Mandarinization: A socio-phonetic account of Nanjing Dialect's new retroflex vowel

Evan Hugh Coles-Harris
University of Colorado Boulder

1 Introduction

This paper is an acoustic investigation of one aspect of dialectal convergence as it is taking place in the Chinese city of Nanjing. The study compares the retroflex vowel as it is produced by younger and older speakers of the local speech variety, Nanjing Dialect (ND), and the national standard variety, Modern Standard Mandarin (MSM), in order to determine whether there has been a change in apparent-time (Labov, 1965). I discuss first the motivation for the study, namely a prevalent narrative in Nanjing that states that the local variety is "mandarinizing." Second, I will discuss and compare the acoustic facts of ND's retroflex vowel as it is used by younger and older speakers. Third, I will argue that there has been a shift among ND speakers such that younger speakers' retroflex vowels are more MSM-like than older speakers'. Finally, I will discuss the implications of this case study for our understanding of contact-induced varietal convergence.

Key terms: sociophonetics, Chinese linguistics, convergence, retroflex vowel

1.1 *Motivation*

This research was inspired in part by anecdotal reports of dialectal convergence that I heard from local friends while I was living and working in Nanjing, the capital of China's Jiangsu Province. When asked to speak the local dialect, many of my friends would refuse, citing the feeling that their ND was "inauthentic," despite early and long-term exposure to the variety both in and outside the home. They claimed that their ND had undergone *pǔhuà*, or "mandarinization," a Chinese term which has come to be fairly widely used among the general population to denote the structural convergence of small local dialects with MSM, the country's official language.

Texas Linguistics Forum 59: 30-38
Proceedings of the 24th Annual Symposium about Language and Society-Austin
April 15-16, 2016

One set of anecdotal reports in particular motivated the decision to consider the retroflex vowel in detail. In 2014, I retrieved the following comments from http://phonemica.net, an interactive and crowd-sourced corpus dedicated to the documentation of all of China's many speech varieties. The particular recording I found these comments on was one of two ND recordings on the site at the time.

> 2013-07-13 15:17 [commenter 1] This Nanjing Dialect feels mandarinized.
>
> 2013-07-25 05:03 [commenter 2] I've found that it's mostly university students that record [for this website], university students are relatively cultured, and they speak Mandarin relatively well, so it's inevitable that there'd be some mandarinization.

And since a more recent site update:

> 2015-05-25 02:22 [commenter 3] There's already a lot of mandarinization [in this recording], retroflexes are differentiated, n and l are also differentiated, it's not very authentic. (Phonemica, 2015)

The first and second comments establish that the ND in the recording is mandarinized, and that this mandarinization is viewed as the likely result of extensive contact with MSM on the part of the speaker. The last comment and others like it (e.g. Wang, 2015; Ge, 2016) make clear that maintaining a distinction between a set of dental syllables (i.e. *zi* [tsɿ], *ci* [tsʰɿ], *si* [sɿ]) and a similar set of retroflex syllables (i.e. *zhi* [tʂɻ], *chi* [tʂʰɻ], *shi* [ʂɻ]) is a salient marker of mandarinization in ND. MSM maintains this distinction in near-minimal pairs like *cí* [tsʰǐ] "word" / *chí* [tʂʰɻ̌] "ruler," while authentic ND is said to have only the dental set of syllables, producing [tsʰǐ] for both. If mandarinization is indeed taking place among the youth of the city, then they should produce [ɻ] while older speakers produce [ɿ] in contexts that require [ɻ] in MSM.

Please note that I'm using the dental and retroflex apical vowels [ɿ] and [ɻ] to transcribe what have also sometimes been transcribed as the syllabic fricatives [z̩] and [ʐ̩] (as in Duanmu, 2007), or as the approximants [ɹ̩] and [ɻ̩] (as in Lee-Kim, 2014). This might be slightly controversial, as there is disagreement about how best to represent these sounds (see Lee-Kim, 2014, for an overview of that debate), but which of these representations is most appropriate is only tangential to the argument of this paper, and so I will be using one set of representations, but the reader may choose to imagine another set in their place. I have been using and will continue to use the symbols [ɿ] and [ɻ] for the time being. Also note that when I use Chinese words, I have provided both the Chinese pinyin transcription in italics and an IPA transcription in brackets where relevant.

