

JULY 21 2023

# Bilingual phonological contrast perception: The influence of Quanzhou Southern Min on Mandarin non-sibilant fricative discrimination

Caihong Weng; Ioana Chitoran; Alexander Martin



*JASA Express Lett* 3, 075202 (2023)

<https://doi.org/10.1121/10.0020266>



View  
Online



Export  
Citation

CrossMark



**ASA**

Advance your science and career as a member of the  
**Acoustical Society of America**

[LEARN MORE](#)

# Bilingual phonological contrast perception: The influence of Quanzhou Southern Min on Mandarin non-sibilant fricative discrimination

Caihong Weng,<sup>1,a)</sup> Ioana Chitoran,<sup>1</sup> and Alexander Martin<sup>2</sup>

<sup>1</sup>UFR Linguistique, Université Paris Cité, CLILLAC-ARP, Paris, France

<sup>2</sup>Center for Language and Cognition, University of Groningen, Groningen, The Netherlands

[caihong.weng@etu.u-paris.fr](mailto:caihong.weng@etu.u-paris.fr); [ioana.chitoran@u-paris.fr](mailto:ioana.chitoran@u-paris.fr); [alexander.martin@rug.nl](mailto:alexander.martin@rug.nl)

**Abstract:** This study explores the discrimination of Mandarin non-sibilant fricatives by bilingual speakers ( $N=40$ ) of Quanzhou Southern Min (L1) and Mandarin (L2) in different phonological contexts, including rounded vowels and the glide [w]. The results of the ABX discrimination task indicate significant contextual effects of the following sound, in line with predictions based on the Perceptual Assimilation model (PAM) [Best (1995). *J. Phon.* **20**(3), 305–330]. Additionally, the observed result could not be fully explained by the acoustic distance between stimuli, and discrimination ability was better for speakers with more exposure to and use of Mandarin. © 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

[Editor: Qian-Jie Fu]

<https://doi.org/10.1121/10.0020266>

Received: 20 March 2023 Accepted: 26 June 2023 Published Online: 21 July 2023

## 1. Introduction

The acquisition of a new phonological system in second language (L2) learning is a fundamental issue in L2 research. Our ability to perceive speech is significantly influenced by our first language (L1): an ample body of research in L2 learning has demonstrated that L2 speech is perceived through the filter of one's L1 (see, e.g., Best, 1993; Escudero, 2005; Flege and Fletcher, 1992; Robinett and Schachter, 1983). This research has shown that aspects of both the phonology and the phonetics of the L1 can affect the perception of L2 contrasts.

From a purely phonological perspective, many L2 perception studies have shown that L1 phonological structure influences the perception of L2 sounds. A well-known example is that L1 Japanese speakers tend to have difficulty discriminating the L2 English [r]~[l] contrast, since this contrast is absent in their L1 phonological system (MacKain *et al.*, 1981). On the other hand, L2 perception can also be influenced by the phonetic differences between L1 and L2 sounds. According to Flege's Speech Learning model (Flege, 1995), the larger the acoustic-phonetic distance between an L1 and an L2 sound, the easier it will be for the L2 learner to establish a new category for the L2 sound, though this position has been challenged. For example, Strange *et al.* (2005) and Strange *et al.* (2004) demonstrate that acoustic similarity between English and German vowels does not always predict perceptual similarity. In a perceptual assimilation task, English-speaking listeners were presented with multiple tokens of German vowels and were asked to indicate which English category each token was most similar to. The researchers found that front rounded vowels (which are allophonic variants of back rounded vowels in the variety of English spoken by the participants), did not always show predictable perceptual similarity patterns in line with acoustic distance. Hence, while acoustic distance undoubtedly plays a central role in perception, phonological patterns in one's L1 clearly also have a role to play.

