DIALECT CONTACT ACROSS THREE GENERATIONS: A SOCIOPHONETIC ANALYSIS OF VARIATION IN [p^h, t^h, k^h, h] IN A CONTACT VARIETY IN HOHHOT, CHINA

Xuan Wang

Hong Kong Polytechnic University
xuan.xw.wang@polyu.edu.hk

ABSTRACT

This paper presents a sociophonetic analysis of a linguistic feature in the formation of a contact variety in Hohhot, a Chinese immigrant city. The new mixed variety was induced by contact between the local Jìn-dialect-speaking community and the Mandarin-speaking migrant community. The linguistic feature examined was the insertion of a period of frication (often [x]) after [p^h, t^h, k^h, h]. Linguistic production of 67 speakers (1957 tokens) from three generations of the local and migrant communities were collected and analyzed by perceptual coding (verified by acoustic measurements). Results of mixed effects regression show that this variable is constrained by both social and linguistic factors. Typical Jìn speakers in the older generation is significantly more likely to use the [x] variants than non-Jìn-speaking migrants, while in the middle and younger generation, when the contact variety has formed, the two communities showed convergence in their probability of using this linguistic feature.

Keywords: dialect contact, sociophonetics, plosives and fricatives, Mandarin varieties.

1. INTRODUCTION

The outcome of mixed varieties induced by language/dialect contact has been studied at length in contact linguistics and sociolinguistics. e.g. [11, 13, 12] The phonetic outcome resulting from the mixing or convergence of different linguistic systems has been found to be very complex in different sociocultural scenarios. This paper examines a situation where a linguistic feature is found in both source dialects/languages in contact, but differs in frequency of use. The locality examined is a Chinese immigrant city, Hohhot. It is the provincial capital of Inner Mongolia, located in north central China. The urban area of Hohhot is considered to be composed of the New Town and the Old Town. Residents of the Old Town are mainly the locally-born community, speaking a sub-variety of Jìn dialect, while residents of the New Town are mostly migrants from all parts of China, who arrived at Hohhot in the 1950s and 1960s because of government policy. Linguistically, they speak dialects from their original hometowns but use Pǔtōnghuà (standard Mandarin) as a lingua franca in Hohhot. This dialect contact situation between the local and migrant communities led to the formation of a new urban variety, Hū Pǔ 呼普 [8], or Hohhot Pǔtōnghuà, which has absorbed linguistic elements from both Jìn dialect and Pǔtōnghuà. This paper, using a variationist sociolinguistic approach, examines a linguistic variable in three generations of the two communities in Hohhot.

The linguistic feature examined is the fricative variable. That is, for the plosives [p^h, t^h, k^h] and the glottal fricative [h], speakers in Hohhot tend to produce them with a velar fricative [x] following. This is, essentially, on the plosives, a period of frication. [x] could also be realized as [χ] or [ç] because of assimilation to different places of articulation. This linguistic variable has not been widely studied in previous literature. Some early linguists such as Karlgren [7] and Chao [3] reported this type of plosives in major Jìn-speaking areas, while more recently, Hou [5] listed it as one of the key features of Jìn dialects. However, this feature is not unique in Jìn. Evidence from Karlgren [7] and Duanmu [4] has shown that the feature of combining [p^h, t^h, k^h,h] with the fricative [x] could also be heard in Pǔtōnghuà and in non-Jìn areas like Beijing, but maybe not as often as, or as noticeably as, in Jìn-speaking areas. Discussion of the linguistic and social distribution of this feature in previous literature was also limited. Chao [3] claimed that in this type of plosive, aspiration was the major function, and the insertion of fricatives [x] was only random. Duanmu [4] reported that the [x] variants were often realised before a back vowel, indicating a co-articulation effect. However, the descriptions have traditionally been only based on these linguists' personal observations.

This leads to the following questions of interests: 1) What is the linguistic constraints operating on this fricative variable? Is it simply as what [3] has described as random variation? Or is it a co-articulation effect proposed by [4] which only relates to the place of articulation of the following vowel? 2) Will this variable be conditioned by social factors? Will the three generations of the migrant and the local communities use this feature differently? And since this feature may differ in frequency of use in Jìn dialect and Pǔtōnghuà, what is the linguistic outcome of dialect contact in Hū Pǔ?

2. METHOD

2.1 Participants

Data were collected from 67 speakers, with 35 from the New Town and 32 from the Old Town. Participants were evenly distributed between male and female, and across three age groups: the older (born in 1930-1941), mid-aged (born in 1949-1962) and the younger generations (born in 1985-1995). All participants were interviewed and recorded either at their own homes or a quiet environment using a Sony ICD-TX50 digital voice recorder.

