Phonetic Convergence in Mandarin

By

Haoru Zhang

B.A., University of Virginia, 2016

Thesis

Submitted in partial fulfillment of the requirements for the Degree of Master of Arts in the Department of Cognitive, Linguistic, and Psychological Sciences at Brown University

PROVIDENCE, RHODE ISLAND

MAY 2019

This thesis by Haoru Zhang is accepted in its present form by the Department of Cognitive, Linguistic, and Psychological Sciences as satisfying the thesis requirements for the degree of Master of Arts

Date________________ _________________________________.
Chelsea Sanker, Advisor

Approved by the Graduate Council

Date________________ _________________________________.
Andrew G. Cambell, Dean of the Graduate School

Table of contents

1. Introduction 1

2. Background 3
   - 2.1 Convergence 4
   - 2.2 Measures of phonetic convergence 6
   - 2.3 Mandarin tones 8
   - 2.4 Experimental methods for phonetic convergence 10

3. Procedure 12
   - 3.1 Participants 12
   - 3.2 Stimuli 13
   - 3.3 Task 14
   - 3.4 Data analysis 16
   - 3.5 Hypotheses and predictions 17

4. Results 19
   - 4.1 F0 (pitch) 19
   - 4.2 F1 and F2 24
   - 4.3 VOT 29
   - 4.4 Vowel duration 31

5. Discussion 35

6. Possible factors and future directions 38

7. Conclusion 41

Works Cited 42

Appendix 1 – list of stimuli 47

Appendix 2 – participant questionnaire 48

1 Introduction

Convergence discusses the phenomenon that people voluntarily or involuntarily adapt to others' various behaviors in a communicative setting (Giles & Coupland, 1991). *Phonetic* convergence, by its name, focuses on the phonetic features in linguistic communication, such as fundamental frequency (Babel & Bulatov, 2012; De Looze et al., 2011; Bulatov, 2009; Collins, 1998), formants and duration in vowels (Sonderegger, Bane & Graff, 2017; Kim, 2012; Evans & Iverson, 2007), and voice onset time (hereafter VOT; Yu et al., 2013; Nielsen, 2011; Shockley et al., 2004).

Although a wide range of phonetic *variables* have been identified to be subject to convergence effects, the majority of previous studies focused on a few *languages* when investigating this phenomenon. Most studies have investigated English (e.g, Gregory & Hoyt 1982, Giles, Coupland & Coupland 1991, Goldinger 1998, Pardo 2006, Babel 2012, etc.), while a few studies investigated Dutch (e.g., Adank et al. 2010, Mitterer & Ernestus 2008), Spanish (e.g., Cibelli 2009 on formant and vowel duration convergence in Spanish-English bilinguals, Simonet 2011 on intonational convergence in early bilinguals, Baluka & Koops 2015 on VOT convergence in bilingual code-switching), and Arabic (e.g., Gregory et al. 1993 on suprasegmentals in interviews, Khattab 2013 on phonetic accommodation strategies in bilingual children). To date, the effects of phonetic convergence on a tonal language have been little explored, with a few studies seeking comparisons with previous English-based results from a corpus perspective (e.g., Xia, Levitan & Hirschberg 2014). The current study contributes to existing convergence literature with controlled experimental data from a widely spoken tonal language. Tonal languages provide additional insights especially on the role of fundamental frequency in phonetic convergence, and can help address the question of whether or not being phonologically contrastive may affect the degree of a particular feature exhibiting convergence. The specific language of interest in this study is Mandarin. It is hypothesized that convergence patterns may differ by tonal categories within a certain variable, such as F0 and vowel formants.

Another contribution of the current study is examining the mechanism of convergence, based on how the extent of convergence effects is influenced by differences in test procedures. Specifically, I will compare results from a mere exposure task and a shadowing task, both of which have been popular methods in previous research. The key difference between these two task conditions is the additional self-priming in the shadowing group, as participants repeat immediately after the model talker. Thus, any difference in convergence is revealed from the task comparison may be driven by this difference in priming. Although previous research has not directly compared these two measures, asking this question would benefit future research in 1) equating findings on similar measures across task condition and 2) asking questions about the underlying mechanisms that drive convergence.

The current paper will first address previous findings on phonetic convergence, briefly discussing the role of social processes in the present theories and identifying the major variables subjected to convergence. Next, it will outline the experiment procedures that have been conducted, and discuss the results from these experiments. Phonetic convergence as found in previous studies was partially replicated, i.e., convergent patterns were found in F0, some vowels, and durational measures. Comparison of results from the two paradigms showed no difference except in F2, i.e., both paradigms revealed significant convergence and only in F2 an interaction between group membership and convergence was found. Finally, the paper will address the significance of the results from this experiment as it relates to the nature of a tonal language and the differences in experimental procedures, suggesting that having phonological contrasts can shape convergence at the sub-categorical level, e.g., the overall significance in F0 is shaped by tonal categories.

In short, the results suggest that convergence in Mandarin is significant overall but also shaped by the categories of tones specifically, and the effects of self-priming as evident from the different test procedures was not found significant.

2 Background

The phenomenon of convergence can be broadly defined as people generally exhibiting similar behaviors in a conversation or interaction with one another, and to linguists the interest lies especially in the convergence effects in speech events. Research from many decades indicate that these effects occur across various settings of utterances and linguistic structures (e.g., Giles et al. 1973 on face-to face conversation, Gregory & Hoyt 1982 on phone interviews, Pardo 2006 on repetitions of recorded speech, etc.). The early literatures on convergence maintain that speakers generally change their speech production behaviors in response to variabilities in the input they received to better achieve communicative goals. Giles (1973) divides these behaviors into two categories, either *convergence* or *divergence*; i.e. the receiver role in a conversation either *reduces* or *emphasizes* the similarities with the sender, respectively. Subsequent studies following this bilateral distinction have developed theories and models accounting for the motivations behind convergence or divergence, such as the communication accommodation theory (hereafter CAT) first raised by Giles, Coupland & Coupland (1991), the

Vocal Channel Social Status Model (hereafter VOCSTAT) from Gregory et al. (2001), the episodic theory from Goldinger (1998) and subsequent exemplar theories (e.g., Pierrehumbert 2001, Schweitzer & Walsh 2016), etc. The latter theories acknowledge the social influences on convergence but also began to take automatic convergence and subconscious motivations into account from investigating in non-social settings and on the phonetic level.