With this backdrop in mind, the current study examined bilinguals raised in the Southern Min (闽南话) and Mandarin speaking community of Quanzhou, China. Speakers in Quanzhou are bilingual in Quanzhou Southern Min (泉州闽南话, henceforth QSM) and Mandarin. They start learning Mandarin at age six, as required by government policy since 1956. QSM and Mandarin have overlapping but different vowel and consonant inventories, providing an ideal setting for comparing segmental contrasts. Even though Quanzhou residents are highly skilled QSM and Mandarin speakers, there is considerable variation in their L2 Mandarin speech. For example, the Mandarin phonemic contrast between labial and velar fricatives is particularly difficult to learn, since QSM has neither /f/ nor /x/ (Kwok, 2018). In QSM, there is only one non-sibilant, glottal fricative, /h/, whereas Standard Mandarin has both /f/ and /x/ (Duanmu, 2007). The variation between velar and labial fricatives is well attested in this bilingual QSM–Mandarin community: speakers frequently merge the two categories /f/ and /x/ into one single velar or glottal category in production, pronouncing Mandarin words like 发 /fā/, “to

<sup>a)</sup> Author to whom correspondence should be addressed.

send out,” as [huā]. Previous studies of this variation focused specifically on how Southern Min speakers produce these Mandarin sounds (e.g., Peng, 1993); however, Southern Min speakers’ perception has not yet been investigated. The present study, therefore, focuses on the variation in perception between L2 Mandarin velar and labial fricatives in this bilingual community.

An important source of variation in the perception of fricative contrasts is coarticulation with the following sounds. For example, Mann and Repp (1980) studied the perception of the English [s]~[ʃ] contrast and demonstrated the influence of the following vowel on the perception of fricatives. Listeners in that study perceived more instances of [s] in the context of [u] than in the context of [a]. This outcome reflects a perceptual adjustment for the coarticulatory effect of rounded vowels on preceding fricatives, which results in a lowering of the spectrum of frication noise. Regarding labial and velar fricatives in particular, Greenlee (1992) demonstrated by way of a discrimination task that speakers of Chicano Spanish made more perceptual confusions between [f] and [x] when the following vowel was rounded. Listeners heard velar fricatives as labial more often in the context of a following rounded vowel since, in that context, both fricatives have more similar formant transitions into the rounded vowel, making it more difficult for the listener to discriminate the two sounds.

In addition to the vowel context, the presence of a glide can also influence fricative perception. Ohala and Lorentz (1977) showed that when the glide [w], which has both labial and velar articulations, occurs adjacent to a fricative, the labial articulation of [w] is more perceptually salient. This means that when a fricative such as [x] is followed by [w], the frication of [x] combines with the labiality of [w], yielding a predominantly labial percept, such as [f], rather than a velar one. This phenomenon is in fact attested cross-linguistically. Mazzaro (2011) showed in an AX (same–different) discrimination task that the discrimination accuracy of the [f] and [x] contrast in Argentine Spanish is lower when the velar fricative is followed by [w], in line with the Greenlee (1992) findings concerning [u] for speakers of Chicano Spanish. These results all indicate that the coarticulation with the following glide contributes to the variation in the perception of the fricative.

Thus, the aim of the present study was to compare the discrimination of the contrast [f]~[x] in different phonological contexts by bilingual speakers of L1 QSM and L2 Mandarin. Our research questions are as follows: (1) Is the L1 phonological system more likely to be the source of L2 perceptual constraints, rather than the acoustic-phonetic differences between the two fricatives? (2) If so, what exactly may be driving the effect? Based on our research questions, we hypothesize that phonological properties of the L1 shape perception of L2 speech. We predict that speakers’ ability to discriminate the contrast should be affected by a following labial glide or rounding on the following vowel. As a theoretical framework, we adopt the Perceptual Assimilation model (Best, 1995), which makes explicit predictions about discrimination differences for different types of non-native phonological contrasts. For example, when listeners associate two L2 phones with two different L1 phonemes (i.e., Two-Category assimilation), this contrast is predicted to be easy to discriminate. However, when listeners associate two L2 phones to a single L1 phonological category (i.e., Single-Category assimilation), the contrast is predicted to be hard to discriminate.