1.2 *Prior Work*

Scholarly support for the notion that mandarinization is happening at all in ND comes from Bao (1980/2010) and Chen and Wiltshire (2011), who described phonological changes under the influence of MSM in the variety's consonant and tone systems, respectively. Acoustic studies of ND vowel qualities are, as far as I know, non-existent, and so there is no documentation of how the vowel system may be being affected by processes of mandarinization. There are fortunately several studies of MSM that have

done acoustic work on the MSM retroflex vowel that help us establish some context for this study.

Critically for our purposes, Howie (1976), Zee and Lee (2001), and Lee-Kim (2014) all confirmed that the MSM retroflex vowel [ɻ] is distinguished from the dental vowel [ɿ] not only by lowered F3--which one would expect from increased retroflexion (Hamann, 2003)--but also by an increased F2. Knowing this and the fact that mandarinization in ND consists in part of [ɿ] shifting to [ɻ] allows us to search for mandarinization in the presence of particular acoustic patterns among ND speaking youth vis-a-vis ND speaking elders, namely lowered F3 and raised F2 in [ɻ].

1.3 *Hypotheses*

The aim of this research is to determine whether there is any evidence of mandarinization in the retroflex vowels of younger ND speakers. To that end, the hypotheses of this paper are formulated as follows:

1. Younger ND speakers' retroflex vowels will show significantly raised F2s relative to older ND speakers'.
2. Younger ND speakers' retroflex vowels will show significantly lowered F3s relative to older ND speakers'.
3. The F2s of younger ND speakers' retroflex vowels will be significantly closer to Mandarin norms than will older ND speakers'.
4. The F3s of younger ND speakers' retroflex vowels will be significantly closer to Mandarin norms than will older ND speakers'.

Testing hypotheses 1 and 2 will indicate whether there has been a change in apparent-time in ND, and suggest whether any change is in the direction of MSM norms. Testing hypotheses 3 and 4 will confirm whether any change is proceeding towards an MSM target vowel, as suggested by local narratives about mandarinization.

2 Methods

The project uses a subset of phonetic data I collected for an as-yet unpublished study meant to test the hypothesis that mandarinization would be apparent in ND's vowel system. The subset consists of the retroflex vowels collected for that larger project.

2.1 *Word List Preparation*

The larger project of which this was a part employed a wordlist containing forty-five Chinese ideographs consisting of five homophonous monosyllables for each of nine target ND monophthongs. The relevant subset of words for this study consisted of five homophonous ideographs that, in MSM, are pronounced *chí* [tʂʰɻ̌]. The ideographs chosen mean "hold," "ruler," "run fast," "pool," and "spoon." These particular ideographs were selected for two reasons. First, they should participate in the ND dental/retroflex alternation, rendering any difference between populations that use one variant or the other visible, allowing me to test my hypotheses. Second, they are common characters that should be known to both ND and MSM speakers with moderate or greater levels of education.

2.2 *Subject Selection*

Potential subjects were pre-selected from personal acquaintances or via snowball sampling. Those who self-identified as speaking either MSM or ND and expressed some willingness to participate were asked to do so. A short survey was administered to collect demographic information and determine whether or not potential subjects were suitable to participate. Because most subjects had at least some proficiency in both MSM and ND, they were asked which felt more natural to speak, and were then asked to record using whichever variety they identified. Several questions were included as exclusion criteria, but in the end no subjects were excluded from recording.

