Acquiring new contrasts in L2 vs. L3: Chinese dialectal speakers' acquisition of Mandarin and English phonetic contrasts

Tianqi Ju, Beiwei Gu, Fumo Huang, Xinyi Liu, Cen Yue, Yao Yao

The Hong Kong Polytechnic University
tianqi.ju@connect.polyu.hk, beiwei.gu@connect.polyu.hk, flora.huang@connect.polyu.hk, xinyisein.liu@connect.polyu.hk, cen-mia.yue@connect.polyu.hk, ctyaoyao@polyu.edu.hk

ABSTRACT

A central question in L3 phonology is whether the acquisition of L3 sounds is more influenced by the sound systems of L1, L2, or a combination of the two. The current study investigates two groups of Chinese dialectal speakers' acquisition of new phonemic contrasts in Mandarin (L2) and English (L3). Both dialectal groups (Northeastern, Southwestern) have a neutralized /s/-/sh/ contrast in their native dialects and have to acquire the contrast for L2 Mandarin (/s/-/ʂ/) and L3 English (/s/-/ʃ/). The Southwestern speakers' native dialects also lack the /n/-/l/ distinction, which is present in both L2 and L3. Although an L2-to-L3 transfer would be facilitative in all three cases, we only found definitive evidence for L2-to-L3 transfer in the production of the /s/-/sh/ contrast by Southwestern speakers, mainly among the male speakers. We discuss the results of selective L2-to-L3 transfer with the L2 Status Factor Model and the Scalpel Model.

Keywords: L3 phonology, Chinese dialects, phonetic transfer, L2 Status Factor Model, Scalpel Model.

1. INTRODUCTION

When a learner sets out to acquire a third language (L3), the acquisition could be influenced by previously acquired first (L1) and/or second (L2) languages. Various proposals have been made to address which language system(s) would serve as the primary source of linguistic transfer to the L3: the language typologically closer to the L3 (the Typological Primacy Model (TPM); [1]–[4]), the later acquired language (the L2 Status Factor Model; [5]), the language that would lead to facilitative transfer (the Cumulative-Enhancement Model (CEM); [6]), the language with a similar linguistic property (the Linguistic Proximity Model (LPM); [7]), or the language selected by a number of cognitive and experiential factors (the Scalpel Model; [8]). While both the TPM and the L2 Status Factor Model assume wholesale transfer from one of the existing languages to the L3, the latter three models all posit property-by-property transfer, where different linguistic properties can be transferred from different language systems. The LPM and the Scalpel Model also allow the possibility of non-facilitative transfer, while CEM considers facilitation a necessary condition.

In this study, we examine two groups of Chinese dialectal speakers, who speak a regional Chinese dialect as L1, Standard Mandarin (or "Mandarin" in short) as L2 and English as L3. The two groups differ in the L1 dialect: one group speak Northeastern Chinese dialects, and the other group speak Southwestern dialects. The dialects spoken in Northeastern China (Heilongjiang, Jilin, and Liaoning provinces), which belong to the Mandarin dialectal family, are highly similar to Standard Mandarin in pronunciation except for certain consonants and tones ([9]). Anecdotal evidence suggests that many Northeastern speakers are not aware of the phonetic differences between their native dialect and Standard Mandarin and consider themselves native speakers of Standard Mandarin. Southwestern dialects, on the other hand, are a much more mixed category. The Southwestern speakers in this study mainly come from Sichuan, Hunan, and Hubei provinces, where the regional dialects include both Southwest Mandarin (a member of the Mandarin dialectal family) and non-Mandarin dialects (e.g., Xiang dialects). The Southwestern dialects are overall farther from Standard Mandarin—in pronunciation, lexicon, syntax, *etc.*—than the Northeastern dialects are.

