Phonetic Drift in Fricatives and the Effects of L2 Experience on L1 Phonetic Categories

JOHN MATTHEWS, TAKAKO KAWASAKI
KUNIYOSHI TANAKA, MASAKI TAKEUCHI

1. Introduction

Phonetic drift refers to a gradual shift in phonetic categories that happens naturally over time.[1] The drift occurs along various acoustic parameters, substantively changing the acoustic realization of category prototypes and moving the boundaries that separate phonetic categories from one another. Originally, the term referred to diachronic change that occurs as languages are imperfectly transmitted successively from generation to generation (Garrett and Johnson, 2013; Sapir, 1921). More recently, however, the term has come to apply also to synchronic shifts in phonetic categories observed among individuals (Houde and Jordan, 1998). With the sudden and dramatic change that occurs to one's linguistic environment in immersion education, language learners abroad have demonstrated this kind of drift in native language phonetic categories under pressure to accommodate new segmental categories within their representational space for phonetic categories (Chang, 2011; Guion, 2003).

Some of the most heavily studied phonetic categories both in speech production and speech perception research are the voiced and voiceless

1) A preliminary account of the work discussed here was presented at the *New Sounds* Conference held at Waseda University in August, 2019. We grateful acknowledge financial support from the Japan Society for the Promotion of Science (JSPS) 学術研究助成基金助成金 No. 18K00848.

obstruent stops. Differences in Voice Onset Time (VOT) consistently and reliably distinguish two, and sometimes three, categories along the continuum. Notably, speakers of different languages have been shown to place the acoustic boundary between categories at different absolute values along the scale. French speakers, for example, perceive a categorical change between voiced and voiceless segments at VOTs around zero (from -5 msec to +5 msec) (Caramazza & Yeni-Komshian, 1973) and produce segments from either category on opposite sides of the same boundary (Kessingerf & Blumstein, 1997). English speakers place the boundary in the area of 30-40 msec VOT (Winitz, LaRiviere & Herriman, 1975). These boundaries are the classic example of categorical perception, in which listeners exhibit rather weak perceptual sensitivity to contrasts within the categories on either side of the boundary but heightened sensitivity to contrasts of equal acoustic distance that straddle the boundary.

Immersion in a second language environment can lead to phonetic drift in native language categories, a gradual ebb of the perceptual boundary between categories under influence from new phonetic surroundings. English native speakers immersed in French, for example, show measurable changes in the VOT boundary between voiced and voiceless stops (Tice & Woodley, 2012). In addition, boundaries between vowel categories along F1 and F2 formant frequency dimensions have also been shown to exhibit phonetic drift under similar conditions. (Chang, 2013; Guion, 2003). These effects can sometimes arise very rapidly during the initial period of immersion and then subside as the learner tunes into two distinct sets of acoustic realizations for phonetic categories.

In the present research we extend the study of phonetic drift to contrasts in place of articulation among voiceless fricatives. In particular, we are interested in the interaction between two similar, but distinct, post-alveolar sibilant fricatives in Japanese [ɕ] and English [ʃ], and their effects on the nearby voiceless alveolar fricative [s], which occurs in both languages. Following previous research we investigate whether studying abroad, immersed in an English-language environment, leads Japanese learners of English to modify a native language phonetic category alongside the establishment of a new, distinct L2 post-alveolar fricative.

In a perception experiment reported elsewhere (Kawasaki, Tanaka, Takeuchi & Matthews, 2019), we presented an identification task to two groups of late Japanese-English bilinguals: one with experience studying abroad in an English-speaking country (SA) and one with no study-abroad experience (NoSA). With an array of real words displayed on a computer monitor, participants heard a nonsense CV-syllable masked in noise and then clicked on the word whose initial segment matched the consonant they heard at the start of the syllable. Figure 1 presents the resulting Confusion Matrices with audio stimulus items listed down the left side of each matrix and the corresponding identification judgments across the top.

