conditions in terms of the confusion patterns, all more likely misperceiving labiodentals as interdentals than as alveolars ($p < .05$ for all groups).

For the interdentals in the V condition, both the Mandarin and Korean groups' interdental misperception was more biased towards the alveolars than labiodentals (Mandarin: 30% versus 9%, $p < .001$; Korean: 25% versus 5%, $p < .005$), while the English natives misperceived the interdentals as labiodentals or alveolars to a more similar degree (16% versus 11%, $p > .716$). In both the A and AVc

grasping auditory cues for the labiodentals may have preceded visual cues. Moreover, the confusion patterns for visual perception show that, compared to native English

Mandarin perceivers in their L1). On the other hand, for the Korean perceivers, the auditory perception of these nonnative sounds is relatively good, which may lead to their lack of use or less accurate use of the visual domain. These findings suggest that the perception and acquisition of L2 sounds in the auditory and visual domain may not occur in parallel, and may even take place in a complementary manner.

5. Concluding remarks and future directions

In sum, while L2 learners make use of both auditory and visual information in perceiving nonnative speech sounds, their perception is influenced by the interaction of the AV speech categories in their L1 and L2. In future research, the correlation of visual perceptual distance between L1 and L2 visual categories, and the corresponding level of difficulty in acquisition (e.g., the more dissimilar the L1 and L2 visual categories are, the easier the formation of an L2 category) need to be more extensively studied and even quantified, as also suggested by the auditory speech learning research (e.g., Flege, 2007; Strange, 1999; Strange, Yamada, Kubo, Trent, & Nishi, 2001). Furthermore, the current study indicates that L2 auditory and visual speech cues may not be acquired simultaneously; rather, AV learning may even occur in a complementary manner. Subsequent research and theories should take into account visual and auditory relationships of L2 speech sounds with corresponding L1 sounds.

Additionally, as the current study focuses on the effect of L1, it has left unaddressed a number of factors that may also affect the perception of L2 speech contrasts, such as L2 phonetic context and variability (Hardison, 2005b), length of residence (LOR) in an L2 country and L2 input (Flege, 2007), attention (Guion & Pederson, 2007), etc. One factor worth noting is phonetic context. For example, previous research has shown that vowel context influences the neighboring sounds not only for native AV perception (e.g., Benguerel & Pichora-Fuller, 1982; Daniloff & Moll, 1968) but also for nonnative perception (e.g., Hardison, 2005b). Since the vowels used in the current study exist in the native phonetic inventories of all three native groups, no interaction of L1 group and vowel context was

Binnie, C. A., Montgomeery, A. A., & Jackson, P. L. (1974). Auditory and visual contributions to the perception of consonants. Journal of Speech and Hearing Research, 17, 619–630.

Burnham, D., & Dodd, B. (1998). Familiarity and novelty in infant cross-language studies: Factors, problems, and a possible solution. In C. Rovee-Collier (Ed.), Advances in infancy research (pp. 170–187).

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. Developmental Psychobiology, 45, 204–220.

Burnham, D., Lau, S., Tam, H., & Schoknecht, C. (2001). Visual discrimination of Cantonese tone by tonal but non-Cantonese speakers, and by non-tonal language speakers. In D. Massaro, J. Light, & K. Geraci (Eds.), Proceedings of the international conference on auditory–visual speech processing (pp. 155–160).

Chen, Y., & 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 (pp. 2177–2180). Saarbrueken, Germany.

Chen, T., & Massaro, D. W. (2004). Mandarin speech perception by ear and eye follows a universal principle. Perception and Psychophysics, 66, 820–836.

Cienkowski, K. M., & Carney, A. E. (2002). Auditory–visual speech perception and aging. Ear and Hearing, 23, 439–449.

Daniloff, R. G., & Moll, K. (1968). Coarticulation of lip rounding. Journal of Speech and Hearing Research, 11, 707–721.

Davis, C., & Kim, J. (2004). Audio–-visual interactions with intact clearly audible speech. Quarterly Journal of Experimental Psychology, 57A, 1103–1121.

De Gelder, B., & Vroomen, J. (1992). Auditory and visual speech perception in alphabetic and non-alphabetic Chinese/Dutch bilinguals. In R. J. Harris (Ed.), Cognitive processing in Bilinguals (pp. 413–426). Amsterdam: Elsevier.

Dodd, B. (1977). The role of vision in the perception of speech. Perception, 6, 31–40.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. Journal of Speech Language and Hearing Research, 12, 423–425.

Erdener, V., & Burnham, D. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. Language Learning, 55, 191–228.

Faulkner, A., & Rosen, S. (1999). Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception. Journal of the Acoustical Society of America, 106, 2063–2073.

Flege, J. E. (1980). Phonetic approximation in second language acquisition. Language Learning, 30, 117–134.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), Speech perception and linguistic experience (pp. 233–273). Baltimore: York.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole, & J. Hualde (Eds.), Laboratory phonology, Vol. 9. Berlin: Mouton de Gruyter.

Flege, J. E., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second language learning. Journal of Memory and Language, 41, 78–104.

Green, K. P., & Kuhl, P. K. (1989). The role of visual information in the processing of place and manner features in speech perception. second K. , 30

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. Developmental Science, 6, 191–196.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd, & R. Campbell (Eds.), Hearing by eye: The psychology of lip-reading (pp. 97–113). London: Erlbaum.

Riney, T. J., & Flege, J. E. (1998). Changes over time in global foreign accent and liquid identifiability and accuracy. Studies in Second Language Acquisition, 20, 213–244.

Sams, M., Manninen, P., Surakka, V., Helin, P., & Katto, R. (1998). McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context. Speech Communication, 26, 75–87.

Schmidt, A. M. (1996). Cross-language identification of consonants, Part 1: Korean perception of English. Journal of the Acoustical Society of America, 99, 3201–3211.

Sekiyama, K. (1994). Differences in auditory–visual speech perception between Japanese and Americans: McGurk effect as a function of incompatibility. Journal of the Acoustical Society of Japan (E), 15, 143–158.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. Perception and Psychophysics, 59, 73–80.

Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. Journal of the Acoustical Society of America, 90, 1797–1805.

Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. Journal of Phonetics, 21, 427–444.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory–visual speech perception examined by fMRI and PET. Neuroscience Research, 47, 277–287.

Sekiyama, K., Tohkura, Y., & Umeda, M. (1996). A few factors which affect the degree of incorporating lip-read information into speech perception. In Proceedings of the 4th International Conference on Spoken Language Processing (pp. 1481–1484). Philadelphia, PA.

Soto-Faraco, S., Navarra, J., Voloumanos, A., Sebastián-Gallés, N., Weikum, & Werker, J. F. (2007). Discriminating languages by speechreading. Perception and Psychophysics, 69, 218–231.