2.1 Convergence

The phenomenon of convergence has been given many terms in the literature, such as "accommodation" (e.g., Giles, Coupland & Coupland 1991, Babel 2009), "adaptation" (e.g., Lawson, Scobbie & Stuart-Smith 2011), "entrainment" (e.g., Levitan et al. 2012, Xia et al. 2014), "imitation" (e.g., Decety et al. 2002, Adank, Hagoort & Bekkering 2010), etc. The terms speak to the many possible explanations of the mechanisms behind convergence that has been raised. For example, the CAT theory claims that speakers may *voluntarily* modify their speech behaviors for the purpose of adjusting their social distance with the conversation partner(s) – hence the term "accommodation". The changes are motivated by the speaker's judgements of the others' favorability, mutual intelligibility, perceived amount of similarities amongst each other, alignment of goals for the conversation, etc. The CAT theory thus has a "socio-psychological core" (Giles, Coupland & Coupland 1991), basing its theory on the assumption that speakers are not only constantly reviewing the social relationships with one another but also able to subjectively manipulate their speech behaviors according to the changing social relationships. Followers of the CAT theory have shown that convergence patterns can be influenced by speaker gender, SES status, education level, race, etc. (see Giles & Ogay 2007 and Babel 2012 for a comprehensive review).

When focusing on the linguistic level, on the other hand, the term “phonetic convergence” is used in general consensus to emphasize the acoustic similarities found in speakers’ utterances driven by an automatic mechanism. Edlund et al. (2009) defines the term *convergence* as the meeting of two parameters at a shared point of points (that they “become more similar” over time). Somewhat contrasting with theories of accommodation, episodic or exemplar theorists such as Goldinger (1998) suggest that listeners store “episodes” of phonetic details and use them as “examples” in their memory, with the more recently established exemplars having greater salience among the cloud of exemplars for a particular phoneme or word. These exemplars guide the perception and production of speakers’ utterance, and thus convergence occurs when an influx of input alters the most recent exemplars.

Although theories such as the exemplar model have been successful for capturing phonological convergence and theories of accommodation have been successful at capturing the socio-psychological factors of convergence, it remains unclear why convergence patterns may differ by the measures taken and whether convergence consistently occurs at token-level, phonetic-level, or phonological level and above. Studies supporting the various theories have generally focused on a few variables from a variety of measurements and thus failed to consistently find similar patterns in the same measurements across studies. Sanker (2015) discusses the correlation of amount of convergence across measures and conversational pairs, presenting challenges to the theories accounting for the underlying mechanism of convergence. What remains unclear, then, is whether or not phonologically contrastive measurements differ from non-contrastive ones in their patterns of convergence; for example, whether or not the convergence found in F0 maximum and vowel formants in English are driven by similar mechanisms, for the former is not phonologically contrastive in English but the latter is.

2.2 Measures of phonetic convergence

To date, few studies have investigated phonological contrastiveness as a variable affecting convergence. Mitterer & Ernestus (2008) argued for contrastiveness playing a role in that a phonetic detail is more likely to be imitated if it is phonologically relevant. They found that the difference of whether or not there is pre-voicing on Dutch consonants was imitated (a phonologically relevant contrast) while a difference in the amount of pre-voicing (not phonologically contrastive) was not. They also found that /r/-colored segments (alveolar vs. uvular trill) that are acoustically distinct but phonologically allophones in Dutch were not differentiated in the speakers' convergence patterns, i.e., both segments were mapped to the same representation. However, their study is limited to only one pair of segments for each of their arguments and thus the results are difficult to generalize even within the language. Similarly, Adank, Hagoort & Bekkering (2010) used contrastive vowel length in Dutch as a variable for convergence and demonstrated that vowel duration convergence occurs in a language with contrastive vowel length, but did not discuss the potential link between the exhibited convergence patterns and the contrastiveness. Nielson (2011) hypothesized in a study on VOT that convergence would be selectively suppressed if the direction of imitation would result in a reduced distance between phonological categories, as evident from participants converging to artificially lengthened but not shortened positive VOT in English. D'Imperio et al.

(2014) conducted shadowing experiments at the phrase-level of contrastive intonation in Italian and found convergence in both phonetic and phonological intonation representations, as reflected from details of tonal alignment and pitch scaling. Finally, Podlipský and Šimáčková (2015) commented on the relationship between phonological contrast preservation and the perceptual salience of phonetic details, and presented evidence consistent with Nielson (2011) but contrary to Mitterer and Ernestus (2008) (although the three studies investigated different languages). Podlipský and Šimáčková hypothesized in addition to Nielson's argument that the likelihood of imitation of any given feature could depend on its perceptual saliency; for example, the lengthening of a target segment is more easily detected in perception than its shortening.

In sum, the current literatures offer suggestive evidence on the relationship between phonological contrast and convergence effects from various angles, but remain rather sporadic in the measures and languages that were investigated. The current study will attempt to expand in this stream of literature by investigating the details of convergence patterns within phonologically contrastive measurements of Mandarin. It will first compare the convergence patterns across measures in Mandarin with those found in other languages in general, and then discuss the potential differences of these patterns in the particular contrast of tonal categories (F0) specifically.