In total, sixteen speakers were recorded for this study, twelve ND speakers and four MSM speakers. Each dialect group was evenly split between "older" (age 52-68) and "younger" (age 20-31) subjects. Prior research (Jiang, 2009) has found these to be appropriate ages to study cross-sectionally to test for generational differences in ND. Among ND speakers, the average age of younger subjects was 25, and the average age of older subjects was 58. Among MSM speakers, the average age of younger subjects was 23, and the average age of older subjects was 59.

2.3 *Recording*

Stimuli consisted of the target ideographs couched twice in a carrier sentence. These sentences were presented to each subject one at a time in random order via PsyScope X B77 (Cohen, MacWhinney, Flatt, & Provost, 1993). Subjects were asked to read the stimuli out loud, and were allowed to do so at their own pace.

Subjects were recorded using a Marantz PMD660 hand-held compact flash recorder and an external Audio-Technica AT897 compact shotgun condenser microphone connected by a three-pin XLR cable. Recordings were monaural .WAV files sampled at 48khz so as to provide high quality audio for acoustic analysis.

2.4 *Measurements*

Vowel tokens were segmented and their formants were automatically measured in Praat (Boersma & Weenink, 2014). Raw formant values were then imported into R (R Core Team, 2013) and examined. Any suspicious values were double-checked by visual inspection of the token's spectrogram, and corrected if they were found to be faulty. In total, 156 usable tokens of the (potentially) retroflex vowel were analyzed.

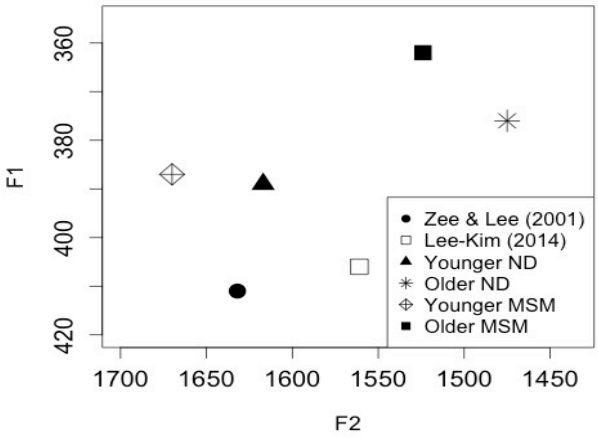
2.5 *Analysis*

Before conducting acoustic analysis, I listened to my data and made impressionistic judgements as to how many tokens were retroflex versus dental. I am not a native speaker of either of these varieties, but at the time that the study was conducted, I had been formally studying MSM for 14 years and had 4 years' worth of exposure to ND, and so neither am I a naïve listener. Of tokens produced by older ND speakers, 38/58 (65.5%) were unambiguously retroflexed, while the remaining were either dental or ambiguous. Of tokens produced by younger ND speakers, 57/58 (98.3%) were retroflex, and the

remaining token was dental. This would indeed seem to indicate an inter-generational change towards MSM norms. Interestingly, only 15/20 (75%) of the tokens produced by older MSM speakers sounded unambiguously retroflex, while the other 25% seemed to be dental or possibly intermediate productions. Younger MSM speakers produced what I heard to be retroflex vowels for 20/20 (100%) tokens, suggesting that a similar shift is taking place among the MSM speakers in Nanjing. Such a shift may be the result of the fact that MSM has only been the national standard since 1956 (Ramsey, 1987), and may not have been the first variety acquired by my older MSM consultants.

To test the hypothesis that younger ND speakers' F2 would be not just higher than older ND speakers' but also significantly closer to Mandarin norms, it was necessary to establish exactly what those norms are. I had originally planned to calculate distances from a norm consisting of averaged values of both the older and younger MSM speakers recorded for this study, but I abandoned the idea when I determined during acoustic analysis that there are in fact significant differences in all three formants between the younger and older MSM speaking consultants (F1: $t(37.233) = 2.433$ p = 0.0199, CI_{95} 4.183435, 45.808350; F2: $t(32.86) = 3.8831$ p = 0.0004707, CI_{95} 69.38528, 222.16724; F3: $t(37.954) = -2.8705$ p = 0.006666, CI_{95} -341.62828, -59.04603). This fact makes it problematic to treat these two groups as a homogeneous set of "MSM speakers" against which to measure any vowel shift in ND. Furthermore, there were no significant differences within age-groups across dialects. Figure 1 below illustrates this problem. Notice that the demographic groups that consist of younger speakers (including the typical values reported in the literature) cluster higher on the F1 and F2 axes than the groups that consist of older speakers.