Poorest performance was observed by members of both groups in the perception of the English labiodental fricative [f], with greater difficulty before the vowel [i] than [ɑ]. While it may be tempting to attribute this difficulty to the absence of this articulation from spoken Japanese, the superior performance on the equally unfamiliar interdental [θ] shows that something more must be at play. While it lies beyond the scope of the present study, it seems likely that the acoustic similarity between voiceless non-sibilant fricatives [f] and [θ] could lead to single-category (SC) assimilation in the sense of the Perceptual Assimilation Model (Best, 1994, 1995) or to equivalence classification in the sense of the Speech Learning Model (Flege, 1987).

The more relevant observations for our purposes here are the disparities in identification accuracy for the alveolar sibilant fricative [s] between NoSA and SA groups and between the two vowel contexts [i] and [ɑ]. The NoSA group accurately identified [si] at a rate of only 45.98%. This is precisely the phonetic context where [s] fails to surface in Japanese, undergoing alternation with [ʃ] when the sequence is created through morphological concatenation. There is no corresponding disparity in identification accuracy

Figure 1. Confusion matrices: Stimulus × Identification

NoSA group: before vowel [ɑ]

	fa	*sa*	*ʃa*	*ta*	*θa*
fa	**73.56%**	4.60%	0.00%	0.00%	21.84%
sa	0.00%	**87.36%**	0.00%	0.00%	12.64%
ʃa	0.00%	0.00%	**100.00%**	0.00%	0.00%
θa	20.69%	12.64%	0.00%	1.15%	**65.52%**

NoSA group: before vowel [i]

	fi	*si*	*ʃi*	*ti*	*θi*
fi	**27.59%**	11.49%	12.64%	2.30%	45.98%
si	0.00%	**45.98%**	36.78%	0.00%	16.09%
ʃi	0.00%	17.24%	**77.01%**	0.00%	2.30%
θi	12.64%	5.75%	10.34%	1.15%	**68.97%**

SA group: before vowel [ɑ]

	fa	*sa*	*ʃa*	*ta*	*θa*
fa	**74.36%**	2.56%	0.00%	0.00%	23.08%
sa	0.00%	**92.31%**	0.00%	0.00%	7.69%
ʃa	0.00%	0.00%	**100.00%**	0.00%	0.00%
θa	2.56%	7.69%	0.00%	2.56%	**87.18%**

SA group: before vowel [i]

	fi	*si*	*ʃi*	*ti*	*θi*
fi	**28.21%**	2.56%	2.56%	0.00%	66.67%
si	0.00%	**71.79%**	10.26%	0.00%	17.95%
ʃi	0.00%	17.95%	**82.05%**	0.00%	0.00%
θi	15.38%	5.13%	0.00%	5.13%	**74.36%**

souce: Kawasaki, Tanaka, Takeuchi, & Matthews, 2019: 705

before the vowel [ɑ], where accuracy rates are similar for both groups (87.36% and 92.31%) and no one misheard [s] as [ʃ]. It appears that the source of interference with accurate perception is the phonological neutralization of the contrast that occurs in Japanese. Notably, however, the SA group demonstrates substantially greater accuracy (71.79%) and only one-third the rate of misidentifying [si] as [ʃi] (10.26% vs. 36.78%), showing that the extensive exposure to spoken English afforded by immersion may lead to overcoming the interference effect of a learner's L1 phonology.

2. Experiment

The mapping of phonetic categories applies both to perception and to production. Phonetic drift in speech production leads to measurable changes in the pronunciation of native language categories under pressure from developing categorical representations for segments in an individual's L2. We therefore conducted the following experiment to measure phonetic drift in the pronunciation of voiceless sibilant fricatives among Japanese-English bilinguals with different amounts of experience surrounded by natural spoken language input.

2.1. Method

To determine whether fricatives exhibit the same kind of phonetic drift that has been observed for obstruent stops and vowel categories, we elicited English and Japanese voiceless sibilant fricatives from a diverse group of Japanese-English bilinguals. In case there might be an interaction between any observed phonetic drift and spoken language exposure, we divided participants into three groups based on their amount of experience immersed in an English-speaking environment. The elicited forms were then submitted to careful acoustic analysis to measure the effects of L2 exposure on L1 phonetic categories.