Besides F0, other phonetic variables that have been identified as able to present convergence effects, including formant values (mostly F1 and F2, e.g., Evans & Iverson 2007; Kim 2012; Babel, 2012; Sonderegger, Bane & Graff 2017), voice onset time (VOT, e.g., Shockley et al. 2004; Nielsen 2011; Yu et al. 2013; Balukas & Koops 2014), vowel duration (e.g., Evans & Iverson 2007; Pardo 2010; Adank et al. 2010; Podlipský & Šimáčková 2015), speech rate (Cohen

Priva et al. 2017; Schweitzer & Walsh 2016; De Looze et al. 2011), turn-taking and pause duration (Edlund et al. 2009; Street 1984; Gregory & Hoyt 1982), etc. The current study, given its goal of first verifying previously found convergence patterns in an alternative language, will focus on monosyllabic Mandarin words given and repeated in isolation, similar to the design of Kim (2012). Thus, the variables able to be tested are F0, F1, F2, VOT and vowel duration.

2.3 Mandarin tones

The “standard” variety of Mandarin has four contrastive tonal categories: tone 1 is a high-pitched flat tone; tone 2 is rising from mid to high; tone 3 is a dipping tone, going from mid-high to low and then back to high when in isolation[1]; tone 4 is falling from high to low. Traditionally, Chao (1930) denotes the them as High (55), Rising (35), Low (21), and Falling (51). The Beijing dialect, the main dialect investigated in the current study, is a close approximation of the “standard” dialect especially with regards to tone. Figure 1 shows the pitch contours in the Beijing variety, respectively.

[1] Tone 3 is subjected to tone sandhi, in which the final rising either disappears or becomes the entire tone in connected speech. In Figure 1, tone 3 is shown in the first case, but nonetheless has the flatter slope and lesser range when compared with tone 4. Since the study asks participants to produce monosyllabic words in isolation, the final rising is generally present in the current data. In traditional denotation, tone 3 with final rising is denoted with superscript 213.

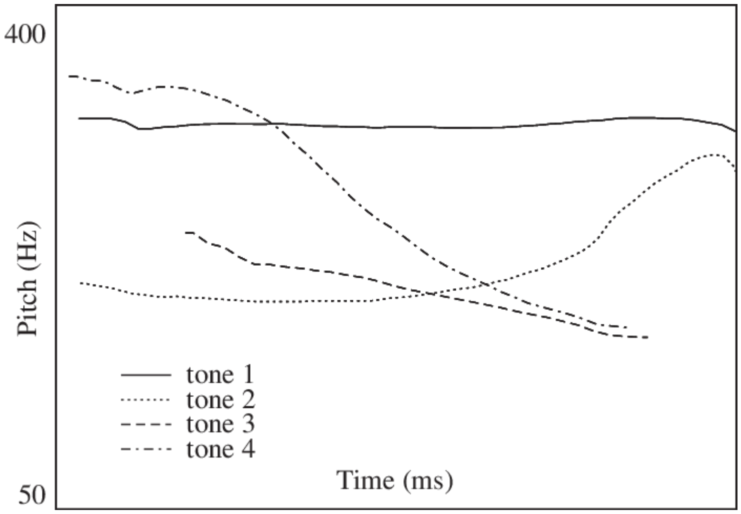


Figure 1. pitch contour for Mandarin tones (from Nixon et al. 2014)

The notion of “standard” Mandarin in the context of the current paper refers to the standardized pronunciations of Mandarin words as they are uniformly taught in schools. The “standard” is mostly based on the Beijing variety and is completely consistent with the Beijing dialect in terms of pitch and vowel qualities. Differences in these variables can arise in other dialects of Mandarin, and participants have generally indicated exposure to at least one other dialectal variety in their daily lives (e.g. in school or at home). However, the participants all come from universities in Beijing where the standard variety is predominately spoken, and have all received K-12 education in the “standard” dialect. The heavy reliance on the “standard” dialect in school settings could create a potential confound in the natural activation of the standardized pronunciations whenever a written form is presented, and thus produce at a reading register that may differ from the natural speaking register (see Christensen (1994) for a more thorough discussion on the differences between written and spoken Mandarin); this will be revisited later in the discussion as it may have an effect on the results of convergence.

Tone and pitch accents are generally reflected as F0 contours. F0 has been identified as a variable subjected to convergence effects by a number of studies on non-tonal languages (e.g., Collins 1998; Bulatov 2009; De Looze et al. 2011; Simonet 2011; Babel & Bulatov 2011; Kim 2012; Pardo 2013; etc.). These studies undertook various measurements of F0, such as the maximum (Kim 2012; Pardo 2013), median (De Looze et al. 2011), mean (Collins 1998), range (Vaughan 2011), or a combination of the above (Simonet 2011; Xia et al. 2014). However, as the majority of them focused on non-contrastive F0, the discussion of convergence often remained on the suprasegmental level of intonational patterns rather than the F0 in isolated words in the sources mentioned above (except for Babel and Bulatov 2011 which focused on individual word-level F0 and Kim 2012 which focused on vowel F0). Xia et al. (2014) was the only study that investigated both local (within two pauses) and global (across conversation) F0 contours with a comparison between a tonal and a non-tonal language (Mandarin and American English), and they found similar patterns across conditions and across languages; however, they did not hypothesize whether having the phonological difference may or may not have influenced convergence since they did not investigate at the word / syllable level.