Figure 1, Typical formant values for [ɻ], non-normalized



In place of my MSM speakers, I searched the literature for typical formant values of [ɻ] in MSM. I found three sets, which are listed in Table 2 below. Zee and Lee (2001) and Lee-Kim (2014) both consulted younger (college- or graduate school-aged) male speakers. Howie (1976) consulted one male speaker of unstated age. All authors' consultants were from northern China, where the speech varieties upon which the standard is based are spoken and have been spoken since well before 1956. One would hope that this means both older and younger speakers maintain a "standard" speaking style, but this is unfortunately impossible to determine at this time. To solve the problem of establishing a

norm, I opted to reference Lee-Kim's values because they are the most recent. This approach is still problematic, in that it establishes a norm based entirely upon one age group, but it is still preferable to using an average of my own data at this point. It would seem that much more work needs to be done on the issue of what serves as "standard" MSM in different parts of China, and among different groups of speakers, but that work is beyond the scope of this paper.

Table 2, Typical formant values for [ɻ] as reported in the literature

Formant (Hz)	Howie (1976)	Zee & Lee (2001)	Lee-Kim (2014)
F3	3680	1962	2319
F2	2040	1632	1561
F1	400	411	406

3 Results

In hypotheses 1 and 2, I predicted that younger ND speakers' retroflex vowels would show significantly raised F2s and significantly lowered F3s relative to older ND speakers'. Table 3 below lists mean formant values for each of the demographic groups examined in this study. Note that younger F2s are indeed higher and F3s lower than older F2s and F3s. MSM values are listed solely for the curious reader's benefit; I will be comparing my ND values to Lee-Kim (2014) below.

Table 3, Mean formant values of /ɻ/ by demographic group

	Younger ND	Older ND	Younger MSM	Older MSM
n	58	58	20	20
F3	2615 Hz (SD: 262.3)	2734 Hz (SD: 299.2)	2556 Hz (SD: 216.8)	2756 Hz (SD: 224.5)
F2	1617 Hz (SD: 101.6)	1475 Hz (SD: 179.4)	1670 Hz (SD: 92.3)	1524 Hz (SD: 140.2)
F1	389 Hz (SD: 32.4)	376 Hz (SD: 35.7)	387 Hz (SD: 34.7)	362 Hz (SD: 30.1)

A linear mixed-effects model predicting ND speakers' F2 values from age group with random effects for speaker and word confirms that for F2 we do see a significant increase from elders to youths (estimate: 144.6 Hz, *t*-value=3.51), supporting hypothesis 1. The age group model was a better fit than and significantly different from a null model when compared by anova ($\chi^2(1)=8.5423$, $p=0.00347$), allowing us to reject the null hypothesis that F2 would show no difference between the two age groups.

A similar model meant to test hypothesis 2 by predicting F3 values fails to explain the data significantly better than a null model ($\chi^2(1)=1.3574$, $p=0.244$), and so we fail to reject the null hypothesis that there is no difference between the F3s of the two age groups. However, it is worth noting that, while there is not a significant difference between an age group-based model and a null model, the mean F3 of younger ND [ɻ] is 119 Hz lower than the mean F3 of older ND [ɻ]. It is possible that there is a nascent change in progress--one that has yet to reach the consistency necessary for statistical significance--and that the retroflex vowel is mandarinizing one formant at a time.