Table 1. Participants Groups

Group	**SA**	**NoSA**	**SB**
n	7	4	2
ages	19–23, 33, 47 m = 25.3	19–21 m = 20	14–17 m = 16.5

2.1.1. Participants

Thirteen participants in three groups performed an elicitation task, producing target forms both in isolation and within a carrier phrase. They included 7 participants who had studied abroad in an English-speaking country for a range of durations from approximately one month to more than 10 years (the SA group). We recruited another 4 participants with no experience abroad (the NoSA group), although they did have some degree of experience with English in Japan as they all major in English linguistics. Finally, a small group of two simultaneous bilinguals (the SB group) was included to compare the effects of late age immersion with life-long exposure to both phonetic inventories. Both individuals were born and raised in Japan in a family where both parents are English native speakers and the language spoken in the home is exclusively English. Also, both individuals have attended all of their schooling since kindergarten exclusively in Japanese.

2.1.2. Materials

The distinct places of articulation for the sibilant fricatives under investigation are measured through an acoustic correlate known as Center of Gravity (Hanulíková and Weber, 2010). Both English and Japanese contain an alveolar fricative [s] with CoG near 7000 Hz, but the two languages differ in their realization of post-alveolar fricatives, broadly transcribed as /ʃ/. Japanese contains an alveopalatal fricative [ɕ] with CoG between 5000 and 6000 Hz, whereas English contains the palatoalveolar fricative [ʃ] with CoG near 4000 Hz (Figure 2).

Elicitation materials contained real English and Japanese words with initial voiceless sibilant and non-sibilant fricatives as well as stops before a high or a

Figure 2. Sibilants along Center of Gravity (CoG) scale（Hertz）

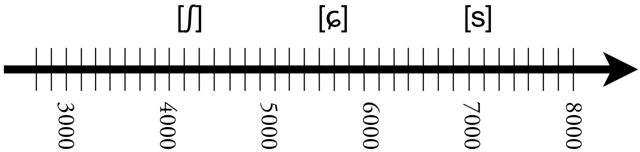


Table 2. Items elicited for analysis

Language	***Words***
English	sand, shine, sink, shield, tank, hide, fan, thank, kind, tint, hill, fill, thief, keen
Japanese	san（酸）, shain（社員）, shin（芯）, ten（点）, han（班）, kan（缶）, tiin（ティーン）, hin（品）, kin（金）

low vowel（Table 2）. Only items containing voiceless sibilant fricatives were submitted to acoustic analysis.

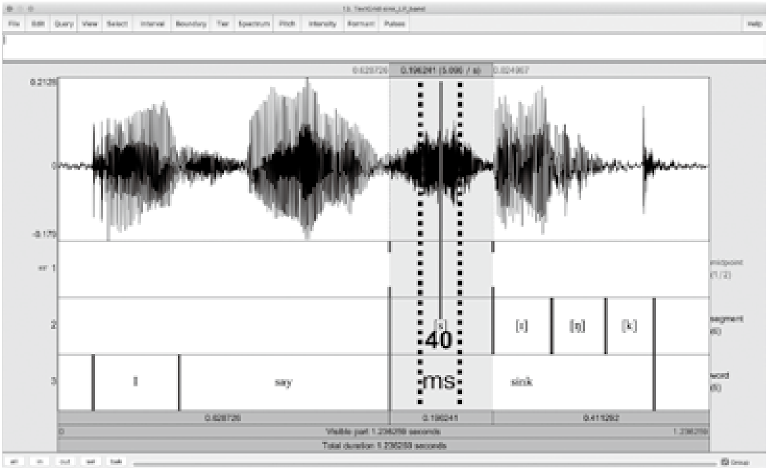
The two vowel environments are important because Japanese phonology neutralizes the contrast between alveolar and post-alveolar obstruents before the high front vowel（i.e., ［i］）while maintaining the contrast, including minimal pairs, before low vowels（i.e., ［a］）.