Following PAM, we predict that the [f]~[x] contrast in the context of an unrounded vowel (e.g., [fa] vs [xa]) should form a Two-Category assimilation, in which [fa] will be perceived as [h<sup>w</sup>a] (with the labiality perceived as labialization of the fricative, and the percept thus being assimilated to the QSM /h<sup>w</sup>/ category), while [xa] will be perceived as [ha] and the fricative will be assimilated to the native /h/ category. This contrast should be easy for L1 QSM speakers to discriminate. However, the [f]~[x] contrast in the presence of the glide [w] (e.g., [fa]~[x<sup>w</sup>a]) or in the context of a rounded vowel (e.g., [fu]~[xu]) should form cases of Single-Category assimilation. We predict that both [fa] and [x<sup>w</sup>a] will be perceived as [h<sup>w</sup>a] and assimilated to the same /h<sup>w</sup>/ category. Similarly, both [fu] and [xu] will be perceived as [hu] and assimilated to the same /h/ category. The contrast in both of these contexts should be more difficult to discriminate. These predictions are based on the gradient assimilation levels of PAM, where discrimination in the case of Two-Category assimilation is expected to be better than discrimination in the case of Single-Category assimilation.

## 2. Method

The experiment consisted of an ABX discrimination task in which participants had to decide whether the third stimulus (X) is the same as the first stimulus (A) or the second stimulus (B). In this task, we tested participants’ perception of the Mandarin [f]~[x] contrast in different phonological contexts. All materials including stimuli, anonymized data, and analysis notebook are available at the following link: [https://osf.io/abnj5/?view\\_only=882026443e5d431f9961193c2e745fb6](https://osf.io/abnj5/?view_only=882026443e5d431f9961193c2e745fb6).

### 2.1 Stimuli

Stimuli consisted of a total of nine disyllabic nonce words of the form consonant-vowel-consonant-vowel (CVCV). The initial consonant was drawn from the set {[f], [x], [m], [s], [ts<sup>h</sup>], [p<sup>h</sup>]}. Vowels were identical and were either [a] or [u] ([i] was not included as neither [si] nor [fi] is phonotactically well-formed in Mandarin). The labial glide [w] was also presented in the stimuli. All the nonce word stimuli were realized with high level tone (tone 1) on both syllables. All ended with the same second syllable (-tā: 搭), and could be represented orthographically as two Simplified Chinese characters. We tested a total of five sets of contrasts. The complete set of 5 items for three [f]~[x] contrasts is shown in Table 1. Two more contrasts [m]~[s] and [ts<sup>h</sup>]~[p<sup>h</sup>] in the context of [a] were also included as control contrast and practice contrast, respectively.

Table 1. Non-word items used for the contrast of /f/~x/ and their respective perceptual assimilation predictions.

Context	L2 Mandarin stimuli	Predicted QSM percept	PAM prediction
/a/	/fātā~xātā/	/h <sup>w</sup> ātā~hātā/	Two-Category
/w/	/fātā~x <sup>w</sup> ātā/	/h <sup>w</sup> ātā~h <sup>w</sup> ātā/	Single-Category
/u/	/fūtā~xūtā/	/hūtā~hūtā/	Single-Category

All stimuli were recorded by four male and two female L1 Mandarin speakers. Each item was embedded in the carrier sentence: “请阅读单词X八遍,” “Please read the word X eight times.” The recordings were made in a sound-attenuated booth at a sampling rate of 44.1 kHz, using a MacBook Air laptop (Apple, Cupertino, CA) with Praat version 6.2 (Boersma and Weenink, 2022), an Audio-technica ATM33a (Audio-technica, Stow, OH) microphone and a USBPre 2 audio mixer by Sound Devices (Reedsburg, WI). Each sentence was read three times. The first author listened to all the recordings and selected the second reading of each speaker to include in the experiment. The second recording (of the three total) was chosen because it sounded more natural, while the first recording included occasional pauses and the third recording was sometimes rushed. All stimuli were segmented and annotated automatically using the Montreal Forced Aligner (McAuliffe et al., 2017). We then verified the output manually and corrected the segmentation in Praat. The peak amplitude of all segmented stimuli was normalized using a Praat script.