2.2. Data collection

Twenty disyllabic words with [p^h, t^h, k^h, h] initial sounds in the first syllables were selected as target words[1]. As some of these words only existed in colloquial, unwritten forms, speakers' production of the words were collected through a picture description task and a scheduled interview to elicit more natural linguistic data. Altogether 1957 tokens were collected, and all tokens were extracted and hand coded from the recordings using Praat [2].

2.3 Coding of linguistic data

For the plosives [p^h, t^h, k^h] tokens, the duration from the burst of the plosive to the end of the fricative was labeled as an interval. And for [h] tokens, the duration of the entire fricative was labeled. An impressionistic method was first applied on the coding of these tokens. Each token was labeled by auditory analysis as whether or not a period of frication, usually [x], was heard in the production, which formed a binary linguistic variable. The coding process, however, was sometimes difficult because many of the tokens were not straight-forward in terms of whether I could hear a [x] or not. Therefore, some acoustic measurements were then made in Praat in order to justify the perceptual coding.

The following six acoustic measures were taken for each token in Praat: 1) Center of gravity (hereafter, COG), which is the average frequency of the entire spectrum. 2) Standard deviation (hereafter, SD), which indicates how far each frequency value is from COG. By inspecting visually from the spectrograms, the spectral energy is often concentrated at higher frequencies for [x] variants, so they are expected to have higher COG and SD values. 3) Skewness. 4) Kurtosis. The measure of skewness and kurtosis are both related to the shape of the acoustic spectra of fricatives. In the spectral slices for [h] tokens, the amplitudes on the higher frequencies around 10kHz often had an obvious drop, which makes the shape more skewed and more different from a Gaussian shape. Thus, [p^h, t^h, k^h, h] tokens were predicted to have higher skewness and kurtosis values than the corresponding [x] variants. 5) Normalized intensity, which measures the amount of energy in the fricative. The fricative [h] produced at the glottis is very difficult to produce with a high amplitude, so the intensity measures are likely to be higher for [x] tokens. 6) Duration. [x] tokens are expected to be longer, because there might be more acoustic energy involved when the turbulent airflow was escaping through a narrower channel [6].

All measures were made over the middle 60% of the labeled interval of each token. To corroborate the perceptual coding, a mixed effects regression model was run to check whether these acoustic measures were significantly correlated with the auditory coding decisions. However, before doing this, it is worth noticing that some of the acoustic measurement values are likely to correlate with each other. In order to avoid multicollinearity and reduce the dimension of the six measures, a principal component analysis (PCA) was run on these measures using the *principal*() function in the *psych* package in R [9, 10]. The result revealed four main factors[2] (see Table 1), which explained 96% of the total variance. The first factor PC1 is related to skewness and kurtosis. COG and SD values have high loadings on PC4, while PC2 and PC3 are related to normalized intensity and duration respectively. The proportion of variance explained by each factor is shown in the column *Pro.V*. Four PC scores were then calculated for each token to represent their acoustic values in these four aspects.

Table 1: Four acoustic measurement factors revealed by PCA.

Factors	Representing acoustic features	Pro. V
PC1	skewness and kurtosis	31%
PC4	COG and SD	31%
PC2	normalized intensity	17%
PC3	duration	17%

Then, a binomial mixed-effects model was fitted using the *glmer()* function in the *lme4* library [1] in R, with the perceptual coding as the dependent variable. The four PC scores revealed by PCA were tested as independent variables, as well as two linguistic variables PHO (the phonemes [p^h, t^h, k^h or h]) and FVOWEL (the following vowel [a, i, u, or ə]). The outcome of the best-fit model is shown in Table 2, in which the four PC scores are all significantly correlated with the auditory coding. The result indicates that the auditory coding was justified and could be used for further analysis.

Table 2: Outcome of the best model for correlation between auditory coding and acoustic measures.

	Estimate	p-value	
(Intercept)	-3.21862	6.52e-08	***
FVOWEL_i	-8.35718	0.00131	***
FVOWEL_ə	1.24373	0.00184	***
FVOWEL_u	1.23624	0.00280	***
PHO_h	1.93128	2.57e-05	***
PHO_p	0.61943	0.24253	
PHO_t	1.23901	0.01466	*
PC1	-0.27315	0.00302	**
PC2	1.34279	< 2e-16	***
PC3	0.65606	4.61e-13	***
PC4	0.60071	7.99e-09	***

3. STATISTICAL ANALYSIS

The perceptual coding data were used to investigate the social and linguistic constraints operating on the [x] variable. The 1957 tokens were hand fitted into binomial mixed effects models in R. The dependent variable was the [p^h, t^h, k^h, h] phonemes being realized as [p^x, t^x, k^x, x], that is, with a period of frication. The independent variables included both social and linguistic factors. For linguistic factors, the following vowel (FVOWEL) was included to test the co-articulation effects proposed in [4], and the phoneme itself was also included as a categorical factor (PHO) to see whether [p^h], [t^h], [k^h] and [h] tokens would behave differently. The twenty disyllabic words tested in this study also showed difference in their stress patterns, varying between a weak-strong pattern (with checked tone in the first syllable and full tone in the second syllable) and a strong-weak pattern (with full tone in the first syllable and neutral tone in the second syllable), so stress pattern of each token (STRESS) was also tested as a potential predictor. As for social factors, the effects of AGE (age group), SEX, TOWN (new town or old town), EDU (education level) were tested.