We focus on the acquisition of phonetic contrasts present in both L2 (Mandarin) and L3 (English) but not in L1. Specifically, we examine the acquisition of the /s/-/sh/ contrast by both dialectal groups and the acquisition of /n/-/l/ by the Southwestern group. Regarding the /s/-/sh/ contrast, Standard Mandarin has a phonemic contrast between alveolar and retroflex fricatives/affricates (i.e., /s, ts, ts^{h}/ vs. /ʂ, tʂ, tʂ^{h}/), and English has a similar contrast between alveolar and postalveolar voiceless fricatives (i.e., /s/ vs. /ʃ/). But the contrast is to a large degree neutralized in both Northeastern and Southwestern dialects. Some Northeastern dialects lack the retroflex series altogether, while others may have both series but use them in free variation. Southwestern dialects in general lack the retroflex series. As a result, both Northeastern and Southwestern speakers are known to have difficulty with the /s/-/sh/ contrast when speaking Standard Mandarin and English ([10]–[12]).

In addition, the Southwestern dialects also lack the /n/-/l/ contrast, which is present in both Mandarin and English, leading to difficulty in L2/3 acquisition ([13]).

Since the contrasts under investigation are present in both L2 and L3 but not in L1, influence from L2 would facilitate the acquisition of the contrast in L3, while influence from L1 would not. The potential facilitative transfer from L2 may not necessarily result in higher production accuracy in L3 (compared to L2), because L3 is usually lower in proficiency. However, if the speaker does perform better in L3 than in L2, that would provide definitive evidence for the facilitative transfer from L2 to L3, which is not available during the acquisition of L2. If the direction of transfer is conditioned by typological distance, we expect to see similar results between the two speaker groups, given the similarity of their language backgrounds. If transfer happens in a wholesale manner, results regarding the two phonetic contrasts in Southwestern speakers should be similar.

2. METHODS

2.1. Subjects

Two groups of Chinese speakers participated in this study: 25 from Northeastern China (15F, 10M; 18-30 y.o.) and 28 from Southwestern China (16F, 12M; 18-30 y.o.). All the participants spent at least 15 years before the age of 18 in their home province and had continuous exposure to the regional dialect both at home and in the public when growing up. All the participants identified the regional dialect as their L1. The Mandarin exposure started before the age of six for all participants; English was acquired as a foreign language later in primary and secondary schools. At the time of the study, all the participants had recently (within the past two years) relocated from Mainland China to Hong Kong to study in a university. None of the participants had studied linguistics before; none reported any history of language or speech disorders.

2.2. Stimuli

The critical stimuli include 16 monosyllabic words for each target phoneme (/s/, /sh/, /n/, /l/) in each language (Mandarin, English), totalling 128 critical items. The critical phoneme always appears in the onset. A wide range of vowel contexts (e.g., /a/, /i/, /u/, /ɔ/, /ə/, /ɤ/, /aj/, /aw/, /ow/) loosely shared between Mandarin and English were used to increase variability and facilitate cross-language comparison. One might notice that when followed by a high front vowel like [i], the English /ʃ/ is likely to be assimilated to the alveolo-palatal /ɕ/ in Mandarin, which is acoustically more similar to /s/ than to /ʃ/ ([14]). The current word list has two such English words (*ship*, *sheep*); results reported in the following remain unchanged when tokens of these two words are excluded from analysis. Words with multiple pronunciations and (English) words that were unfamiliar to Chinese learners were avoided.

The 128 critical stimuli contain 26 minimal pairs in the critical phoneme (5-7 pairs per contrast per language); overall, more than 70% of the critical words had a minimal pair neighbour in the critical position in the language. Another 64 monosyllabic words (32 Mandarin, 32 English), which do not contain any of the critical phonemes, were used as fillers.