2.1.3. Procedure

Participants produced items in four blocks: two in English and two in Japanese. All items were pronounced both in isolation and embedded within the carrier phrase "I say ____, again" or "Kore-kara ____ to iimasu." Each list was repeated 3 times.

Audio recordings were digitized at 44.1 kHz, 16-bps, and subsequently passed through an 11 kHz low-pass filter. Center of Gravity was measured with Praat（Boersma & Weenink, 2019）at the center 40 msec span within each fricative, following Jongman, Wayland & Wong（2000）, as illustrated in Figure 3.

Figure 3. Portion of fricative in CoG measurement



2.2. Results

In the graphs that follow, Center of Gravity (CoG) is plotted from 0 to 10K Hertz on both axes in order to compare the degree of deviation between two phonetic categories. Data points that fall along the line from the origin with a slope of 1 exhibit no difference between the categories. The further data points lie from the line, the greater the deviation between categories under comparison.

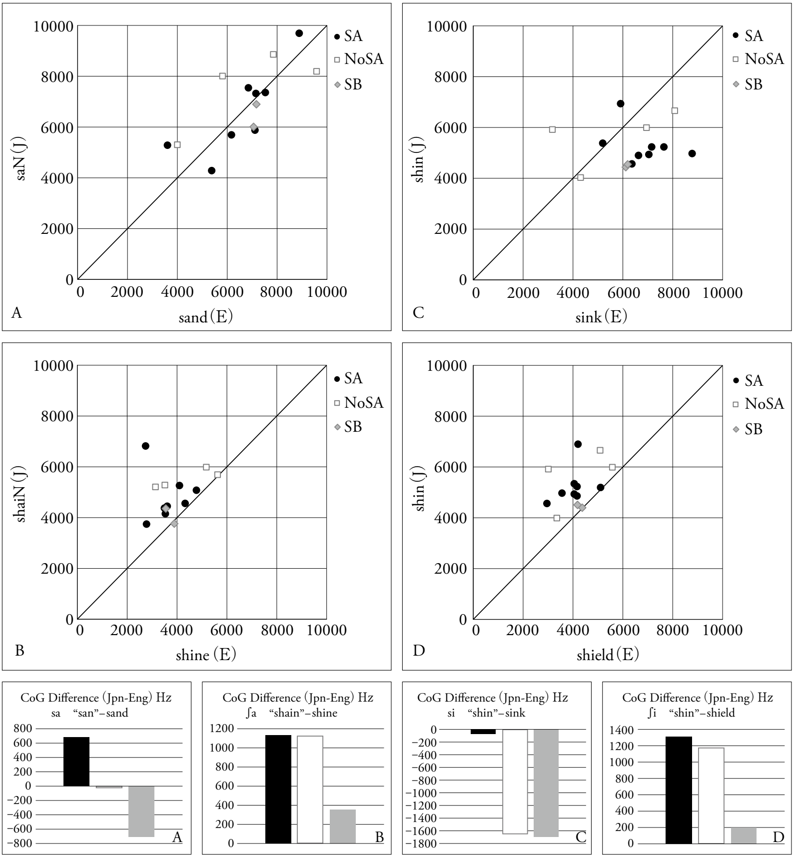
We begin by reporting cross-language comparisons in Figure 4. Measurements for Japanese categories appear on the y-axes and English categories on the x-axes. Panel A compares pronunciations of Japanese and English alveolar [s] before a low vowel and shows that the SA group produces these segments in a nearly identical fashion in the two languages, with some individual variation. Cumulative differences for the group appear in the Mini-Panel A at the bottom of the figure. The simultaneous bilinguals produce distinct articulations in the two languages with a CoG approximately 700 Hz higher for English [s] than for Japanese [s], though the distinction appears to be greater in the pronunciations of one member of the group than the other. One member of the SA group makes this same distinction, too, but the others exhibited the opposite pattern with lower CoG for English [s] than for Japanese.