The model was hand-fitted using forward stepwise selection, with pairwise model selection guided by the Akaike Information Criterion. Speaker and word were included as random intercepts. The best model had the fixed effects of FVOWEL, a two-way interaction between PHO and STRESS, and a three-way interaction between TOWN, SEX and AGE. For random effects, the model had speaker as a random intercept and a by-speaker random slope for PHO and FVOWEL.

Table 3: Output of the best model to explore the linguistic and social constraints operating on the fricative variation.

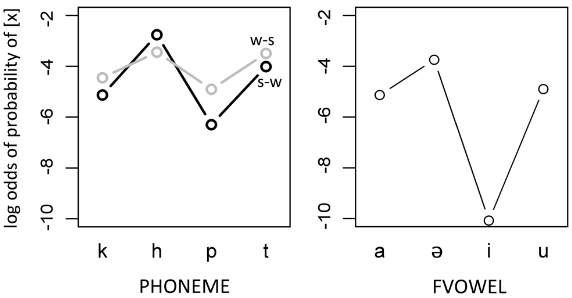
	Estimate	p-value	
(Intercept)	-5.1316	6.73E-05	***
PHO_h	2.3684	1.25E-05	***
PHO_p	-1.1641	0.339269	
PHO_t	1.1186	0.063167	.
STRESS_w-s	0.6794	0.179753	
FVOWEL_ə	1.3858	0.031962	*
FVOWEL_i	-4.9442	0.032886	*
FVOWEL_u	0.2357	0.711095	
SEX_male	1.7479	0.0854	.
TOWN_old	4.687	2.30E-06	***
AGE_mid	2.9927	0.00086	***
AGE_yng	2.2657	0.016019	*
PHO_h:STR_w-s	-1.3622	0.012848	*
PHO_p:STR_w-s	0.7113	0.561737	
PHO_t:STR_w-s	-0.1631	0.795384	
SEX_m:TOWN_old:AGE_mid	4.6378	0.00195	**
SEX_m:TOWN_old:AGE_yng	4.4894	0.007692	**

The output of the best-fit model is presented in Table 3. Plots of the effects of the main factors are shown in Figures 1 and 2. The Y-axes in these plots all represent the log odds of the probability of [p^h, t^h, k^h, h] being realized as [p^x, t^x, k^x, x]. The larger the number, the more likely that [x] tokens are realized.

The left plot in Figure 1 illustrates the effects of the interaction between PHONEME and STRESS. The phonemes [p^h, t^h, k^h] did not show a significant difference in the weak-strong and strong-weak stress patterns in terms of whether a [x] sound is involved. However, for the [h] phoneme, it is more likely to be pronounced as [x] in the strong syllables ($p<0.05$), that is, in the case of a strong-weak stress pattern. This is easily explained from the perspective of articulation. It is more likely that the [h] phoneme is realized as [x], which is more intense and "noisy" acoustically, in a stressed and longer syllable.

The effect of FVOWEL is shown in the right plot in Figure 1. As previously mentioned, [4] proposed that the realization of [x] could be related to the following back vowels, like [u]. From the results, [u] significantly favors the [x] variants compared to the front vowel [i] ($p<0.05$). Also, the vowel [a] is behaving similarly with [u], and [ə] even favors [x] significantly more than [u] ($p<0.001$). This mostly support [4]'s arguments that [x] variants are favored when followed by back vowels.

Figure 1: Plots of the effects of FVOWEL and the interaction between PHONEME and STRESS for the fricative variable.



However, the fricative variable is not simply a co-articulation effect, because the variation was also found to be constrained by social factors in Hohhot. In addition, no interaction was found between the linguistic and social factors, which indicated that the p, t, k, h tokens were behaving similarly in their social distribution. The plots for the three-way interaction between TOWN, SEX and AGE are shown in Figure 2. The X-axes are the three age groups or generations so that we could see the change in this variable over time clearly. The black lines refer to the New Town speakers, and the gray lines represent the Old Town speakers. The two plots show the results from female and male speakers respectively.

Figure 2: Plots of the 3-way interaction between TOWN, SEX and AGE for the fricative variable.

