International Journal of Language and Linguistics
2022; 10(2): 103-110
http://www.sciencepublishinggroup.com/j/ijll
doi: 10.11648/j.ijll.20221002.16
ISSN: 2330-0205 (Print); ISSN: 2330-0221 (Online)

The Influence of Dialect on the Perception and Production of Lax-Tense Vowel Distinction in English Learning

Jie Ji*, Aitong Jiang

Department of English and International Studies, China Foreign Affairs University, Beijing, China

Email address:
jijie@cfau.edu.cn (Jie Ji)
*Corresponding author

To cite this article:
Jie Ji, Aitong Jiang. The Influence of Dialect on the Perception and Production of Lax-Tense Vowel Distinction in English Learning. *International Journal of Language and Linguistics.* Vol. 10, No. 2, 2022, pp. 103-110. doi: 10.11648/j.ijll.20221002.16

Received: February 20, 2022; **Accepted**: March 9, 2022; **Published**: March 18, 2022

Abstract: This acoustic experimental study investigates the influence of dialectal background on the perception and production of lax-tense vowel distinction in English learning from the theoretical standpoint of language transfer. Previous studies usually regard the first language as a source of transfer, with few considering the influence of dialect in the process of transfer, while this study has taken account of participants' dialects and biological genders and look at whether and how Chinese dialectal knowledge is transferred in English learning. A perception and a production experiment are conducted with two groups of participants -- Cantonese Chinese speakers and Mandarin Chinese speakers -- to respectively analyze their perception and production strategy for English lax-tense vowel pairs [I]/[i:] and [ʊ]/[u:]. The study finds out that dialect and gender cause statistically significant difference. The result shows that Cantonese speakers can effectively leverage spectral cues to differentiate English tense vowels from lax vowels, while Mandarin speakers rely heavily on durational cues. The two lingual groups have disparate production result, but no one group is overall better than the other in producing lax-tense vowels. Mandarin and Cantonese participants only differ in the F2 of the [u:] production. Moreover, when gender is considered, Mandarin females can produce native-like [ʊ]; Cantonese males and females can produce native-like [I], Cantonese females native-like [i:] and Cantonese males native-like [u:]. This study confirms that dialect should be considered in transfer study, and further points out that both dialect and gender are significant variables of the transfer mechanism in foreign language acquisition.

Keywords: Transfer, Dialect, Speech Production, Speech Perception

1. Introduction

When a language learner advances to a certain level, his or her accent seems to freeze, not continually growing to become native-like. This phenomenon is known as fossilization, and the accented language is called interlanguage. What is curious about fossilization is that it seems to take on certain patterns, which allows us to differentiate French English to a Japanese one. A perspective to look at this crystallization of language is through the lens of transfer theory.

Language transfer -- the transfer of a feature from one acquired language to another new language -- is a universal step during language acquisition. Transfer happens in all aspects of language learning, from grammar, phonetics, to discourse, syntax, etc. Transfer theory is in turn developed over time to investigate the conditions and outcomes of language transfer. Previous findings show that not all transfers yield favorable language learning results. While positive transfer can help learners get the hang of the new language by leveraging current knowledge of other language(s), negative transfer hinders the process due to dissimilarity or confusing similarity of languages. These studies usually regard L1 (the first language) as a source of transfer, with few considering the influence of dialect in the process of transfer; fewer studies look at whether and how Chinese dialectal knowledge is transferred in English learning.

To understand how Chinese dialects participate in English learning, this study conducts an acoustic experiment on perception and production, to see whether Cantonese, a southern Chinese dialect that contains the lax-tense vowel pairs [I]/[i:] and [ʊ]/[u:] similar to those in English, transfers into English production and perception of Cantonese speakers. Another group of speakers from northern regions of China, where the lax-tense vowel pairs does not exist in their native Mandarin, are also invited to evaluate the effect of dialect transfer.

This article starts with literature review of transfer studies of speech sounds, and then moves to introduce the experiment design. Results and discussion are followed with reflection as well as future research directions.

2. Previous Studies

2.1. Language Transfer of L1 Speech Sounds

As early as 1963, Ausubel asserted that "all learning involves some kind of transfer" [1]. Transfer studies in second language acquisition started in 1950s and since produced theories like Contrastive Analysis Hypothesis, Speech Learning Model, Perceptual Assimilation Model, Second Language Linguistic Perception Model.

Contrastive Analysis Hypothesis used to attribute all errors to transfer, but later a more moderate version was brought up by Oller and Ziahosseiny [2]. The moderate Contrastive Analysis denotes that, similar patterns in second language, rather than dissimilar ones, acquire extra effort to internalize.

In the same vein, Fledge discovered that it is harder for experienced L2 (second language) learners to produce L2 phones similar to existing ones in L1, through a mechanism he referred to as "equivalence classification" [3]. This mechanism would make L2 learners unable to differentiate acoustically different approximants in both perception and production. Fledge later came up with Speech Learning Model to suggest that new phones are much easier to acquire in an authentic way than similar ones are. Study show that learners directly map L2 phones that have a similar but not identical counterpart in L1 to the latter, rather than create a new category for the novel phones [4].