Panels B and D show pronunciations of the post-aveolar fricatives before low and high vowels, respectively. The simultaneous bilinguals (SB) demonstrate use of the same articulation in both languages. Both groups of late bilinguals produce substantially lower CoG for these segments in English than in Japanese as seen in the Mini-Panels B and D at the bottom of Figure 4, though individuals from each group do approach similar articulations in both languages. While both late bilingual groups differ markedly from the SB pair, there is no measurable difference between them, despite the differences in the their experience, or lack of experience, in an English immersion environment.

Panel C presents the analysis of sibilant fricatives before the high front vowel [i], an environment of phonological neutralization in Japanese. As a group, the SA participants resemble the simultaneous bilinguals in producing a clear distinction between [s] in English and [ʃ] in Japanese, unaffected by potential effects of Japanese neutralization when pronouncing English. However, here again, there is some variation, with two individuals producing nearly identical CoG in the two languages or even slightly lower CoG in English than in Japanese. The NoSA group produces roughly the same articulation in both cases, exhibiting a neutralization of the contrast in both languages in this phonetic environment (see Mini-Panel C at the bottom of Figure 4).

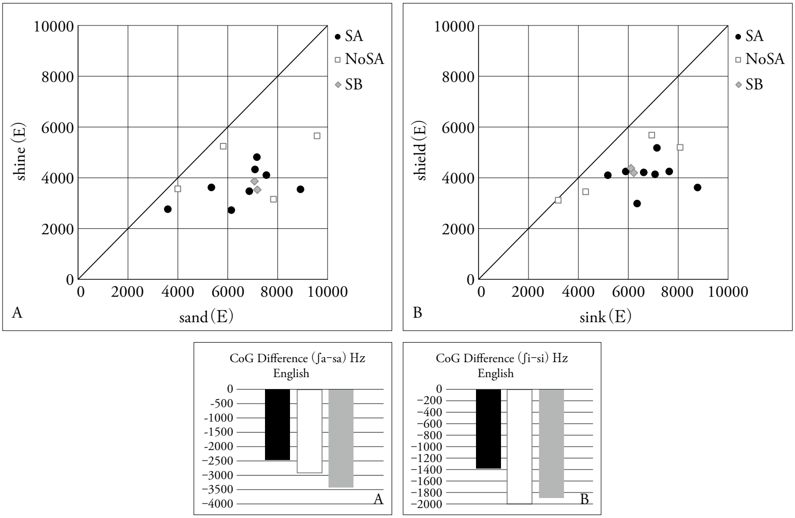
The next set of results, presented in Figure 5, compares pronunciations of [s] and [ʃ] in English to one another. It this case, measurements of CoG for the alveolar fricative [s] are plotted along the x-axes and post-alveolar [ʃ] along the y-axes. For the most part, all participants clearly distinguish between these segments, both before the low vowel (A Panels) and before the high front vowel (B Panels), with the exception of individual members of the NoSA group who produce segments with near identical CoG. Interestingly,

Figure 4. Cross-language comparisons of [s] and [ʃ] in Japanese and English



this behavior is not limited to the phonetic environment in which Japanese neutralizes the contrast (i.e., before [i]). Two members of the group produce very little difference between [s] and [ʃ] before [a] as well, whereas the other two members differentiate the segments with substantial differences in CoG. As a group, the NoSA participants exhibit a bimodal distribution in their pronunciation of the sibilant fricatives, indicating that they do not, in fact, form a homogeneous group (see especially Panel A in Figure 5).