For the female speakers (the left plot in Figure 2), the pattern of the fricative variable shows clearly a converging trend. The older generation of Old Town residents, who are mostly Jìn speakers, are significantly more likely to produce the [x] variants, whereas the older speakers from the migrant community are significantly less likely to produce this feature ($p<0.001$). This result also provides direct evidence for the frequent use of [p^x, t^x, k^x, x] as a Jìn feature, which supports the early linguists' observation [7, 3]. When it comes to the second and third generations, speakers from the Old Town and the New Town seem to have converged to each other and show a similar likelihood of using the [x] variants.

The male speakers, as shown in the right plot in Figure 2, presents a different pattern. The older group speakers from New Town and Old Town do not show a significant difference in their fricative variation. This doesn't necessarily indicate that this feature is chiefly associated with female speakers, but may relate to the fact that I failed to find any older male participants living in the Old Town due to the limit of my social networks. Instead, I interviewed several older Jìn speakers who grew up in the Old Town but later moved to New Town. Many of them were professors from a local university, whose pronunciation may have changed and converged to New Town speakers due to their social networks and higher educational level. The convergence pattern continues in the mid-aged male speakers, in which New Town and Old Town speakers show a similar probability of using the [x] variants. However, the younger speakers present some changing trends. For the New Town speakers, the younger group is significantly less likely to use the Jìn form compared to the mid-aged group ($p<0.05$). It seems that the New Town younger group speakers are leading the change to reduce the use of the [x] variants, and the Old Town younger group speakers are conforming to this trend.

4. CONCLUSION

The fricative variation results show that this variable is not a haphazard or random insertion of [x] as suggested by [3]. The feature is found to be strongly favored before back vowels, which supports [4]'s claim, but it is not simply a co-articulation effect since social factors also come into play. The typical Jìn speakers in the Old Town are significantly more likely to use the [x] variants than the first-generation migrants in the New Town, which further proves the claim that the frequent use of [p^x, t^x, k^x, x] is a Jìn feature. The language change pattern indicates that in the mid-aged and younger generation, when the mixed variety has formed, speakers from New Town and Old Town have converged and reached a similar probability of using the [x] variants, which is also the phonetic outcome of this variable in the mixed variety Hū Pǔ. The method of using auditory coding corroborated by acoustic measures of this variable was proved to be successful in this study, and could be used in future studies to explore this feature in a broader linguistic context.

5. REFERENCES

[1] Bates, D., Maechler, M., & Bolker, B. (2011). lme4: linear mixed-effects models using S4 classes. R

package version 1.1-11. Retrieved March 20, 2016 from http://cran.rproject.org.
[2] Boersma, P., Weenink, D. 2017. Praat: doing phonetics by computer [Computer program] (Version 6.0.26). Retrieved from http://www.praat.org/.
[3] Chao, Y. R. 1935. The types of plosives in Chinese dialects. *Bulletin of Institute of Historical and Philosophy, 5*(4), 515–520.
[4] Duanmu, S. 2007. *The phonology of standard Chinese*. Oxford: Oxford University Press.
[5] Hou, J. 1999. *Studies of modern Jìn dialect*. Beijing: Commercial Press.
[6] Johnson, K. 2002. *Acoustic and Auditory Phonetics*. Wiley.
[7] Karlgren, B. 1915-28. *Études sur la phonologie chinoise [Study of Chinese phonology]*. Leyde et Stockholm.
[8] Puthuval, S., Wang, X. 2017. The Language of Hohhot. In Sybesma, R., Behr, W., Handel, Z. Huang, J. (Eds.), *Encyclopedia of Chinese Langauge and Lingusitics* (pp. 379-386). Leiden: Brill Academic Publishers.
[9] Revelle, W. 2015. Psych: procedures for personality and psychological research. Northwestern University, Evanston, Illinois, USA. R package version: 1.5.8. Retrieved June 15, 2015 from http://CRAN.R-project.org/package=psych.
[10] R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, R version 0.98.1103. Retrieved April 26, 2015 from http://www.R-project.org/.
[11] Thomason, S. G., Kaufman, T. 1988. *Language contact, creolization, and genetic linguistics.* Berkeley: University of California Press.
[12] Trudgill, P. 1986. *Dialects in contact.* Oxford: Blackwell.
[13] Winford, D. 2003. *An introduction to contact linguistics*. Oxford: Blackwell.

[1] This paper is a part of a larger sociolinguistic project looking at a set of disyllabic words in Hohhot, so the [p^h, t^h, k^h, h] tokens were only examined in these selected words, but not in a broader linguistic environment.

[2] The decision was made by considering results of several statistical criteria including scree plot, parallel analysis, Minimum Average Partial criterion (MAP), and Very Simple Structure Criterion (VSS).