2.3. Procedure

The experiment was a word reading task conducted with DMDX ([15]) in a soundproof phonetic lab. Mandarin and English word stimuli were presented in separate blocks. In the Mandarin block, the participant saw both the Chinese character and the pinyin form of the target word, in order to minimize confusion about the citation form due to imperfect lexical knowledge of Mandarin; in the English block, only the spelling form was provided. In each trial, the participant was asked to read the word on the screen in a carrier phrase ([wɔ3 kʰan4 taw4] "*I saw* __]" for Mandarin and "*I read* __" for English). The stimuli within a block were presented in a randomized order, with each item occurring twice; the order of language blocks was balanced across participants, with a 5-minute break between the blocks. After completing both blocks, the participant would complete a survey about their demographic and language backgrounds. A complete experimental session took about 40 minutes.

A total of 6400 (64 words × 2 repetitions × 50 speakers) [n/l] tokens and 6400 [s/sh] tokens were collected. Two of the authors independently listened to all the [n/l] tokens in a randomized order and annotated each as [n] or [l] without accessing any information about the speaker or the target item. A small number of the [n]/[l] tokens may sound ambiguous; in cases of annotation discrepancy (N=118), additional judgement was sought from a third author. The [s/sh] tokens were first annotated manually for the critical fricative interval in Praat ([16]); a Praat script adapted from [17] was then used to automatically measure duration and spectral mean (i.e., spectral center of gravity) of each annotated interval. After excluding tokens of misidentified targets or corrupted recordings, 6386 tokens remained in the n/l dataset and 6361 in the s/sh dataset.

2.4. Statistical analysis

The production data are analyzed separately by phonetic contrast (/s/-/sh/, /n/-/l/) and dialectal group (Northeastern, Southwestern). Mixed-effects models are built to examine the production of each contrast by each speaker group: in each model, fixed-effects include target phoneme, language, speaker sex, and their interactions, and random effects include by-speaker and by-item intercepts. The outcome variables are production accuracy (both contrasts; generalized mixed-effects models) and acoustic measures (log duration and spectral mean (converted to ERB scale); the /s/-/sh/ contrast only; linear mixed-effects models). Production accuracy of [s/sh] tokens is coded using a raw spectral mean of 5000 Hz as the boundary between [s] and [sh], following the norms of Mandarin and English sibilants ([14], [18]). All the models are built with the lme4() package ([19]) in R ([20]). Reporting in the next section focuses on language-related effects.

3. RESULTS

3.1. /s/-/sh/ contrast

Both dialectal groups have significantly less accurate productions of [sh] than [s] ($ps <. 001$; see Fig 1). Northeastern speakers tend to produce the /sh/ sound better in Mandarin (mean accuracy: 76.5% (F); 88.1% (M)) than in English (69.7% (F); 87.3% (M)), but the trend does not reach significance ($p >.05$). On the contrary, Southwestern speakers have overall more accurate [sh] productions in English (74.6% (F); 72.4% (M)) than in Mandarin (67.4% (F); 58.6% (M)), but the trend does not reach significance ($p >.05$).

The models on (log) sibilant duration reveal that both dialectal groups significantly lengthen the English sibilants compared to the Mandarin counterparts ($ps < .01$), suggesting greater effort in reading L3 English words. As predicted, Northeastern speakers' model of spectral means shows a significant lowering of spectral means in [sh] compared to [s] ($\beta_{\text{target=sh}} = -4.73$; $p < .001$), and the difference is greater for male speakers than female speakers ($\beta_{\text{target=sh}} = -0.47$; $p = .002$). The /s/-/sh/ difference does not interact with language.

The model of Southwestern speakers' spectral means also shows significantly lower spectral means for [sh] than for [s]. However, the size of the /s/-/sh/ difference interacts with both speaker sex and language. Female Southwestern speakers are predicted to have a difference of around 5.02 ERB in spectral mean between their [s] and [sh] of both languages ($\beta_{\text{target=sh}} = -5.02$; $p < .001$). Male Southwestern speakers have significantly smaller /s/-