2.2 Procedure

The ABX discrimination task was programmed using the library jsPsych (De Leeuw, 2015). In each ABX trial, participants heard a sequence of three stimuli, with an interstimulus interval of 300 ms. The first two stimuli were different from one another, while the third stimulus (X) was a token either of the first stimulus (A) or the second stimulus (B). The A and B tokens were produced by two male speakers, while X was produced by a female speaker. This ensured that each token was produced by different voices, making the X token acoustically distinct from both the A and B tokens. The intention was to avoid a matching decision based solely on the acoustic similarity of the A and B tokens to the X token. The participants’ task was to determine whether X was a token of A or of B. For instance, a participant might hear /fātā<sub>M1</sub>-/x<sup>w</sup>ātā<sub>M2</sub>-/fātā<sub>F1</sub>, to which they should respond “A.” In another trial, they might hear /fātā<sub>M1</sub>-/x<sup>w</sup>ātā<sub>M2</sub>-/x<sup>w</sup>ātā<sub>F1</sub>, to which they should respond “B.”

After each trial, participants were given 3000 ms to respond. If no response was given in this period, the next trial was presented, and the non-response was counted as an error. With one repetition, a total of 128 trials for four contrasts for the main test session were created by combining the stimuli into ABA, ABB, BAA, and BAB trials. The order of all trials was randomized.

Before the experiment began, participants read an information letter describing the study and agreed to participate in the study. They were asked to wear headphones in a quiet space to ensure optimal auditory presentation. The experiment was divided into two parts. In part one, participants were invited to do practice trials of the [ts<sup>h</sup>ātā] ~[p<sup>h</sup>ātā] contrast, meant to familiarize them with the ABX task. They received feedback on their performance (i.e., whether the response they gave was correct or incorrect). There were 16 such trials. In part two, participants moved on to the main experiment, where they no longer received feedback on their responses. There was a self-timed break in the middle of the main test session. At the end of the task, participants filled out a questionnaire about their language background, age, and living place. The entire experiment lasted approximately 20 min and was conducted entirely online.

2.3 Participants

We recruited 63 volunteer participants through the first author’s personal network. All were bilinguals in QSM and Mandarin aged from 18 to 58 years. Mandarin was used as the metalanguage to conduct the experiment.

We set up an exclusion criterion prior to data collection according to a binomial test (using the 0.05 significance level). The result revealed that for an individual participant to show above-chance performance, they had to respond correctly to at least 21 out of the 32 control trials. Participants who did not perform above chance level on the control contrast were considered off task and their data were excluded from analysis (N = 23). We decided on this strict inclusion criterion given that participants took part in the experiment entirely online and we could otherwise not ensure that they were not distracted while doing the task. The data from 40 participants were thus retained for analysis.

3. Result

The mean accuracy of the 40 retained participants was calculated for each contrast. Figure 1 shows the accuracy of the [f]~[x] contrast in three contexts. Accuracy on [fa]~[xa] (mean = 0.808, standard deviation, SD = 0.187) was generally higher than for [fa]~[x<sup>w</sup>a] (mean = 0.688, SD = 0.152) and [fu]~[xu] (mean = 0.684, SD = 0.193). All analyses were performed using logistic mixed-effect models in R using the lme4 library (Bates et al., 2014).

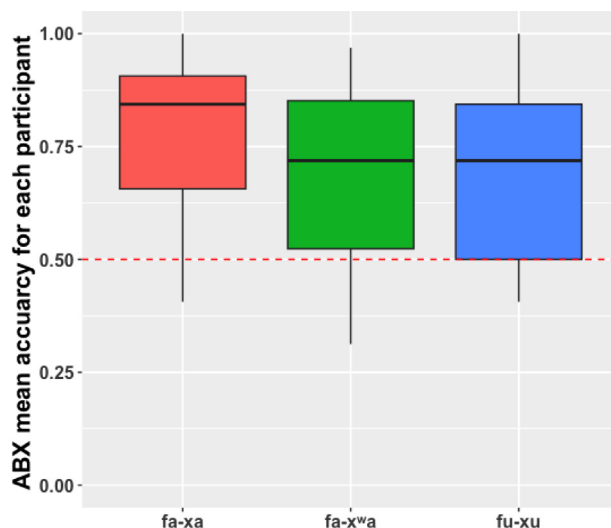


Fig. 1. Proportion correct response on [f]~[x] contrast in three contexts.