Best and Tyler also developed Perceptual Assimilation Model (PAM) to predict L2 assimilation and discrimination [5]. Instead of focusing on individual phones, PAM investigates vowel and consonant contrasts. PAM sorts L2 learning scenario based on the relationship between L1 and L2. According to the PAM model, if a L2 contrast contains one novel phone, then the contrast is expected to be discriminated well; the scenario is dubbed "uncategorized-categorized assimilation (UC)". If a L2 contrast already exists in L1, then it is a "two-category assimilation (TC)", where learners can most accurately distinguish the contrast.

Different from the above models, the Second Language Linguistic Perception (L2LP) model assumes that L2 learners map all L2 sounds to L1 at the beginning of L2 learning, before distinguishing the L2 contrasts [6]. The model has proposed three scenarios to frame L2 perception development: NEW scenario (SC in PAM, when a L2 contrast is mapped to only one L1 category), SIMILAR scenario (TC in PAM) and SUBSET scenario (UC in PAM). It also predicts that the SIMILAR scenario is easiest, and the NEW scenario the hardest, and the SIMILAR scenario requires less effort than the NEW scenario.

2.2. Dialects and Speech Production/Perception

In the above theories, L1 is deemed homogenous, and transfer context is usually set under a simple contrast between L1 and foreign languages. But in an experiment with Iberian and Peruvian Spanish speakers [7], Escudero and Williams took account of the dialectal influence of L1 and found that the acoustic properties of dialect alone could correctly predict participants' perception of the L2 Dutch vowel contrasts, and that the L2 proficiency of participants could not.

O'Brien and Smith [8] have demonstrated that "a speaker's L1 dialect does indeed play a role in the acquisition of L2 sounds" through showing that the production of German [u:]-[y:] contrast by native North American English speakers from three distinct dialect backgrounds varies. Falahuddin and Fitriati [9] have investigated into the influence of mid-east Sundanese dialect in the pronunciation of English. Chládková and Podlipský [10] finds that the more spectrum-dependent Bohemian Czech tend to identify the Dutch [I] as [i:], while the more duration-dependent Moravian Czech as [I]; Moravian Czech also tend to identify both vowels in the /y-Y/ as short vowels, while nearly half of the Bohemian Czech would perceive them to be long vowels. Miller and Grosjean [11] find that monolingual native speakers of Swiss French use both spectral and temporal information to identify manipulated /o/-/ɔ/ contrast, while monolingual native speakers of standard French use only spectral cues. Unlike the phonological system of Swiss French, that of standard French does not stress durational differences. They thus point out that a listener's overall phonological system dictates the way he/she uses phonetic cues and that dialect differences should not be underestimated.

There are few studies investigating the influence of dialectal background on Chinese English-learners' speech production and perception. Hsueh Chu Chen [12] has studied the acoustic timing patterns and perceptual assessment of Chinese learners from three different dialectal backgrounds -- Hong Kong Cantonese, Taiwan Chinese, and Beijing Chinese; he finds that the Hong Kong group perform significantly better in unstressed syllable duration, while the Beijing group perform significantly closer to the native speaker group in speed rate and pause. Alice Y. W. Chan [13] has conducted perception experiments on 40 advanced Cantonese English learners and finds that they have minor problems identifying the English /i:, I/ and /u:, ʊ/ contrasts.

3. Method and Materials

3.1. Research Design and Questions

This study aims to investigate whether dialectal background affects Chinese learners' perception and production of English lax-tense vowels. The study will invite two groups of participants: one group grow up in southern Guangdong province speaking Cantonese, while the other grow up in northern China speaking Mandarin. Although Cantonese and Mandarin are both Chinese dialects, they differ hugely in pronunciation. The English lax vowels [I] and [ʊ] do not exist in both Mandarin and Cantonese phonological systems as phonemes, but they are recognized as allophones in Cantonese that precede nasals and plosives [14]. For example, "[sIk]" and "[sʊk]", in Cantonese meaning "color" and "uncle", have [I] and [ʊ] before the plosive coda [k]. By contrast, there are no phones similar to English lax vowels in Mandarin; it only has the tense [i:] and [u:]. It seems, according to the L2LP model in the last section, that the English lax-tense vowel contrasts are the NEW scenario in Mandarin while the SIMILAR scenario in Cantonese.

Therefore, this study intends to address the following questions:

1. Perception: Will Cantonese speakers perceive lax-tense vowels more accurately than Mandarin speakers? What are the dominant perceptual strategies of identifying [I]/[i:] and [ʊ]/[u:] deployed by Cantonese and Mandarin speakers respectively?
2. Production: Will Cantonese speakers produce lax-tense vowels more precisely than Mandarin speakers? Will speakers from opposite genders perform differently?