Two more linear mixed-effects models were created to test hypotheses 3 and 4 by comparing the F2 and F3 distances of the two age groups' productions from the typical formant values reported in Lee-Kim (2014). The model predicting distance of F2 from the norm by age group with random effects for speaker and word confirms that younger ND speakers productions are closer to Lee-Kim's norms along that dimension (estimate: -72.09, *t*-value=-2.439). The age group model was a better fit than, and significantly different from, a null model when compared by anova ($\chi^2(1)=4.8432$, $p=0.02775$). Once again, however, for F3 an age group-based model fails to outperform a null model ($\chi^2(1)=1.3252$, $p=0.2497$), thereby failing to support hypothesis 4.

4 Discussion

The results of this study have interesting implications for our understanding of how mandarinization is progressing in the Nanjing context specifically, but also for how sound change more generally progresses when different linguistic varieties come into contact. The differences in rates of retroflexion in my impressionistic judgements suggest that there has been an inter-generational change in ND, the linear models predicting raw formant values tell us that that the change is in F2, and the models predicting distance tell us that that change in F2 is indeed toward an MSM target. All of this taken together indicates that contact-induced phonological change can proceed a) gradually as a product of changes in rates of variation, b) at a fine-grained phonetic level, and c) in a principled manner towards a variety-external phonetic target.

Prior research (Bao, 1980/2010; Chen & Wiltshire, 2011; Jiang, 2009) and the narrative told by ND speakers themselves suggest that in the case of ND and MSM contact-induced language change has taken place gradually over the course of generations. This study corroborates those accounts by showing that one aspect of mandarinization, a shift in use from [ɽ] to [ɻ] in certain contexts, is only partial among older ND speakers, but complete among younger speakers. More tellingly, among the five older speakers that do display the shift by using the [ɻ] vowel, three only display it some of the time, varying between [ɽ] and [ɻ], rather than using [ɻ] consistently. Among the youth, however, the shift has solidified and become a permanent phonological feature of ND, and the variable [ɽ]/[ɻ] ceased to be a variable at all, with virtually 100% of younger productions being [ɻ]. The age-graded nature of the shift strongly suggests a gradual change over time.

The phonetic nature of this gradual change is also interesting, in that the change has taken place along only one of the two expected dimensions. For both raw formant measurements and distances from MSM norms, F2 shows evidence of mandarinization, while F3 does not. It is possible that F2 is perceptually more prominent for ND listeners, and so accommodation to the MSM norm is happening along this dimension before any other, but because MSM [ɽ] and [ɻ] are differentiated by F2 and F3 simultaneously, we

cannot determine the relative salience of these dimensions from examinations of the retroflex vowel alone. It may be possible to use another vowel pair which in MSM is differentiated solely by F3 to determine the acuity with which ND speakers attune to that dimension of distinction--[i] and [y] seem like likely candidates, as they differ most markedly in F3, and at least some ND speakers do not distinguish between them. In any case, ND speakers are attuned to MSM F2, suggesting that when vowels converge across varietal boundaries, they are analyzed by listeners at a fine-grained phonetic level.

Finally, the results of this study make apparent that different speakers of the same speech variety, when exposed to a second speech variety, converge upon particular phonetic targets from that second variety. This process of convergence is detectable even in dozens of repetitions of the [ɻ] vowel by many speakers within a community, and must, then, happen according to principles shared by those speakers. One potential explanation for how these principles come to be shared is that--while we don't have any direct evidence as to which came first--increased use of the retroflex vowel and the narrative about its increased use may recursively and dialectically reinforce each other, focusing speakers on the [ɻ] vowel as they hear it used and hear about it being used by members of their community.

5 Conclusion

There are three essential points to take away from this study. First, I have provided quantitative impressionistic and acoustic evidence for the reality of the mandarinization of ND. This is an important point from the perspectives of linguistics, anthropology, and China studies because mandarinization serves to reduce the diversity of human language generally and threatens China's rich cultural heritage. Second, the acoustic patterns of mandarinization were surprising, in that they were visible in only one of two expected dimensions, hinting at some deeper principles of sound change. Third, case studies such as this one are informative and have great potential to help us understand the linguistic processes that underlie the convergence of speech varieties, even when they are inspired by anecdotal reports.