2.4 Experimental methods for phonetic convergence

Goldinger (1998) pioneered the shadowing paradigm for investigating phonetic convergence. The shadowing paradigm refers to subjects immediately repeating what they heard after an utterance was given, often an individual word or a short phrase. Studies on convergence that have employed this methodology include Fowler et al. (2003), Shockley et al. (2004), Mitterer & Ernestus (2008), Babel (2009), Bulatov (2009), Babel & Bulatov (2011), Babel

(2012), to name a few. A repetition or mere exposure paradigm has also been popular amongst studies of convergence. This paradigm lets listeners be continuously exposed to target for a period of time and then produce the utterances altogether. This method can potentially reveal prolonged effects of convergence when comparing between blocks of exposure, and is thus favored by studies zooming in on specific phonetic factors, such as Goldinger & Azuma (2004), Bulatov (2009), Nielson (2011), Abrego-Collier et al. (2011), Kim (2012), etc.

To date, only a couple studies have directly compared across experimental methods used in the context of convergence. Sanker (2015) found a significant effect of task in measures related to turn-taking when participant pairs first go through a guided Q&A task and then held free conversations within each other. Pardo et al. (2018) compared across a shadowing task and a conversational task but only found a moderate relationship between task condition and phonetic convergence in male talkers (and not in female talkers). Although both of these studies focused on the within-subject or within-pair convergence patterns across task, both found that individuals are not entirely consistent, suggesting that task could be a factor influencing convergence. Other studies that has participants go through similar tasks (in terms of actions) but make adjustments in social context (e.g., Babel 2009, Abrego-Collier et al. 2011) have found situational differences, although not discussing task difference strictly. The current study randomly assigned participants to either the shadowing or the exposure paradigm with the same set of stimuli, same amount of repetition for each target word and the same model talker, therefore seeking to compare the two paradigms directly. Asking this question has implications for future experimental design in convergence research as outlined in previous sections.

3 Procedure

The current study builds on the findings of many previous studies on the phonetic convergence of context-free utterances. It seeks to confirm these findings and further explore the possibilities of differing results in a tonal language, namely Mandarin. Specifically, it explores whether or not having the additional contrastive layer in the phonology may lead to tonal category-specific convergence patterns. Finally, the study also seeks to compare two commonly employed experimental methods for phonetic convergence.

3.1 Participants

A total of 42 Mandarin speakers were recruited in Beijing, China, randomly assigned to one of the two paradigm groups (shadowing and exposure). 3 participants from each group were excluded from the data analysis for reasons of poor audio quality and interference during the experiment. This resulted in data from 18 participants in each group. All participants were adult undergraduate or graduate students (age 18-30, 19 female) with normal speech and hearing. They were recruited through social media and flyers from an area of Beijing densely occupied by students in top-tier colleges. All participants received monetary compensation for their time.

A questionnaire was given to the participants after completing the experiment to record their language background information for post-hoc analysis purposes. All participants received L1 input in Mandarin since infancy and throughout early childhood, although some also received L1 input in certain dialectal variations of Mandarin. All participants also received various degrees of L2 English education beginning in early school years (around grade 1-3 of elementary school). However, Mandarin remained the primary language choice in all participants’ daily lives, and none of the participants reported to have used a dialect or language more fluently and frequently than Mandarin at any point or place in their lives so far.

3.2 Stimuli

A set of 58 monosyllabic Consonant-Vowel (CV) Mandarin words was recorded by a 26-year-old female Southern-accented Mandarin model talker who was blind to the purpose of the study. The recording was made in one sequence with an Audio-technica AT2005 USB microphone and the same device is used for all participants. The words were read as a list, one at a time, with slight pauses in between so that listing effects on prosodic structure such as intonation are reduced. Post-experiment feedback from the participants suggest that they thought the model talker’s pronunciation were of the "standard” variety. Acoustically, the only noticeable difference of the model talker’s accent from the “standard” is in the word /ʈʂʰɻ̩/ (吃, to eat) in which the model talker’s pronunciation is slightly fronted and with less retroflex (more similar to [tsʰɹ̩]).

Relatively high-frequency and unambiguous words were chosen to ensure that participants would accurately produce the intended item. While this decision was important for reliable elicitations, it could weaken convergent effects, as various previous studies have found convergence to be more apparent in lower frequency words (e.g. Goldinger 1998, Nielson 2011). The results suggest an interaction between word frequency and the degree of convergence in a couple measurements, which will be discussed in the next section. Polyphones (words with multiple possible pronunciations) were excluded from the word list, as the participants were asked to read from characters at certain points during the experiment. Nasal consonants and diphthongs were also excluded from the stimuli. For the purpose of measuring VOT and vowel qualities, specific accounts of the variety of initial consonants and tonal category of the vowels were taken, so that there is a relatively equal number of each consonant and tone involved. Additional efforts were taken to make sure that there is no nasality in the stimuli words. A complete list of stimuli used in the experiment is attached as appendix 1.

3.3 Task

The experiment was conducted in a quiet room. Participants sat in front of a computer screen with the recording device placed next to the screen. Auditory stimuli were played through Bose noise-canceling headphones (QuietComfort 35) to prevent any Lombard effects from the environment, but participants are still able to hear themselves albeit to a slightly limited extent. Participants were told that they would be pronouncing some words in Mandarin and they agreed to being recorded. They read and signed a consent form given in Mandarin.

Task instructions were given on the computer screen in both English and simplified Chinese characters. The researcher also gave verbal instructions in Mandarin prior to beginning the experiment, as the experiments were conducted in a predominately Mandarin environment. The stimuli were presented in Psychopy (Peirce, 2007).

For both groups, the experiment began with a pre-test section where subjects were instructed to read the stimulus words one at a time as they appeared on the screen (to establish a baseline, or starting distance from the model speaker), and end with a post-test section with another reading of the full list in the same way (read from text, in isolation) as in the pre-test (ending distance from the model speaker). Between the pre-test and the post-test, there were three test sections that go through the set of stimuli for a total of three times, where the instructions to subjects differ by group. For the shadowing group, participants heard a word and immediately repeated it without seeing the Mandarin Characters for the word. The same procedure repeated for the three test sections, further broken down into 9 rounds with 20 words per round; items from the complete 60-word stimuli inventory were randomly assigned into rounds. Therefore, each word was heard and shadowed 3 times. For the exposure group, participants listened to 20 words blocks without seeing anything on the screen, and read out the characters corresponding to the words one by one after hearing all 20 of them. This was then repeated for 9 rounds, so the total amount of exposure to each stimulus remains the same across the two groups. Participants in each group heard the same 20 words in each round, presented in randomized order.