3.2. Participants

The participants are sixteen English Majors in their 3rd and 4th year at college. Eight of them speak Cantonese, while the other eight speak Mandarin and cannot speak or understand Cantonese. To avoid potential interference, the selection of Mandarin-speaking participants follows these criteria: 1) those who were born and raised in northern region, or whose family come from such background; 2) knowing little Cantonese and cannot understand native Cantonese conversation; 3) use Mandarin on a daily basis and have a sound grip of standard Mandarin pronunciation.

3.3. Acoustic Terms

Acoustic experiments cannot stand alone without the concept of "frequency", "source-filter theory" and "formants", so this part is devoted to explaining these concepts and clarifying how they are used in this study.

To start with, sound is made up of repeated sound waves. The frequency of a sound refers to how many times in a second a sound wave repeats itself. According to the source-filter theory, a speech sound boils down to a source sound and a filter component acoustically. The source sound is generated by speakers' vocal fold vibration, which determines the pitch, loudness, and texture of sounds. Then, the source sound is filtered by speakers' vocal tracts above the vocal fold, which attenuates certain frequencies and enhance others. This process gives rise to peaks and valleys in the spectrum of the sound (a graphic representation of sound, which presents amplitude at each frequency), with peaks ("local energy maxima") known as formants [15]. In other words, a formant is a concentration of acoustic energy around a particular frequency in the speech wave. The frequency of the center of the formants are known as formant frequencies.

The two quintessential indicators of this paper's experiments -- F1 and F2, meaning first formant and second formant frequency respectively -- are the first two peaks in the spectrum. F1 and F2 are sufficient in describing the location of vowels, even though more formants are required to synthesize vowels. And they can reflect the frontness and height of vowels. Higher the F1, lower the tongue body; higher the F2, further front the vowel. Therefore, F1 and F2 are used in this experiment to interpret speakers' vowel production strategies. The duration, intensity, and formants of sound, among other properties of sound, can be readily obtained with Praat, a widely-used speech analysis program written by Boersma and Weenink from University of Amsterdam [16].

3.4. Experiment Design

This study consists of two experiments -- perception experiment and production experiment of English lax-tense vowel pairs [I]-[i:] and [ʊ]-[u:].

3.4.1. Perception Experiment

In the perception study, researcher first used Praat to synthesize natural tokens of "sheep" and "who'd" respectively into 6 spectrally linear steps. Natural tokens of "sheep", "ship", "who'd" and "hood" were first extracted from online pronunciation instruction on lax and tense vowels by native speakers. Next, the researcher manually measured F1, F2, F3 and F4 of the natural tokens and then calculate the values of the middle 4 steps. The F1, F2, F3 and F4 values of the natural "sheep" are respectively 504 Hz, 1504 Hz, 3254 Hz and 4068 Hz; those of "ship" are 695 Hz, 2128 Hz, 2725 Hz, 3509 Hz; of "who'd" are 400 Hz, 1429 Hz, 2767 Hz, 3934 Hz; of "hood" are 602 Hz, 1636 Hz, 2964 Hz, 4091 Hz. Therefore, for the "sheep/ship" pair, the differences in formant values of each consecutive tokens are 38.2 Hz, 117.6 Hz, -105.8 Hz and -111.8 Hz; for the "who'd/hood" pair, they are 40.4 Hz, 41.4 Hz, 39.4 Hz and 31.4 Hz. Finally, the 4 formant values of "sheep" and "who'd" were manipulated with Praat Vocal Toolkit to become a "sheep/ship" or a "who'd/hood" continuum of 6 steps. Notably, pitch and intensity remain the same as those of "sheep" and "who'd", as it is assumed that it is the vowel quality and duration that serve as acoustic cues to English tense-lax vowel contrasts.

From the 6 spectral steps, the researcher extracted the first, second and third synthesized tokens -- they will be dubbed as "repeated tokens" thereafter—and resynthesized them into 5 steps. Durational steps were manipulated with Praat manually and only the duration of vowel voicing was changed. The duration continuum is made up of 0.1×, 0.3×, 0.5×, 1× and 2× speed version. The reason why the rest 3 spectral steps were excluded is that identification becomes obvious and easy when their duration is manipulated, but the study would like to put participants to test under an extreme scenario. Thus, the duration continuum is made up of 3 spectral steps, identical to the first three in the spectral continuum, and each spectral step has 5 manipulated temporal steps.

Tokens were joined with respect to how they have been manipulated, i.e., spectrally or durationally. Random number sequences were generated by an online simulator to determine the sequence of tokens. Thus, 4 recordings of randomly played tokens (2pairs*2categories) have been created. In each spectral continuum, there are 6 tokens; in each durational continuum, there are 15 (5duration*3spectra).

Participants received a PowerPoint slide containing 4 online surveys and 4 embedded recordings alongside instruction. Each online survey corresponds to one recording. Two are respectively devoted to durational continuum of "sheep/ship" and "who'd/hood", while other two to spectral continuum of "sheep/ship" and "who'd/hood". In the first two, each token is separated with a 10-second blank audio, when participants are expected to think about what they have just heard and then take a short break. As for the latter two, since there are only 6 tokens within, the breaks are shortened to 5 seconds.