Figure 5. Comparison of English [s] ~ [ʃ] fricative contrast

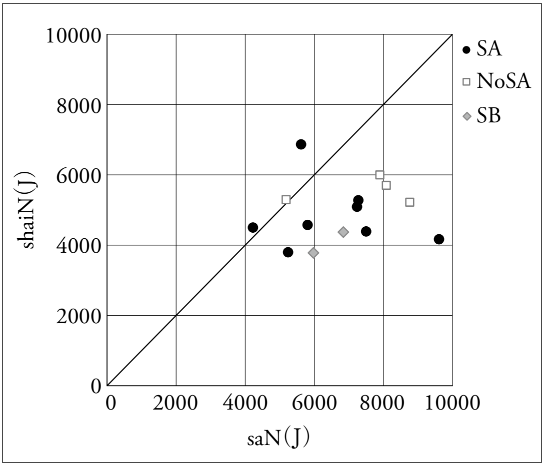


The final set of data from the elicitation task appears in Figure 6. Here we compare CoG measurements for the [s]~[ɕ] contrast in Japanese, with the first along the x-axis and the second along the y-axis. As Japanese neutralizes the contrast before [i], these data are from pronunciations before the low vowel [a] only.

In general, all groups produced the alveolar [s] with measurably higher CoG than the alveopalatal [ɕ] as expected, with exceptions in both late bilingual groups. It may be that other acoustic characteristics distinguish these segments in the speech of those individuals, but we have no data to speak to that possibility. The range of acoustic cues available for distinguishing among sibilant fricatives is not widely understood and remains an area in need of additional research.

In summary, the simultaneous bilinguals (SB) appear to produce near identical segmental articulations for segments shared between their two languages (Figure 4, panels A, B, & D) but clearly differentiate the contrasts within each language (Figures 5 and 6), including in the phonetic

Figure 6. Comparison of Japanese [s] ~ [ɕ] fricative contrast



environment where Japanese phonology neutralizes the contrast (Figure 4, panel C). The SA group exhibit variable performance: sometimes resembling the SB group (Figure 4, panel C; Figure 5, panel B), sometimes resembling the NoSA group (Figure 4, panels B & D), and sometimes falling between the two other groups (Figure 4, panel A; Figure 5, panel A).

3. Discussion

The primary issue to be addressed by these findings is whether the experience of studying abroad, and the corresponding greater experience with spoken English input that immersion affords, leads to phonetic drift in the pronunciation of late Japanese-English bilinguals. In comparison to the NoSA group, participants in the SA group did indeed produce the alveolar fricative [s], a segment in both Japanese and English, with a CoG more similar to simultaneous bilinguals who have been speaking English their whole lives. In addition, we have seen that those with study abroad experience resemble the simultaneous bilinguals in producing distinct articulations before the vowel [i] for English [s] and Japanese [ɕ], escaping the neutralization effects of L1 interference that members of the NoSA group show. Moreover, they more clearly differentiate the English contrast between [s] and [ʃ] than the NoSA group does.

While the simultaneous bilinguals appear to have developed a single post-alveolar fricative that they employ in both languages (contrary to bilinguals reported in Fowler et al., 2008), both groups of late bilinguals show distinct articulations for Japanese [ɕ] and English [ʃ]. Looking at individual data, however, we see that most members of the SA group appear to be drifting toward a common articulation for these two segments, as their [ɕ] approaches CoG values nearly equal to their articulations of [ʃ]. While we are unable to conclude that such merging of phonetic realizations arises as a direct consequence of the greater exposure to spoken English that this group has experienced, it is nevertheless consistent with that analysis.

Without intensive immersion experience, participants in the NoSA group were not expected to have established separate post-alveolar fricatives for Japanese and English. Based on the PAM (Best, 1995) and SLM (Flege, 1995) frameworks, we predicted the similarity between L1 and L2 segments would precipitate equivalence classification (or equivalently, single-category assimilation). Recall, however, that the pool from which these participants were recruited was comprised of Japanese university students majoring in linguistics with coursework in English phonetics. Taking this into consideration, we now speculate that phonetic drift may not arise solely from intensive exposure as experienced in an immersion environment but may also emerge from the heightened perceptual awareness that is brought about through acquired metalinguistic knowledge as well.