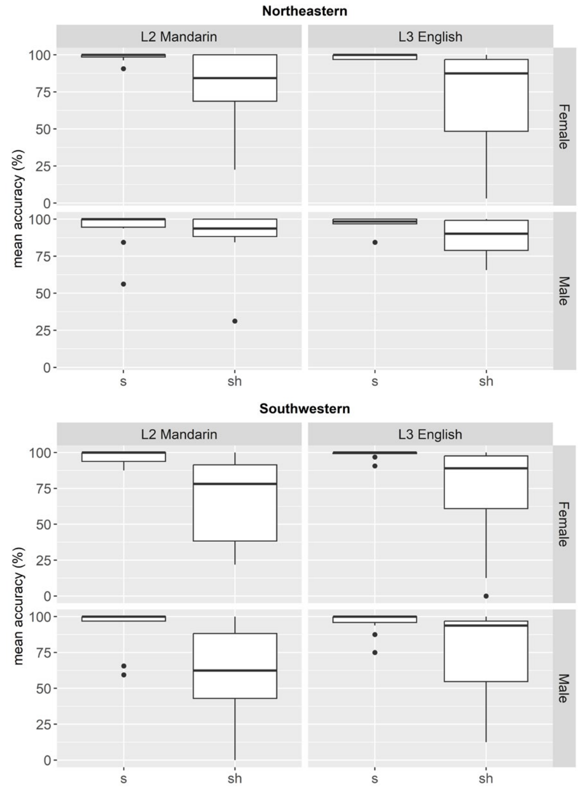

Figure 1: Production accuracy of Northeastern (top) and Southwestern (bottom) speakers' productions of the /s/-/sh/ contrast, separated by language and sex.

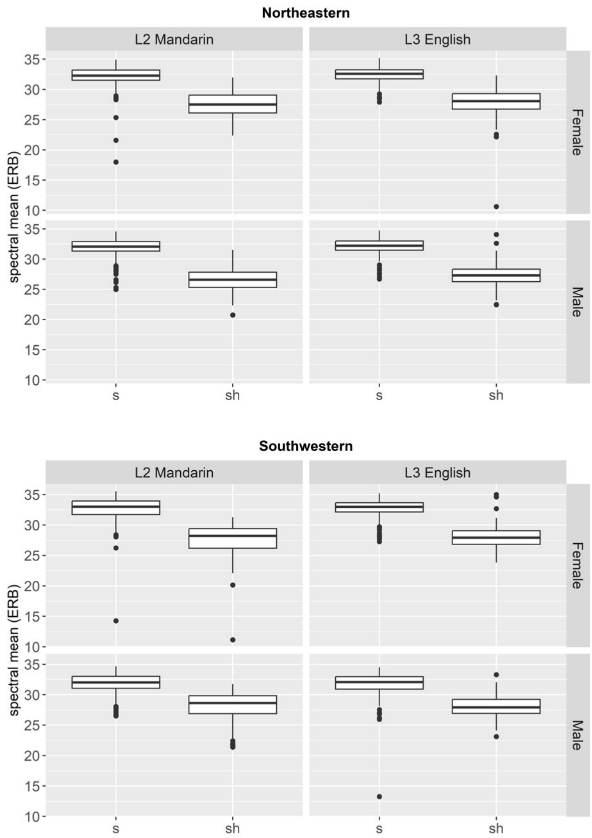

Figure 2: Average spectral means of Northeastern (top) and Southwestern (bottom) speakers' productions of the /s/-/sh/ contrast, separated by language and sex.

/sh/ difference in Mandarin ($\beta_{\text{target=sh:sex=male}}$ = 1.41; $p<.001$) compared to female Southwestern speakers, but they show a marginally significant trend to increase the /s/-/sh/ difference in English ($\beta_{\text{target=sh:sex=Male:language=ENG}} = -0.35$; $p = .06$). Of all four region×sex groups, male Southwestern speakers have the smallest /s/-/sh/ distinction (in terms of spectral means) in Mandarin, which is consistent with their low production accuracy with Mandarin [sh] sound. Meanwhile, male Southwestern speakers are also the only group that show a significantly greater spectral distinction between [s] and [sh] in English than in Mandarin.