We implemented a model including a fixed effect for the factor Context and a random intercept for the factor Participant, including random slopes for the factor Context. In the accuracy scores, the factor Context was found to be a significant predictor ([fa]~[xa] vs [fa]~[x<sup>w</sup>a]:  $\beta = -0.732$ , standard error,  $SE = 0.099$ ,  $z = -7.387$ ,  $p < 0.001$ ; [fa]~[xa] vs [fu]~[xu]:  $\beta = -0.757$ ,  $SE = 0.099$ ,  $z = -7.651$ ,  $p < 0.001$ ).

#### 4. Additional analyses

So far, this study has investigated the extent to which phonological categories between L1 and L2 can predict patterns of L2 perception. However, it is also hypothesized that the acoustic similarity between stimuli may influence participants' decisions, as suggested by [Flege \(1995\)](#). As opposed to relying on phonological indicators, participants may rely on the phonetic similarity of each stimulus. Additionally, [Best and Strange \(1992\)](#) suggest that exposure to L2 input may lead to the reorganization of perceptual assimilation patterns. Several subsequent studies have supported this proposal, demonstrating that increased exposure to an L2 can have a positive impact on L2 perception ([Al Zoubi, 2018](#); [Flege et al., 1997](#); [Jia et al., 2006](#), among others).

Therefore, we additionally investigated the potential impact of phonetic similarity and individual level of bilingualism on L2 perceptual constraints. The aim is to determine to what extent our results could in principle be explained by acoustic distance and additionally to consider the impact that individual levels of bilingualism could have on our results. We will explore these questions in turn.

##### 4.1 Acoustic similarity

To investigate whether the results of our ABX task could be explained solely by the acoustic features of the stimuli, we designed a computational model, following previous work ([Martin and Peperkamp, 2017](#); [Millet et al., 2019](#); [Schatz et al., 2013](#)), which performed our ABX task based exclusively on acoustic distance between stimuli.

For each ABX triplet, the model determined the acoustic distance between A and X and between B and X. If the distance between A and X was smaller than between B and X, it responded A to that trial. If the distance between A and X was larger than between B and X, it responded B. We could then measure the model's accuracy and compare it to the performance of our experimental participants.

For each stimulus, we extracted Mel-frequency cepstral coefficients (MFCCs) using the Librosa library in PYTHON ([McFee et al., 2015](#)). This provided a high-dimensional featural representation of the acoustics of each stimulus which we could use to measure the distance between stimuli. These acoustic features are among the most widely used in modern automatic speech recognition systems ([Cho et al., 2021](#)). We then used the dynamic time warping (DTW) algorithm ([Sakoe and Chiba, 1978](#)), implemented using the DTW-PYTHON library ([Giorgino, 2009](#)), to measure the distance between the featural representations for each stimulus. Our model then used these distances to determine its response to each ABX triplet in our task.

Overall, the model's accuracy was lower than that of the participants in our experiment (mean accuracy of the model = 62%; mean accuracy of the participants = 74%). The model performed better on the [f]~[x] contrast in the context of rounded vowels (e.g., [fu] vs [xu], mean accuracy = 81%) than it did on the same contrast in the context of low vowels (e.g., [fa] vs [xa], mean accuracy = 50%) or in the context of the labial glide (e.g., [fa] vs [x<sup>w</sup>a], mean

accuracy = 56%), indicating that the acoustic distance of the [f]~[x] contrast in the context of rounded vowels was greater than in the other phonological contexts we considered. Importantly, this pattern does not match the pattern we observed in our experimental participants. This suggests, then, that acoustic distance was not the determining factor for the participants in our task. Rather, their response pattern is better captured by their phonological categories, according to the predictions based on the Perceptual Assimilation model as laid out in Sec. 1. We note, however, that our model considered the stimuli as a whole and it is possible that a different analysis that looks specifically at acoustic features known to be associated with fricative perception (e.g., center of gravity) would yield a different result.

#### 4.2 Individual level of bilingualism

Up until now, our findings have been focused on hypothesized language-internal mechanisms. However, individual levels of bilingualism can have a significant impact on perception. To get a measure of L2 exposure and use for our participants, we examined their answers to our post-test language use questionnaire. We determined an overall Mandarin exposure score (ranging from -6 to 6) for each participant based on their self-reported frequency of use of Mandarin. A higher score indicates that the participant has more regular exposure to Mandarin relative to QSM. Participants reported quite a bit of variation in their exposure to and use of Mandarin (mean = 0.48).