Phonemic transcription of "who'd" and "hood" is displayed once in the slides, in case the participants would mistake one for the other; it is at the same time assumed that the participants know the pronunciation of "sheep" and "ship", and so the phonemic transcription is not provided based on the principle of minimizing interference that may come from participant's speculation.

3.4.2. Production Experiment

As for production experiment, participants were asked to record tokens with their phone recorders, which have a sampling rate of 440khz. Each vowel was repeated twice, and there are 8 tokens collected from each participant. After the recorded tokens were sent to the researcher, they were annotated, and their average F1 and F2 values were extracted with Praat script.

1. Participants were instructed to use the following script for warm-up before they recorded:
2. And it's lunch time; but I am still full,
3. That's not good!
4. So I sit in a booth
5. with a book about a cook that always wear a hood.
6. Fix Mike's kite, feed Meg's hen.
7. Mick's men met Mike's team.

Participants were asked to record the following words for vowel collection: good, booth, zoo, book; fix, feed, Mick's, team. Phonetic script was given in order to cross out the possibility that the participant do not know the correct pronunciation of these words.

The F1, F2 and duration of production is then first subject to Independent Samples Test to examine inter-group difference; then, biological sex is taken into account and One-Sample T-Tests are used to compare each gender's production to that of native speakers.

4. Results and Discussion

4.1. Perception

4.1.1. Perception of Sheep/Ship Spectral Continuum

In the identification of "sheep/ship" contrast, Cantonese speakers demonstrate a clear pattern of weighting on spectrum, as seen in Figure 1. Some of them report hearing "sheep" throughout; the others would report "sheep" for the first two steps and then their answer would flip to "ship" for certain steps and onwards.

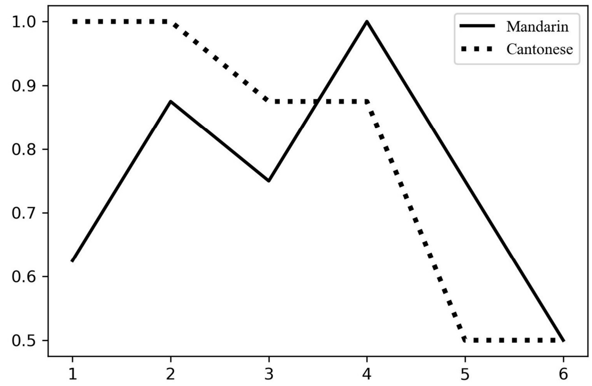

Figure 1. *Perception of Sheep/Ship Spectral Continuum.*

Mandarin speakers, either from an individual or a group level, are messier. Only two participants show the tendency to use spectral cues. Three out of eight have mistaken the first step, which is the natural token of "sheep", for "ship". Meanwhile, all agree that the 4th step should be "sheep". Individuals from Mandarin group seem to have randomly picked answers.

4.1.2. Perception of Who'd/Hood Spectral Continuum

In identifying the "who'd/hood" contrast, the result is more complicated, as seen in Figure 2. Although Cantonese have adopted spectral cue in general, one participant seems to be confused by the final spectrum. It is also obvious that Mandarin speakers are less sensitive to this contrast, as three of them report hearing "who'd" while the spectra are closer to those of "hood" acoustically, and three of them show random pattern.

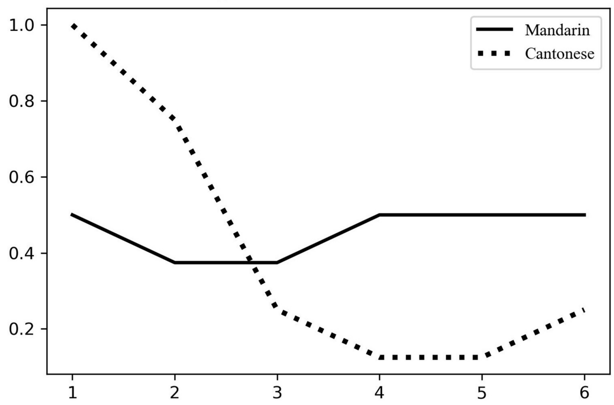

Figure 2. *Perception of Who'd/Hood Spectral Continuum.*

With a closer look, six out of eight in Cantonese group have shown reliance on spectral cues as their report gradually shift to "hood" as the manipulation of vowel quality shifts to "hood". Two participants, one male and one female, have contributed to the general deviation at the final stage in Cantonese group. The male is very sensitive to spectral cues, reporting "hood" except for the first and the last spectrum. His sharp turn is likely to have resulted from lack of knowledge of the American "hood", which is more fronted than "hood" in British English; maybe he has not paid attention to the American pronunciation of [ʊ], therefore, he is confused about what he heard and randomly picked one. The female, in contrast, shows little sensitivity to vowel quality, choosing "who'd" for every token; in fact, she also chose "sheep" for every token in the "sheep/ship" contrast.