3.2. /n/-/l/ contrast

Northeastern speakers, whose native dialects have the /n/-/l/ contrast, produce a clear /n/-/l/ distinction in both languages (mean accuracy > 98%). Southwestern speakers, on the other hand, have more difficulty with the contrast, especially the [l] sound (Fig 3). Both male and female Southwesterners tend to produce the [l] sound less accurately in English (F: 96.3%; M: 82.6%) than in Mandarin (F: 98.2%; M: 89.9%), although the cross-language difference is not significant ($p > .05$).

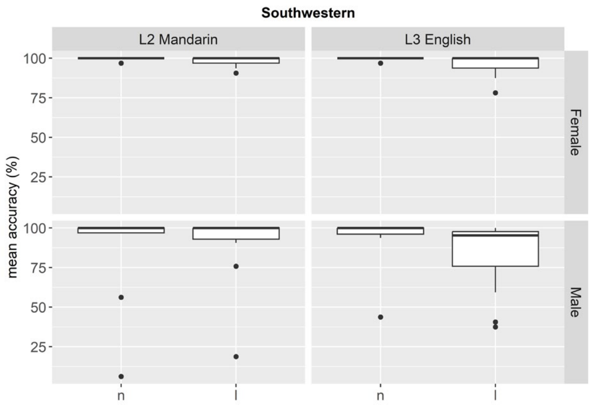

Figure 3: Production accuracy of Southwestern speakers' production of the /n/-/l/ contrast, separated by language and sex.

Taken together, our results show that male Southwestern speakers show a robust pattern of pronouncing the /s/-/sh/ contrast with a greater spectral distinction and higher accuracy in L3 English than in L2 Mandarin. Female Southwestern speakers also tend to pronounce the /s/-/sh/ contrast with higher accuracy in L3 English than in L2 Mandarin, but with no significant spectral enhancement. However, this L3-better-than-L2 pattern is not observed in the other cases examined in this study. Northeastern speakers' production of the /s/-/sh/ contrast shows no evidence of enhancement in L3 over L2, despite the greater articulatory effort as shown in the hyperarticulation of L3 English words. Southwestern speakers' own production of the /n/-/l/ distinction also lacks the L3-better-than-L2 pattern; if anything, the speakers tend to have lower—instead of higher—accuracy of [n/l] tokens in L3 English than in L2 Mandarin.

4. DISCUSSION

As stated above, better performance in L3 than in L2 would provide definitive evidence for facilitative transfer from L2 to L3 in this study. Current results only show facilitative transfer from L2 to L3 in Southwestern Chinese speakers' acquisition of the /s/-/sh/ contrast. There are two possible reasons why Northeastern speakers did not show the L3-better-than-L2 pattern. First, Southwestern speakers are more likely to recognize Mandarin as an L2, whereas Northeastern speakers may not consider Mandarin as distinctly different from their L1, given the close resemblance between the two. Following the L2 Status Factor Model, Southwestern speakers are thus more likely to transfer from Mandarin to English than Northeastern speakers. Second, maybe the Northeastern speakers in this study happen to have lower English proficiency, which would cancel out the facilitation from the transfer from Mandarin. However, male Northeastern speakers actually had much higher /s/-/sh/ production accuracy (>87%) than Southwesterners (52.5-77.5%) in both languages, which undermines the account based on L3 proficiency.

Another question regards why we didn't see the L3-better-than-L2 pattern in Southwestern speakers' production of the /n/-/l/ contrast. Obviously, this cannot be explained by L2 status, L3 proficiency or typological proximity between languages. We think that the answer is probably related to some linguistic, cognitive, or experiential factors, following the Scalpel Model. The discrepancy between the /s/-/sh/ contrast and the /n/-/l/ contrast in Southwestern speakers' data may be explained by articulatory or perceptual features of the two contrasts or the speaker's awareness about these contrasts in the L2 and L3.