The linguistic background questionnaire was given after the completion of the experiment in Mandarin. The questions targeted specific information about the subjects' language use as children, choice of language with relatives and friends, major linguistic environments that they have lived in, self-rated language proficiency and learning skills, etc. An English copy of the questionnaire is attached as appendix 2. Responses indicate that participants generally received Mandarin education starting from kindergarten, and use Mandarin as the dominant variety when there are other dialectal varieties spoken at home or at school (e.g., Cantonese and Wu). Self-rated language proficiency was all at the native level, and the language learning skill were rated in the range of average to excellent (numerically 3-5 on a scale of 1-5). The subjects were not explicitly told the goal of the research until after completing the test procedure and returning the questionnaire.

3.4 Data analysis

Acoustic data of F0, F1, F2, VOT, and vowel duration were obtained from the model talker and the participants' utterance. Recordings were labeled and segmented by hand under consistent criteria, and measurements were taken using automated scripts in Praat (Boersma & Weenink, 2018). The author was the sole handler of segmenting and scripting the acoustic data.

Data points that resided above and below 1.5 times the interquartile range outside of the range within speaker *and* within variable category for the speaker were considered outliers and were omitted. Hand-corrections were made to data points with clear measurement errors from the automated scripts (e.g., when the formant extractor mistakes the F3 for F2). The data were then analyzed with a "difference of difference" approach; that is, comparing the difference of each individual's measurements from the model talker's at pre-test with the same sets of differences at post-test for each item.

The F0, F1 and F2 measurements were first log-transformed and then z-normalized (within speaker and across phonological categories) before calculating the "difference of difference," to eliminate individual variations in physical properties that may influence the values, such as vocal tract length, vocal fold thickness, etc. Similar normalizations have been conducted in studies such as Babel (2009). As previously stated, a baseline difference was established to verify that the participants did not start out at a position that is too similar to the model talker, such that they would not have room for significant convergence, but it was not used as a screening criteria for the data.

3.5 hypotheses and predictions

(1a) As previous literature suggests, variation by measure exists in convergence (e.g., Sanker 2015, Kim 2012, etc.), but it may be shaped purely by exposure (and not phonological contrast). If so, convergence should occur in Mandarin as it does in other languages, as reflected in a decrease in participant-model talked distance from pre-test to post-test, but with varying strength across measure. That is, there would be more evidence for convergence in variables such as F0 since listeners are exposed to them in every word, but less evidence in F1 and F2 for any particular vowel since only some of the stimuli contain that vowel. Even with the breakdown of tone categories, there are still more accounts of exposure to each tone than to each vowel (there are seven different vowels included in the stimuli). Evidence for this type of variance are found in literature for cumulative priming (e.g. Kaschak et al., 2011; Oben and Brône, 2016), where extended exposure to priming accounts for structural or lexical-level alignment, respectively.

(1b) Contrary to (1a), variation *can* be shaped by the phonological contrasts, as contrasts mediate the amount of attention that listeners place on the relevant phonetic details. There is evidence in the literature for participants exhibiting more convergence when they have more ability to pay attention to the conversation or stimuli (e.g., Abel & Babel 2017, Heath 2017). There is also evidence that individual differences in the ability to pay attention to phonetic details affect speech perception and therefore are predictive of convergence (Yu et al., 2013). This would be reflected in the current data if relevant measurements exhibit different convergence patterns when broken down by tonal categories, since tone creates a phonological contrast that speakers are forced to pay attention to. Following this hypothesis, more convergence should be found in F0 in Mandarin than in English, but it would be shaped by the tonal categories such that convergence is stronger to the mean within each category but not to the overall mean.

(2) Hypothesis: convergence is affected by priming from both the speakers themselves and the given input from others. If only the input from others matter, task differences should not occur. That is, both the shadowing and the exposure paradigm should reveal patterns of convergence, and the degree of convergence would not differ. However, studies have established that hearing the feedback from our own production influences our future production (e.g., Reitter et al. 2006), and that self-produced imitation better facilitates comprehension and word-recognition than passive exposure of input from others (e.g., Adank et al. 2010, Nguyen et al. 2012). In the exposure group, not having this feedback from the participant’s own speech can thus lower the effects on convergence. From the angle of attention and focus, the exposure group is also more likely to lose attention within each trial as they hear 20 words each time versus one in the shadowing condition, and the self-priming effects are nonexistent.

It is important to note that both of these hypotheses are based purely on phonetic/linguistic factors, and any social factors such as personal preference towards the model talker, difference in gender, or distance of language backgrounds, etc., have not been considered as primary variables of investigation. Although these psycho-social influences have been well established by many previous studies (as mentioned earlier in the paper), the current study focuses on the acoustic factors of convergence while acknowledging that these effects may be present. Their possible influences on current findings will be briefly discussed later in the paper.

4 Results

As convergence is indicated by a “difference of difference” approach, the data were first examined with t-tests within each variable and each group to confirm the existence of convergence. They were then fitted as predictors of convergence patterns using a linear mixed-effects model, which was conducted in R using the lmer() function from the lme4 package (Bates et al., 2015). Group effects will be mostly discussed in the mixed-effects model as the t-tests were run separately to focus on the difference of within vs. across tonal categories.