For Mandarin group, three out of eight have shown what can be called as "reverse reliance" on spectral cues -- their reports gradually shift from "hood" to "who'd", while the spectra are actually transitioning from "who'd" to "hood". Another participant locates the "who'd" in the 3rd and 4th steps and rendering other steps to be "hood".

To sum up, Cantonese demonstrate good sense of vowel quality in spectral continuum perception overall, while Mandarin do not.

4.1.3. Perception of Sheep/Ship Temporal Continuum

Figure 3 shows that Mandarin group take good advantage of the temporal cue to discern the phones. Controlled for vowel quality, the result shows that in fact, both groups tend to identify tokens as "sheep" when duration is 1× and 2×, as "ship" when duration is shorter. However, Mandarin group show stronger such tendency. Remember that the first three spectral steps are also part of durational experiment, and their identification rates for the repeated tokens increase from 62.5%, 87.5%, 75% in Figure 1 to 87.5%, 87.5% and 87.5% in Figure 3. By contrast, Cantonese group have the same identification rate, displaying stable spectral reliance.

Furthermore, four Mandarin participants have switched their answers for the first spectral step, while three participants do so for the third spectral step and two for the second; meanwhile, only two in the Cantonese group choose different answers from those in 4.1.1. The change takes place in the third spectral step, which has a very nuanced spectrum that stands in between "ship" and "sheep"; considering the durational interference, this change is not interpreted as failure to use spectral cues.

4.1.4. Perception of Who'd/Hood Temporal Continuum

In Figure 4, the identification rates of "who'd" of Mandarin group for the repeated tokens are 62.5%, 50% and 50%, while their identification rates of repeated tokens in Figure 2 drops to 50%, 37.5% and 37.5%. Interestingly, Cantonese group also show disparate identification rate. Their identification rates of "who'd" increases from 87.5%, 50% and 12.5% in Figure 4 to 100%, 75%, 25% in Figure 2. In fact, six out of eight Cantonese participants have switched their answers for the second spectral step, while only one person does so for the third spectral step and none for the first step.

In sum, Mandarin speakers tend to use a combination of duration and spectrum in perception, while Cantonese tend not to. That said, temporal cues have great influence on participants' perception. Moreover, the perception of "who'd/hood" seems more challenging to both groups of participants.

4.1.5. Discussion

In the spectral continuum tests, Cantonese participants are efficient in using spectra to differentiate the lax-tense vowels, while their counterparts do the lexical decision without a clear strategy. Moreover, Mandarin speakers tend to locate the tense vowels at the 3rd and 4th step. While it could well be a coincidence, as they do not utilize vowel quality in perception, it is also probable that they actually identify [i:] and [u:] with higher F1 and F2, that is, with a lower and anterior vowel space. Besides, both Mandarin and Cantonese are less sensitive to the "who'd/hood" contrast than the "sheep/ship" one.

In the durational continuum test, Cantonese speakers use combined strategy, while Mandarin speakers show high propensity to use temporal information.

In sum, Cantonese speakers are better at using spectral cues in perception. This demonstrates that dialectal knowledge can positively transfer into L2 language.