We then calculated each participant's performance on the target contrast in the two difficult phonological contexts (i.e., the context of a rounded vowel or a labial glide) compared to the easier phonological context (i.e., the context of a low vowel), using difference scores. Using these two difference scores (difference in accuracy on [fa]~[xa] vs [fa]~[xu] and difference in accuracy on [fa]~[xa] vs [fa]~[x<sup>w</sup>a]), we looked to see if these differences were related to individuals' self-reported exposure to and use of Mandarin. We performed two correlations in R. We found a negative correlation between the Mandarin exposure score and the contrast accuracy difference score for both comparisons: [fa]~[xa] vs [fa]~[x<sup>w</sup>a] ( $R = -0.142$ ,  $\beta = -0.008$ ,  $SE = 0.003$ ,  $t = -2.199$ ,  $p < 0.05$ ) (see Fig. 2) and [fa]~[xa] vs [fu]~[xu] ( $R = -0.137$ ,  $\beta = -0.008$ ,  $SE = 0.004$ ,  $t = -2.12$ ,  $p < 0.05$ ) (see Fig. 3). This implies that individuals with higher levels of exposure to Mandarin exhibited smaller differences in accuracy when comparing contrast types representing Two-category and Single-category, indicating better discriminability of the difficult L2 contrasts. In other words, bilingual speakers' individual L2 exposure level had a positive impact on the discriminability of L2 contrasts (i.e., more exposure = smaller difference score). Participants who reported a high level of exposure to Mandarin demonstrated a greater overall accuracy, reducing the difference in performance between the different phonological contexts.

#### 5. Discussion

In this study, we used an ABX discrimination task to assess the ability of QSM bilinguals to perceive the L2 Mandarin non-sibilant fricatives in three different phonological contexts: in the presence of the unrounded vowel [a], the rounded vowel [u], and the labial glide [w]. For the three contexts we evaluated, we found a significant difference in accuracy by context type, in line with our predictions. QSM listeners had difficulty discriminating the Mandarin non-sibilant fricatives [f] and [x] in the context of the rounded vowel [u] and in the presence of the labial glide [w]. This was despite the fact

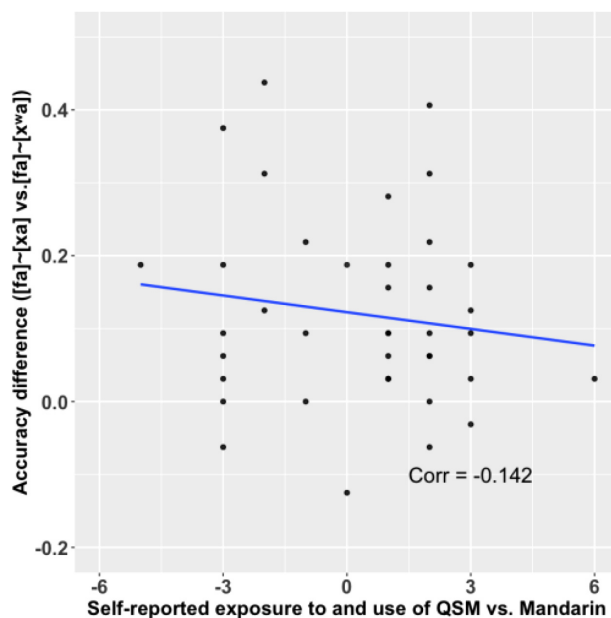


Fig. 2. [fa]~[xa] vs [fa]~[x<sup>w</sup>a] accuracy difference and L2 use/exposure level.