4.1 F0 (pitch)

To thoroughly explore the potential differences between a tonal and non-tonal language, I investigated the F0 maximum and minimum within the sonorant section of each word. F0-maximum is a common measurement for F0 in convergence studies (e.g., Kim 2012; Pardo 2013) and together with F0-minimum, the total range has been investigated in convergence settings as well (e.g., Vaughan 2011).

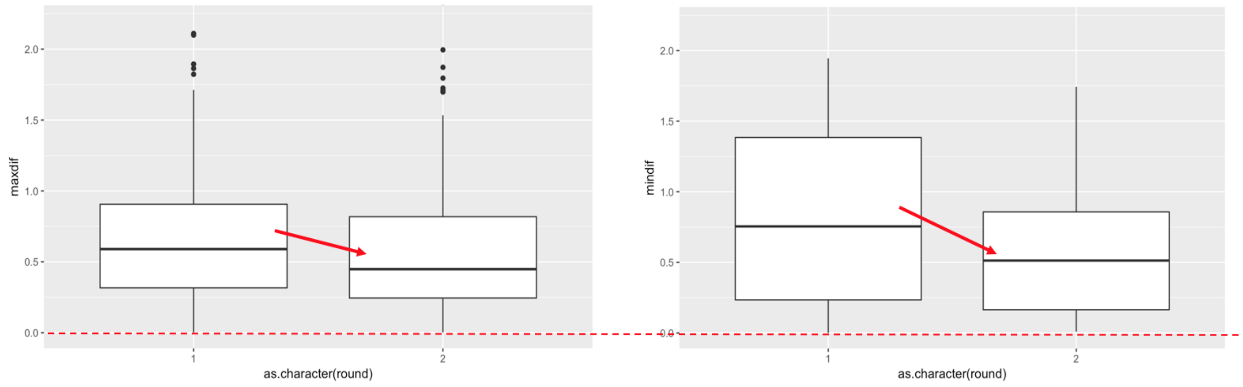
The analysis was first carried out across and then within each of the four tonal categories of Mandarin for both data groups. The baseline differences were significant across categories in F0-maximum but not in F0-minimum for both groups. Convergence was found in

F0-maximum *across* tonal categories (t = 2.641, p = 0.008), consistent with previous works in English and other languages. When broken down to each tonal category, the strongest evidence for convergence in F0 was found in the maxima. Table 1 shows the pre-test average difference (baseline), post-test average difference for both groups in each tonal category.

In F0-maximum, statistically significant convergence patterns occurred in tone 4 for both groups and a convergent trend was found in all of the other three tones except for tone 3 in the exposure group (see figure 2). A convergent trend is indicated by the average differences between the subjects and the model talker *decreasing* from pre-test to post-test. F0-minimum showed a significant convergence in tone 3 in the shadowing group only.

F0 minimum						
Tone	**Shadowing**			**Exposure**		
	Pre-test avg. diff.	Post-test avg. diff.	p-value	Pre-test avg. diff.	Post-test avg. diff.	p-value
1	0.499	0.499	1	0.502	0.508	0.902
2	0.325	0.344	0.638	0.328	0.36	0.371
3	1.02	0.362	0.009**	0.847	0.814	0.553
4	0.584	0.612	0.565	0.648	0.71	0.221
F0 maximum						
Tone	**Shadowing**			**Exposure**		
	Pre-test avg. diff.	Post-test avg. diff.	p-value	Pre-test avg. diff.	Post-test avg. diff.	p-value
1	0.782	0.727	0.466	0.812	0.707	0.158
2	0.419	0.381	0.198	0.448	0.388	0.049*
3	0.799	0.719	0.383	0.681	0.703	0.77
4	0.562	0.493	0.035*	0.577	0.465	0.0003***

Table 1. Average speaker-model talker difference (in seconds) pre-test, post-test, and p-value of the t-test of the *difference between the two average differences* in F0-minimum and maximum for each group. If converging, the post-test average differences should be less than the ones in pre-test.



F0-**max** difference, Tone 4

F0-**min** difference, Tone 3

Figure 2. Speaker-model talker difference pre-test vs. post-test (y-axis) in F0-max in Tone 4 (left) and in F0-min in Tone 3 (right) in shadowing group

It is arguable that the result from the t-tests of F0-maximum is potentially reflecting a pattern of the saliency of the characteristic that "defines" a certain tonal category, or the characteristic of a category that speakers must pay attention to the most. That is, given their respective contours, for tone 4 to be contrastively differentiated from tone 3, the height at which the tone begins is much more important in perception and production compared to the end point of the tone or its length. Thus, the significance of the convergence found in particular measurements within a tonal category may be reflecting that such characteristics are more "defining" than others for this particular tone. This difference in convergence patterns across tones speak to the hypothesis (1b), namely that what people pay attention to is what they tend to converge more on. Particularly, it echoes the argument from Podlipský & Šimáčková (2015) in that the perceptual saliency of a certain contrast affects how much participants converge to the particular contrast.