Finally, it should be pointed out that phonetic drift could be instantiated in two different ways. The first can be thought of as a kind of attraction, where phonetic categories in a learner's native language acquire acoustic characteristics that resemble elements in the second language input that surrounds them. Another type of phonetic drift can be envisioned whereby native language categories exhibit a kind of dispersion from one another under pressure from the insertion of a new category between existing categories. This would be the condition that English native speakers face when acquiring Japanese. Their native language fricatives [s] and [ʃ] represent distinct categories along the CoG scale (as presented in Figure 2 above). The introduction of Japanese [ç] amid those categories could drive each of the native categories to drift in opposite directions to accommodate the new category. We consider this to be an interesting new direction for research on phonetic drift which we hope to pursue in future experimental work.

References

Best, C. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In C.Goodman & H. Nusbaum (Eds), *The development of speech perception*, 167–224. Cambridge: The MIT Press.

Best, C.T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*, pp. 171–204. Baltimore, MD: York Press.

Boersma, P. & D. Weenink (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.01, retrieved 14 August 2019.

Caramazza, A. & G. H. Yeni-Komshian (1973) The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* 54, 421.

Chang, C.B. (2013) A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, 41, 520–533.

Chang, C.B. (2011) Systematic drift of L1 vowels in novice L2 learners. Proceedings of the International Congress of Phonetic Sciences XVII, pp. 428–431. August 17–21, 2011: Hong Kong.

Flege, J. E. (1987). The production of "new" and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47–65.

Flege, J. E. (1995). Second language speech learning: Theory, problems, findings. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*, pp. 233–277. Baltimore, MD: York Press.

Fowler, C.A., Sramko, V., Ostry, D.J., Rowland, S.A., & P. Hallé (2008). Cross-language phonetic influences on the speech of French-English bilinguals. *Journal of Phonetics*, 36, 649–663.

Garrett, A. & K. Johnson (2013) Phonetic bias in sound change. In A.C.L. Yu (Ed.), *Origins of Sound Change: Approaches to phonologization*, pp. 51 – 97. Oxford: Oxford University Press.

Gong, J., Lecumberri, M.L.G., & M. Cooke (2016) Can intensive exposure to foreign language sounds affect the perception of native sounds? Proceedings of INTERSPEECH 2016, pp. 883－887. September 8－12, 2016: San Francisco.

Guion, S.G.（2003）The vowel systems of Quichua-Spanish bilinguals: Age of acquisition effects on the mutual influence of the first and second languages. *Phonetica*, 60, 98－128.

Hanulíková, A., & Weber, A.（2010）. Production of English interdental fricatives by Dutch, German, and English speakers. *New Sounds 2010: Sixth International Symposium on the Acquisition of Second Language Speech*, pp. 173－178. Poznan: Adam Mickiewicz University.

Houde, J. & M. Jordan（1998）Sensorimotor adaptation in speech production. *Science* 279: 1213－1216.

Jongman, A., Wayland, R., & S. Wong（2000）Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, 108（3）, 1252－1263.

Kartunshina, N., Frauenfelder, U.H., & N. Golestani（2016）How and when does the second language influence the production of native sounds: A literature review. *Language Learning* 66（suppl. 2）, 155－186.

Kawasaki, T., Tanaka, K., Takeuchi, M., & J. Matthews（2019）Reconstruction of the Perceptual Map for Fricatives in L2 Acquisition. Presentation at the 36th Annual Meeting of the Japan Cognitive Science Society. Hamamatsu（Shizuoka）, Japan. September 5－7, 2019.

Kessingerf, R.H. & S.E. Blumstein（1997）Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics* 25（2）, 143－168.

Sapir, E.（1921/2004）. *Language: An introduction to the study of speech*. Mineola, NY: Dover Publications.

Tice, M. & M. Woodley（2012）Paguettes & bastries: Novice French learners show shifts in native phoneme boundaries. *UC Berkeley Phonology Lab Annual Report* 8（8）, 72－75.

Winitz, H., LaRiviere, C. & E. Herriman（1975）Variations in VOT for English initial stops. *Journal of Phonetics* 3（1）, 41－52.