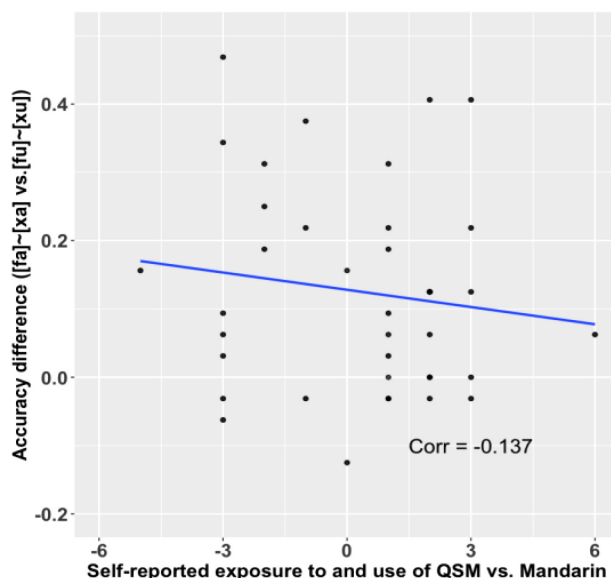


Fig. 3. [fa]~[xa] vs [fu]~[xu] accuracy difference and L2 use/exposure level.

that, according to the analysis described above, the acoustic distances between our stimuli were actually greater in the context of [u] than in the context of [a]: our computational model performed better on trials comparing [fu] and [xu] than on trials comparing [fa] and [xa]. The outcome of our experiment is, however, in line with the predictions of PAM’s performance levels: Two-Category assimilation > Single-Category assimilation, and is also in harmony with previous work highlighting the influence of the vocalic context on obstruent perception (Mann and Repp, 1980) and the influence of labiovelars in particular (e.g., Ohala and Lorentz, 1977).

We additionally examined whether the individual degree of L2 exposure further affected bilinguals’ perception of L2 sounds. We focused on individual-level exposure to Mandarin and found a significant negative correlation between individual-level exposure to Mandarin and differences in accuracy in the different phonological contexts. This result shows that individual-level exposure to L2 is crucial to bilingual speakers’ capacity to discriminate L2 contrasts. The more exposure to the L2, the better individuals are at perceiving L2 contrasts (see also Flege et al., 1997).

This study examined only the perception perspective. An avenue for future work is the inclusion of production data. It would be interesting to consider individual-level production alongside perception: What is the relationship between perception and production? Are the speakers who can better perceive the contrast in Mandarin also the ones who can better produce it? Follow up work may give more precise answers to these questions.

## 6. Conclusion

In conclusion, the results of our present study on the discrimination of the L2 Mandarin [f]~[x] contrast by bilingual speakers of QSM and Mandarin have provided further insight into cross-language transfer with bilingual speakers from L1 to L2, highlighting the phonological rather than the phonetic level. The outcome of the discriminability of the L2 Mandarin fricatives shows that the phonemic context of rounded vowels and the labial glide /w/, as well as individual L2 exposure level have a significant impact on the perception of the L2 fricative contrast. The predictions of the Perceptual Assimilation model provided the theoretical framework for our hypotheses, and indeed Two-Category Assimilation was better discriminated than Single-Category Assimilation.

## Acknowledgements

The authors would like to thank two anonymous reviewers for their valuable feedback and comments. This study received partial funding from the Agence Nationale de la Recherche (grant ANR-10-LABX-0083-LabEx EFL). The authors have no conflict of interest to disclose.

## References

- Al Zoubi, S. M. (2018). “The impact of exposure to English language on language acquisition,” *J. Appl. Ling. Lang. Res.* 5(4), 151–162.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). “Fitting linear mixed-effects models using lme4,” [arXiv:1406.5823](https://arxiv.org/abs/1406.5823).
- Best, C. T. (1993). “Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development,” in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (Springer, New York), pp. 289–304.

- Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience* (Timonium, New York), pp. 171–206.
- Best, C. T., and Strange, W. (1992). "Effects of phonological and phonetic factors on cross-language perception of approximants," *J. Phon.* **20**(3), 305–330.
- Boersma, P., and Weenink, D. (2022). "Praat: Doing phonetics by computer version 6.2.06 [computer program]," <http://www.praat.org/> (Last viewed January 21, 2022).
- Cho, S., Nevler, N., Parjane, N., Cieri, C., Liberman, M., Grossman, M., and Cousins, K. A. (2021). "Automated analysis of digitized letter fluency data," *Front. Psychol.* **12**, 654214.
- De Leeuw, J. R. (2015). "jspsych: A javascript library for creating behavioral experiments in a web browser," *Behav. Res.* **47**(1), 1–12.
- Duanmu, S. (2007). *The Phonology of Standard Chinese* (Oxford University Press, Oxford, UK).
- Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization* (Netherlands Graduate School of Linguistics, Amsterdam, Netherlands).
- Flege, J. E. (1995). "Second language speech learning: Theory, findings, and problems," *Speech Percept. Ling Exp.* **92**, 233–277.
- Flege, J. E., Bohn, O.-S., and Jang, S. (1997). "Effects of experience on non-native speakers' production and perception of English vowels," *J. Phon.* **25**(4), 437–470.
- Flege, J. E., and Fletcher, K. L. (1992). "Talker and listener effects on degree of perceived foreign accent," *J. Acoust. Soc. Am.* **91**(1), 370–389.
- Giorgino, T. (2009). "Computing and visualizing dynamic time warping alignments in R: The dtw package," *J. Stat. Softw.* **31**, 1–24.
- Greenlee, M. (1992). "Perception and production of voiceless Spanish fricatives by Chicano children and adults," *Lang. Speech* **35**(1–2), 173–187.
- Jia, G., Strange, W., Wu, Y., Collado, J., and Guan, Q. (2006). "Perception and production of English vowels by mandarin speakers: Age-related differences vary with amount of L2 exposure," *J. Acoust. Soc. Am.* **119**(2), 1118–1130.
- Kwok, B.-C. (2018). *Southern Min: Comparative Phonology and Subgrouping* (Routledge, New York).
- MacKain, K. S., Best, C. T., and Strange, W. (1981). "Categorical perception of English /r/ and /l/ by Japanese bilinguals," *Appl. Psycholing.* **2**(4), 369–390.
- Mann, V. A., and Repp, B. H. (1980). "Influence of vocalic context on perception of the [ʃ]-[s] distinction," *Percept. Psychophys.* **28**(3), 213–228.
- Martin, A., and Peperkamp, S. (2017). "Assessing the distinctiveness of phonological features in word recognition: Prelexical and lexical influences," *J. Phon.* **62**, 1–11.
- Mazzaro, N. (2011). "Experimental approaches to sound variation: A sociophonetic study of labial and velar fricatives and approximants in argentine Spanish," Ph.D. thesis, University of Toronto, Toronto, Canada.
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). "Montreal forced aligner: Trainable text-speech alignment using Kaldi," in *Proceedings of Interspeech*, August 20–24, Stockholm, Sweden, pp. 498–502.
- McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., and Nieto, O. (2015). "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th Python in Science Conference*, July 6–12, Austin, TX, pp. 18–25.
- Millet, J., Jurov, N., and Dunbar, E. (2019). "Comparing unsupervised speech learning directly to human performance in speech perception," in *CogSci 2019-41st Annual Meeting of Cognitive Science Society*, July 24–27, Montreal, Canada.
- Ohala, J., and Lorentz, J. (1977). "The story of [w]: An exercise in the phonetic explanation for sound patterns," *BLS* **3**, 577–599.
- Peng, S.-H. (1993). "Cross-language influence on the production of mandarin /f /and /x/ and Taiwanese /h/ by native speakers of Taiwanese Amoy," *Phonetica* **50**(4), 245–260.
- Robinett, B. W., and Schachter, J. (1983). *Second Language Learning: Contrastive Analysis, Error Analysis, and Related Aspects* (University of Michigan Press, Ann Arbor, MI).
- Sakoe, H., and Chiba, S. (1978). "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust. Speech Signal Process.* **26**(1), 43–49.
- Schatz, T., Peddinti, V., Bach, F., Jansen, A., Hermansky, H., and Dupoux, E. (2013). "Evaluating speech features with the minimal-pair ABX task: Analysis of the classical MFC/PLP pipeline," in *INTERSPEECH 2013: 14th Annual Conference of the International Speech Communication Association*, August 25–29, Lyon, France, pp. 1–5.
- Strange, W., Bohn, O.-S., Nishi, K., and Trent, S. A. (2005). "Contextual variation in the acoustic and perceptual similarity of north German and American English vowels," *J. Acoust. Soc. Am.* **118**(3), 1751–1762.
- Strange, W., Bohn, O.-S., Trent, S. A., and Nishi, K. (2004). "Acoustic and perceptual similarity of north German and American English vowels," *J. Acoust. Soc. Am.* **115**(4), 1791–1807.