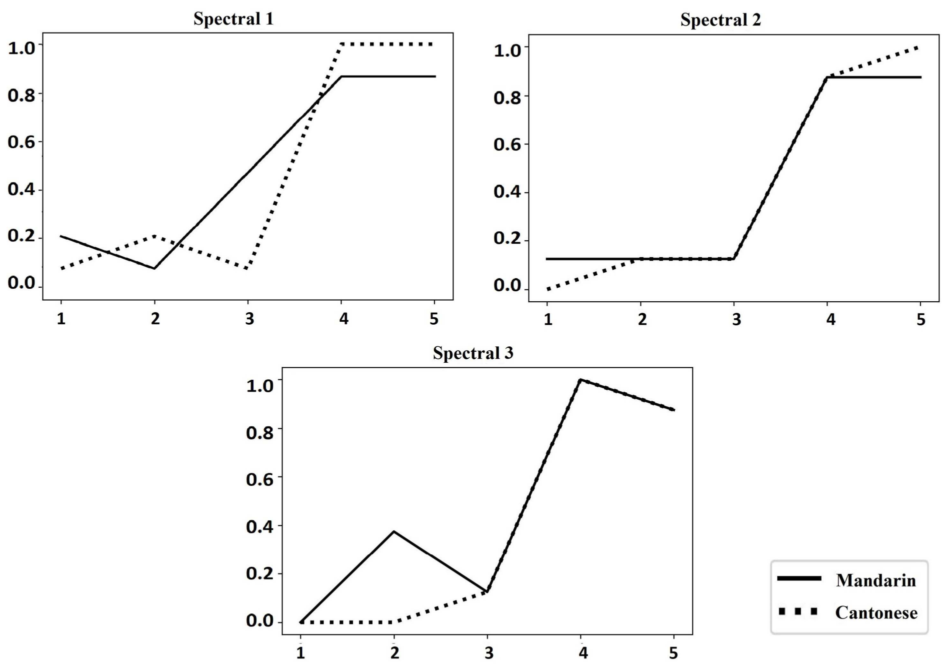

Figure 3. *Perception of Sheep/Ship Temporal Continuum.*

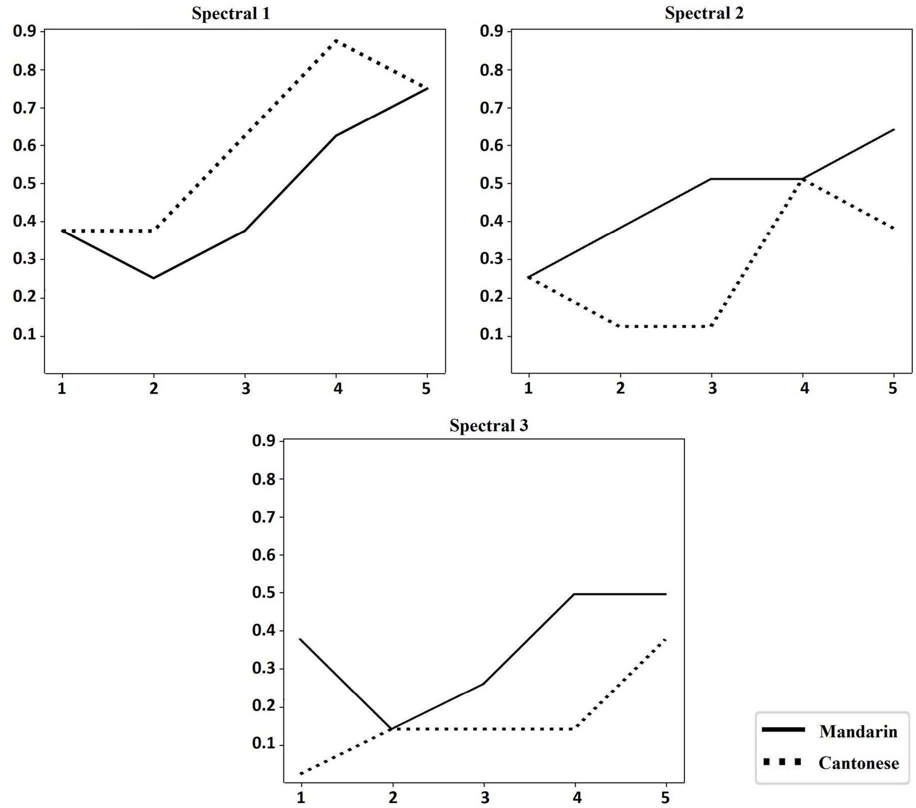

Figure 4. *Perception of Who'd/Hood Temporal Continuum.*

4.2. Production

4.2.1. Production of Cantonese and Mandarin Groups

The result of Independent Samples Test, as seen in Table 1, indicates that in a group level, there is no significant difference between the Cantonese and Mandarin groups, with respect to duration in two groups' production of lax-tense vowels (P>0.05). In terms of F1, statistical significance is also not found between two groups' production (P>0.05).

The only difference found in a group level is that of F2 of the tense [u:] (P=0.012<0.05), with Cantonese group producing [u:] of higher F2 than Mandarin group (means of Mandarin=1199.5Hz; means of Cantonese=1493.25Hz). In other words, Cantonese speaker would pronounce [u:] in a more fronted position compared to the Mandarin group.

In sum, no group inclines to use duration as a production strategy more than the other, and no one group is overall better than the other in producing lax-tense vowels.

Table 1. *Independent Samples Test result*[1] *on Duration, F1 and F2.*

	Levene's Test for Equality of Variances			T-test for Equality of Means		
		F	Sig.	t	Sig. (2-tailed)	Mean Difference
I-duration	Equal variances assumed	3.714	0.075	-0.895	0.386	-0.01400
i:-duration	Equal variances assumed	0.288	0.600	-0.590	0.565	-0.01888
ʊ-duration	Equal variances assumed	0.274	0.609	-1.169	0.262	-0.02856
u:-duration	Equal variances assumed	0.256	0.621	-1.971	0.069	-0.07513
I-F1	Equal variances not assumed	5.311	0.037	-0.848	0.415	-39.00000
i:-F1	Equal variances assumed	0.849	0.372	-1.859	0.084	-45.06250
ʊ-F1	Equal variances assumed	0.005	0.944	-0.733	0.476	-24.87500
u:-F1	Equal variances assumed	0.337	0.571	-1.617	0.128	-35.25000
I-F2	Equal variances assumed	0.395	0.540	-0.522	0.610	-50.50000
i:-F2	Equal variances not assumed	6.015	0.028	0.592	0.567	111.50000
ʊ-F2	Equal variances assumed	4.009	0.065	-1.614	0.129	-156.87500
u:-F2	Equal variances assumed	0.285	0.602	-2.898	0.012*	-293.75000

* Sig. (p value)<0.05

1 The original output in the Independent Samples Test table includes two rows: "Equal variances assumed" and "Equal variances not assumed". If Levene's test indicates that the variances are not equal across the two groups (i.e. p<0.05), only "Equal not variances assumed" line for the t test is reserved; otherwise only the "Equal variances assumed" line is reserved in this table for simplicity. We also delete the "df", "Std. Error Difference" and "95% Confidence Interval" rows of the original SPSS output.