More research deserves to be done on this and related topics. As mentioned above, there are issues that I plan to pursue relating directly to the ND vowel system, but beyond that there are the issues of what speech varieties "count" as MSM, how varietal convergence plays out in different contexts, and what principles govern that process.

References

Bao, M.-W. (2010). *Liushi nian lai Nanjing fangyin xiang Putonghua kaolong de kaocha* [Examination of the drawing close of Nanjing dialectal pronunciation to Mandarin over the last sixty years]. In Gu, Qian (Ed.), *Bao Mingwei yuyanxue wenji* [The Collected Linguistic Works of Bao Ming-Wei] (pp. 15–22). Nanjing, China: Nanjing University Press. (Original work published 1980)

Boersma, P. & Weenink, D. (2014). Praat: Doing phonetics by computer (Version 6.0.14) [computer software]. Available from http://www.fon.hum.uva.nl/praat/download_mac.html

Chen, S. & Wiltshire, C. (2011). Differences of tone realization between younger and older speakers of Nanjing Dialect. In Z. Jing-Schmidt (Ed.), *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23)* (Vol. 2, pp. 105–122). Eugene: University of Oregon Press.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope X B77: A new graphic interactive environment for designing psychology experiments (Version 1.1.1) [computer software]. Available from http://www.macupdate.com/app/mac/23133/psyscope

Duanmu, S. (2007). *The Phonology of Standard Chinese* (2nd ed.). The Phonology of the World's Languages. Oxford, U.K.: Oxford University Press.

Ge Lun Ai Xi Liang Da Keng (2016, January 4). *Nei baan ning hak hou nan ngong gau* [You damn Nanjingers are so stupid]. Message posted to http://tieba.baidu.com/p/4240212854

Hamann, S. (2003). *The phonetics and phonology of retroflexes* (Doctoral dissertation). Retrieved from http://www.lotpublications.nl/the-phonetics-and-phonology-of-retroflexes-the-phonetics-and-phonology-of-retroflexes.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge, U.K.: Cambridge University Press.

Jiang, W. (2009). *R-suffixation in Nanjing Dialect: A rule-based analysis and its feature geometry explanation* (Master's thesis). The Chinese University of Hong Kong.

Labov, W. (1965). On the mechanism of linguistic change. In C. W. Kreidler (Ed.), *Report of the Sixteenth Annual Round Table Meeting on Linguistics and Language Studies* (pp. 91–114). Monograph Series on Languages and Linguistics, Institute of Languages and Linguistics. Washington, D.C.: Georgetown University Press.

Lee-Kim, S.-I. (2014). Revisiting mandarin 'apical vowels': An articulatory and acoustic study. *Journal of the International Phonetic Association*, 44(3), 261–282.

Phonemica (2015). *Nanjing shi de yi ge gushi* [A story of Nanjing City]. Retrieved from http://phonemica.net/entry?mid=543746312bd553930782ab9b&story=0

R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (Version 3.0.2) [computer software].

Ramsey, S. R. (1987). *The languages of China*. Princeton, NJ: Princeton University Press.

Wang Xia Pie (2015, September 17). *Nanjing Fangyan suan shi wandan le* [Nanjing Dialect is as good as done for]. Message posted to http://tieba.baidu.com/p/4048426496

Zee, E. & Lee, W.-S. (2001). An acoustical analysis of the vowels in Beijing Mandarin. In P. Dalsgaard, B. Lindberg, H. Benner, and Z.-H. Tan (Eds.), *INTERSPEECH* (pp. 643–646). Aalborg, Denmark: International Speech Communication Association.

Evan Hugh Coles-Harris
Department of Linguistics
Hellems 290
295 UCB
Boulder, CO 80309
evan.colesharris@colorado.edu