To further confirm the effects of tonal categories, a linear mixed-effect regression was run with the participants' post-test F0 maximum as the variable and fixed effects of the participants' pre-test F0-maximum, model talker's F0-maximum, tonal categories (as factors), group condition (shadowing vs. repetition), interaction terms (group and pre-test, group and model talker's values, tonal categories and model talker's values), and random effects of participant (speaker) and word item (sound) as predictors (see table 3). The first interaction term denotes the effects of the task condition on how much the participants relied on their own baseline production. The second interaction term shows how much the similarities to the model talker are because of the task condition. The third interaction investigates whether each tonal category influences the effects from the model talker.

```
Formula: maxpost ~ 1 + maxpre + m_f0_max + as.factor(tone) + group +
    group * maxpre + group * m_f0_max + as.factor(tone) * m_f0_max +
    (1 + m_f0_max + as.factor(tone) | speaker) + (1 | sound)
```

```
                                   Estimate   Std. Error    df        t      Pr(>|t|)
Mean of tone levels                3.298e-02  3.322e-02  7.530e+01   0.993  0.32399
Pre-test F0-max                    2.106e-01  2.478e-02  1.620e+03   8.497  < 2e-16***
Model F0-max                       1.302e-01  3.181e-02  6.510e+01   4.093  0.00012***
Tone 1                             6.755e-02  3.872e-02  7.130e+01   1.744  0.08540 .
Tone 2                            -1.206e-01  4.055e-02  6.800e+01  -2.974  0.00406 **
Tone 3                            -2.350e-01  6.867e-02  6.670e+01  -3.422  0.00107 **
Shadowing group                   -1.354e-02  2.552e-02  3.550e+01  -0.531  0.59894
Pre-test F0-max:shadowing group    4.172e-02  2.205e-02  9.736e+02   1.892  0.05881 .
Model F0-max:shadowing group      -8.756e-03  1.608e-02  3.230e+01  -0.544  0.58988
Model F0-max:Tone 1               -1.039e-01  3.618e-02  6.350e+01  -2.871  0.00555 **
Model F0-max:Tone 2                1.820e-01  6.793e-02  5.390e+01   2.679  0.00978 **
Model F0-max:Tone 3               -4.660e-02  3.982e-02  7.200e+01  -1.170  0.24582
```

Table 3. Linear mixed-effects (lmer) model for F0-maximum across groups. The intercept for tone is mean of tone levels and the intercept for group is the exposure condition. All results were calculated by the lme4 package.

The results suggest that patterns revealed from the t-tests are confirmed, and that convergence is found as the variable for the model talker's F0-maximum values is significant (beta = 0.13, standard error (SE) = 0.032, t = 4.093, p = 0.00012).

It would seem that tone factors are significant in determining convergence, for the interactions (*model F0-max: tone 1, model F0-max: tone 2*) show that there is difference across tones. As the comparison is conducted with the average of the tones, it is unclear whether specific patterns exist with any specific tone, but the regression nonetheless demonstrates the existence of categorical differences.

Group (task condition) was not significant in F0-maximum, as evident from the interaction term *model F0-max:shadowing group.* This interaction reveals how much the group difference affects the effects from the model talker. The result indicates that participants from both groups were converging and that they were not differing in the patterns of convergence.

The same regression analysis is run on F0-minimum with the following results:

```
Formula: minpost ~ 1 + minpre + m_f0_min + as.factor(tone) + group +
    group * minpre + group * m_f0_min + as.factor(tone) * m_f0_min +
    (1 + m_f0_min + as.factor(tone) | speaker) + (1 | sound)
```

```
                                  Estimate   Std. Error    df        t      Pr(>|t|)
Mean of tone levels                8.388e-02  4.378e-02  8.000e+01  1.916  0.0590 .
Pre-test F0-min                    1.231e-01  2.567e-02  1.647e+03  4.794  1.78e-06***
Model F0-min                       9.316e-02  8.282e-02  4.950e+01  1.125  0.2660
Tone 1                             6.209e-01  8.460e-02  6.540e+01  7.340  4.22e-10***
Tone 2                            -8.394e-02  8.921e-02  5.250e+01 -0.941  0.3510
Tone 3                            -4.072e-01  9.526e-02  7.020e+01 -4.275  5.92e-05***
Shadowing group                   -5.119e-02  2.527e-02  3.480e+01 -2.025  0.0505 .
Pre-test F0-min:shadowing group    3.496e-02  2.411e-02  1.134e+03  1.450  0.1473
Model F0-min:shadowing group       1.873e-02  2.599e-02  3.740e+01  0.721  0.4757
Model F0-min:Tone 1                6.704e-02  8.577e-02  4.590e+01  0.782  0.4385
Model F0-min:Tone 2                1.551e-01  2.259e-01  4.290e+01  0.687  0.4960
Model F0-min:Tone 3               -5.216e-03  8.862e-02  5.290e+01 -0.059  0.9533
```

Table 4. Linear mixed-effects (lmer) model for F0-minimum across groups. The intercept for tone is mean of tone levels and the intercept for group is the exposure condition.

Here it is worth noting that the main effects of convergence are not significant (*model F0-min*), consistent with what the t-test suggests overall. The significant convergence in tone 3 in the t-test may be a result of repeated measures as the regression does not show differences across tonal categories (sum contrasts were used so the intercept is not any one tone in particular).

The existing convergent patterns across categories in F0-max but not F0-min may simply suggest that the subjects are converging to a higher pitch from the model talker, even though that is not always the case with every participant's data. As people normally speak at a lower pitch range relative to our own maximum (Russel & Stathopoulos 1988), there is a lot more room to move upwards than downwards when shifting within our pitch range. The differences in convergence *within* tonal categories may be explained by a within-category change in the target or exemplar. Having the phonological contrast did not induce a change in the degree of *subjectivity* to convergence, but that the convergence is "cut up" with different targets for each category.