4.2.2. A Close Look at Different Genders

However, if taking biological sex into account, the data tells a different story. We divide the speakers into four groups: Cantonese male, Mandarin male, Cantonese female, and Mandarin female, and then cross-examine their production of lax-tense vowels with that of native speakers' using One-Sample t-Tests. Only their F1 and F2 are examined here; as many participants speculated on the purpose of experiment and intendedly prolonged the duration, their duration values are deemed unsuitable for this comparison. The F1 and F2 of American native speakers, as shown in Tables 2 and 3, are retrieved from *Journal of the Acoustical Society of America* [17].

Table 2. *One Samples Test result of male.*

	Native male	Cantonese male		Mandarin male	
		t	p	t	p
I-F1	427	1.838	0.207	-6.582	0.007**
I-F2	2034	-1.098	0.387	0.819	0.473
i:-F1	342	-0.133	0.906	-4.208	0.025*
i:-F2	2322	-5.676	0.030*	0.285	0.794
ʊ-F1	469	-12.651	0.006**	-4.68	0.018*
ʊ-F2	1122	0.426	0.712	0.941	0.416
u:-F1	378	0.361	0.753	-1.059	0.367
u:-F2	997	3.167	0.087	5.53	0.012*

Table 3. *One Samples Test result of female.*

	Native female	Cantonese female		Mandarin female	
		t	p	t	p
I-F1	483	1.741	0.157	2.061	0.131
I-F2	2365	-1.219	0.29	-5.823	0.010*
i:-F1	437	-1.12	0.325	-4.448	0.021*
i:-F2	2761	-0.984	0.381	-1.835	0.164
ʊ-F1	519	-0.941	0.4	-0.909	0.43
ʊ-F2	1225	4.012	0.016*	0.15	0.89
u:-F1	459	-1.386	0.238	-3.634	0.036*
u:-F2	1105	3.811	0.019*	-0.165	0.88

* p<0.05, ** p<0.01

As seen from Tables 2 and 3, both Cantonese males and females can produce native-like [I], while Mandarin learners cannot. For [i:], Cantonese females have native-like production, Cantonese males have lower F2 than native speakers, while Mandarin group has lower F1. When it comes to [u:], Cantonese males also stand out in terms of native-likeness in a group level, but Cantonese females have lower F1. Remarkably, Cantonese females can differentiate the [I]/[i:] contrast well in both production and perception.

Mandarin females stand out in production of [ʊ], showing no difference in comparison to native speakers; Cantonese learners cannot produce native-like [ʊ]. Specifically, Cantonese males have lower average F1 when compared to native speakers, while Cantonese females have higher F2. The fact that Mandarin group outperform Cantonese group in the production of [ʊ] can be justified by Contrastive Analysis and Speech Learning Model mentioned in 2.1, which assumes that new phones are much easier to acquire than similar ones are, while the fact that the Cantonese outperform the Mandarin group in the production of [I] does not apply to these models. It could be justified by the L2LP model, which predicts that the SIMILAR scenario is easiest and the NEW scenario the hardest, since the English lax-tense vowel contrasts are the NEW scenario in Mandarin while the SIMILAR scenario in Cantonese.

However, why Cantonese group fail in producing a native [ʊ] but succeed in a native [I] is quite puzzling, since the two lax vowels are both recognized as allophones in Cantonese that precede nasals and plosives. One explanation is that the experiment did not use British English to measure their production but used American English as a yardstick. Further study could develop on this basis to test whether the same result stands when controlled for accent preference.

Moreover, it can be inferred from the result that transferring to perception is easier than to production. Cantonese have very clear pattern of using spectral cues, showing sound grip of vowel qualities, although they do not necessarily produce those phones in a native way.

5. Conclusion

This study is an acoustic analysis of the perception and production of English lax-tense vowel pairs [I]/[i:] and [ʊ]/[u:] by two groups of Chinese learners, one with Cantonese background and one with Mandarin. The result shows that Cantonese speakers can use spectral cues effectively to differentiate tense vowels from the lax vowels, indicating dialect's positive transfer into L2 language, while Mandarin speakers rely heavily on durational cues. It also shows that Cantonese speakers are not overall better than Mandarin speakers in terms of lax and tense vowel production, indicating that transferring to production might be harder than to perception.