To summarize, our results showcase the selectivity of facilitative transfer from L2 to L3 in L3 phonology. Our findings are most compatible with the L2 Status Factor Model and the Scalpel Model. We hope to further investigate the effects of cognitive and experiential factors in L3 acquisition in future research.

7. REFERENCES

[1] Rothman. J. 2011. L3 syntactic transfer selectivity and typological determinacy: The typological

primacy model. *Second Lang. Res.*, 27 (1), 107–127.
[2] Rothman, J. 2010. On the typological economy of syntactic transfer: Word order and relative clause high/low attachment preference in L3 Brazilian Portuguese. *IRAL - Int. Rev. Appl. Linguist. Lang. Teach.*, 48 (2–3), 245–273.
[3] Rothman, J. 2013. Cognitive economy, non-redundancy and typological primacy in L3 acquisition: Initial stages of L3 Romance and beyond. In: S. Baauw, F. Drijkoningen, L. Meroni, and M. Pinto (eds) *Romance Languages and Linguistic Theory 2011. Selected papers from 'Going Romance' Utrecht 2011*. John Benjamins, 217–247.
[4] Rothman, J. 2015. Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Biling. Lang. Cogn.*, 18 (2), 179–190.
[5] Bardel, C. and Falk, Y. 2007. The role of the second language in third language acquisition: The case of Germanic syntax. *Second Lang. Res.*, 23 (4), 459–484.
[6] Flynn, S., Foley, C., Vinnitskaya, I. 2004. The Cumulative-Enhancement model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *Int. J. Multiling.*, 1 (1), 3–16.
[7] Westergaard, M, Mitrofanova, N., Mykhaylyk, R. Rodina, Y. 2017. Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *Int. J. Biling.*, 21 (6), 666–682.
[8] Slabakova, R. 2017. The scalpel model of third language acquisition. *Int. J. Biling.*, 21 (6), 651–665.
[9] Liu, B. 2009. *Shenyang fangyan de yuyin fenxi* [Phonetic analysis of Shenyang dialect]. *Wenjiao Ziliao*, 2009.April, 54–55.
[10] Gong, C, Tian, Y. 2008. *Dongbei fangyan dui yinghan yuyin xide chansheng de yingxiang* [Influence of Dongbei dialect on the acquisition of English and Mandarin pronunciations]. *Jiaoyan Tansuo*, 10, 30–31.
[11] Zhang, L. 2018. An acoustic study on the English fricative production by Chinese Northeastern EFL learners. Thesis.
[12] You, S., Kan, P., Wu, J.-Y., Yang, Z.-Y. 2009. On the language transfer of Chinese dialects to English pronunciation: Comparison of Jinan, Chengdu and Pingxiang Dialects. *J. Gannan Norm. Univ.*, 2009.5, 54–58.
[13] Zhang, W., Levis, J. M. 2021. The Southwestern Mandarin /n/-/l/ Merger: Effects on Production in Standard Mandarin and English. *Front. Commun.*, 6 (August), 1–22.
[14] Chang, C. B., Yao, Y., Haynes, E. F., Rhodes, R. 2011. Production of phonetic and phonological contrast by heritage speakers of Mandarin. *J. Acoust. Soc. Am.*, 129 (6), 3964–3980.
[15] Forster, K., Forster, J. C. 2003. DMDX: A Windows display program with millisecond accuracy. *Behav. Res. Methods, Instruments, Comput.*, 35 (1), 116–124.
[16] Boersma, P., Weenink, D. 2022. Praat: Doing phonetics by computer.
[17] DiCanio, C. 2021. Spectral moments of fricative spectra script in Praat.
[18] Jongman, A., Wayland, R., Wong. S. 2000. Acoustic characteristics of English fricatives. *J. Acouistical Soc. Am.*, 108 (3), 1252–1263.
[19] Bates, D. M., Maechler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.*, 67, 1–48.
[20] R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.