Furthermore, sexual difference has always been dismissed in previous studies, but this study has demonstrated that gender is a meaningful variable. Production-wise, men and women in each dialectal group have shown different patterns in their production. Those variances would have been overlooked had gender not been considered.

The result of this study calls for new paradigm of transfer research. The result has shown that identifying the difference between languages as the condition for transfer is far from enough. Dialectal and gender backgrounds are all at play, and their influence should be further investigated. Additionally, although acoustic experiments usually have a small size, this study suggests that an expanded experiment size is needed because sexual difference also causes variability.

Acknowledgements

The research for this paper was financially supported by Beijing Municipal Education Commission ([2019] 552) and Central University Research Fund (3162020ZYQA01). The quantitive analyses are empowered by Hongchen Xu.

References

[1] Major, Roy C. (2008). 3. Transfer in Second Language Phonology: A Review. In Studies in Bilingualism, edited by Jette G. Hansen Edwards & Mary L. Zampini, 36: 63-94. Amsterdam: John Benjamins Publishing Company. doi: 10.1075/sibil.36.05maj.

[2] Oller Jr, John W., & Seid M. Ziahosseiny. (1970). The contrastive analysis hypothesis and spelling errors. Language learning, 20 (2), 183-189. doi: 10.1111/j.1467-1770.1970.tb00475.x.

[3] Flege, James Emil. (1987). The Production of ‘New’ and ‘Similar’ Phones in a Foreign Language: Evidence for the Effect of Equivalence Classification. Journal of Phonetics, 15 (1), 47-65. doi: 10.1016/S0095-4470(19)30537-6.

[4] Mayr, Robert, & Paola Escudero. (2010). Explaining Individual Variation in L2 Perception: Rounded Vowels in English Learners of German. Bilingualism: Language and Cognition, 13 (3), 279-297. doi: 10.1017/S1366728909990022.

[5] Tyler, Michael D., Catherine T. Best, Alice Faber, & Andrea G. Levitt. (2014). Perceptual Assimilation and Discrimination of Non-Native Vowel Contrasts. Phonetica, 71 (1), 4-21. doi: 10.1159/000356237.

[6] van Leussen, Jan-Willem, & Paola Escudero. (2015). Learning to Perceive and Recognize a Second Language: The L2LP Model Revised. Frontiers in Psychology 6. doi: 10.3389/fpsyg.2015.01000.

[7] Escudero, Paola, & Daniel Williams. (2012). Native Dialect Influences Second-Language Vowel Perception: Peruvian versus Iberian Spanish Learners of Dutch. The Journal of the Acoustical Society of America, 131 (5), EL406-12. doi: 10.1121/1.3701708.

[8] O’brien, Mary Grantham, & Laura Catharine Smith. (2010). Role of First Language Dialect in the Production of Second Language German Vowels. IRAL - International Review of Applied Linguistics in Language Teaching, 48 (4). doi: 10.1515/iral.2010.013.

[9] Falahuddin, M. A., Saleh, M., & Fitriati, S. W. (2019). The Influence of Mid-East Sundanese Dialect (L1) in The Pronounciation of English Among English Department Students at Universitas Majalengka. English Education Journal, 9 (2), 157-163. doi: 10.15294/EEJ.V9I2.28396.

[10] Chládková, Kateřina, & Václav Jonáš Podlipský. (2011). Native dialect matters: Perceptual assimilation of Dutch vowels by Czech listeners. The Journal of the Acoustical Society of America, 130, EL186-192. doi: 10.1121/1.3629135.

[11] Miller, Joanne L., & Francois Grosjean. (1997). Dialect Effects in Vowel Perception: The Role of Temporal Information in French. Language and Speech, 40 (3), 277-88. doi: 10.1177/002383099704000304.

[12] Chen, Hsueh Chu. (2015). Acoustic Analyses and Intelligibility Assessments of Timing Patterns Among Chinese English Learners with Different Dialect Backgrounds. Journal of Psycholinguistic Research, 44 (6), 749-73. doi: 10.1007/s10936-014-9315-5.

[13] Chan, Alice Y. W. (2011). The Perception of English Speech Sounds by Cantonese ESL Learners in Hong Kong. TESOL Quarterly, 45 (4), 718-48. doi: 10.5054/tq.2011.268056.

[14] Zee, Eric. (1991). Chinese (Hong Kong Cantonese). Journal of the International Phonetic Association, 21 (1), 46-48. doi: 10.1017/S0025100300006058.

[15] Jongman, Allard. (2005). Acoustic Phonetics. In Encyclopedia of Linguistics, edited by Philipp Strazny, 1-2. New York: Routledge.

[16] Boersma, Paul & David Weenink. Praat: Doing Phonetics by Computer. https://www.fon.hum.uva.nl/praat/.

[17] Hillenbrand, James, Laura A. Getty, Michael J. Clark, & Kimberlee Wheeler. (1995). Acoustic Characteristics of American English Vowels. The Journal of the Acoustical Society of America, 97 (5), 3099-3111. doi: 10.1121/1.411872.