4.2 F1 & F2

There are a total of 7 different vowels included in the stimuli set. Only one statistically significant convergence was found across tones in the t-tests (in F1 of /o/ in the shadowing group; see table 5 for F1, table 6 for F2, and figure 3). This may suggest that a lack of overall convergence effect. However, when broken down by tones, more trends emerge consistently within each tonal category and vowel.[2]

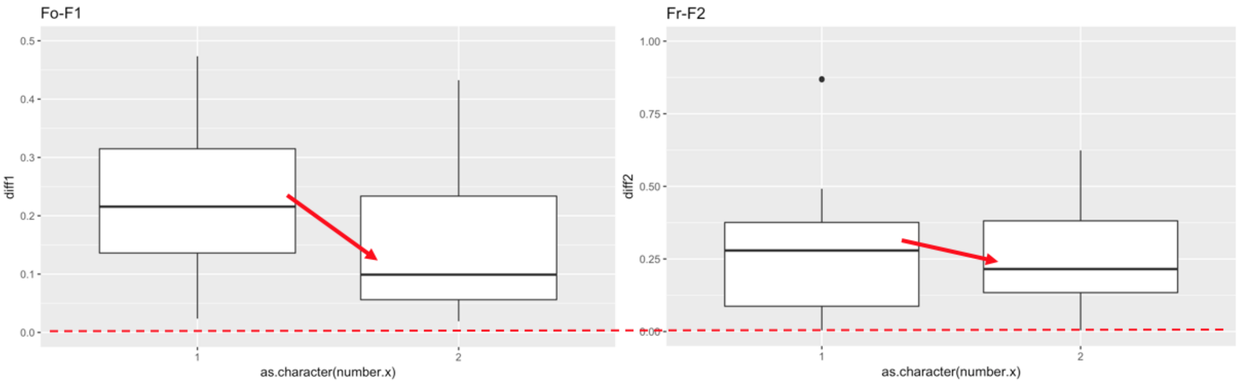
[2] A by-item analysis will also be conducted in the next phrase of the study to investigate whether the strength of the convergence effects differ at a lexical or phoneme level. Previous studies have targeted particular lexical tokens that are regarded as more susceptible to accent/dialectal variations. Although this was not among the main hypotheses in the current study, there could nonetheless be lexical-level effects that were masked by the grouping by tone or by vowel.

F1						
Vowel	**Shadowing**			**Exposure**		
	Pre-test avg. diff.	Post-test avg. diff.	p-value	Pre-test avg. diff.	Post-test avg. diff.	p-value
a	0.26	0.287	0.396	0.3342	0.3402	0.9097
ɤ	0.284	0.272	0.614	0.3104	0.2843	0.4123
i	0.173	0.194	0.275	0.235	0.2984	0.3622
o	0.368	0.183	0.003**	0.4366	0.3192	0.2601
u	0.271	0.316	0.12	0.2315	0.2519	0.4357
y	0.224	0.244	0.465	0.2685	0.3073	0.5686
ɻ̩	0.209	0.411	0.195	0.2005	0.2773	0.5478
ɚ	0.353	0.292	0.637	0.2473	0.1669	0.1071

Table 5. Average speaker-model talker difference (in seconds) pre-test, post-test, and p-value of the t-test of the difference between the two average differences in F1 for both groups across tones.

F2						
Vowel	**Shadowing**			**Exposure**		
	Pre-test avg. diff.	Post-test avg. diff.	p-value	Pre-test avg. diff.	Post-test avg. diff.	p-value
a	0.257	0.255	0.925	0.2866	0.3316	0.2935
ɤ	0.183	0.227	0.015	0.2314	0.243	0.7436
i	0.157	0.162	0.802	0.2341	0.2511	0.7962
o	0.349	0.271	0.193	0.5543	0.3253	0.1845
u	0.92	1.015	0.451	1.0884	1.0886	0.9987
y	0.241	0.25	0.724	0.2499	0.2802	0.5927
ɻ̩	0.194	0.146	0.32	0.2848	0.2243	0.2978
ɚ	0.304	0.245	0.3	0.2289	0.2786	0.5322

Table 6. Average speaker-model talker difference (in seconds) pre-test, post-test, and p-value of the t-test of the difference between the two average differences in F2 for both groups across tones.



differences in F1 in vowel /o/

differences in F2 in vowel /ɚ/

Figure 3. examples of speaker-model talker difference pre-test vs. post-test (y-axis) in shadowing group

A linear mixed-effect regression was run with the participants' pre-test F1 and F2 (separately), model talker's F1 and F2 (respectively), vowel categories (as factors), group condition (shadowing vs. repetition), interaction terms (group and pre-test, group and model talker's values, tonal categories and model talker's values), and random effects of participant (speaker) and word item (sound) as predictors. The interaction terms denote similar inquiries as those in F0 regressions. The intercept of the current regression is the mean of the vowels (using the sum contrast instead of treatment in the R package *lme4*).

```
Formula: F1_post ~ 1 + F1_pre + m_F1 + as.factor(vowel) + group + group * F1_pre +
        group * m_F1 + as.factor(vowel) * m_F1 + (1 + m_F1 + as.factor(vowel) |
        speaker) + (1 | sound)
```

```
                                   Estimate   Std. Error    df        t      Pr(>|t|)
Mean of vowels                      3.766e-02  3.879e-02  3.767e+02  0.971   0.332265
Pre-test F1                         2.504e-01  2.554e-02  1.833e+03  9.804   < 2e-16 ***
Model F1                            4.537e-01  9.651e-02  3.400e+01  4.700   4.19e-05***
Vowel 1                             8.822e-01  1.549e-01  1.660e+01  5.696   2.85e-05***
Vowel 2                             4.764e-02  4.927e-02  1.697e+02  0.967   0.334995
Vowel 3                            -3.730e-01  8.386e-02  2.210e+01 -4.448   0.000200***
Vowel 4                            -1.523e-02  7.889e-02  1.223e+02 -0.193   0.847224
Vowel 5                             3.731e-01  1.546e-01  5.636e+02  2.413   0.016156 *
Vowel 6                            -2.935e-01  6.135e-02  1.389e+02 -4.784   4.33e-06***
Shadowing group                    -1.523e-02  1.147e-02  4.020e+01 -1.327   0.191832
```

```
Pre-test F1: shadowing group      -1.834e-02  2.235e-02  9.420e+02 -0.821  0.411984
Model F1:shadowing group           4.081e-03  2.542e-02  1.545e+02  0.161  0.872666
Model F1: vowel 1                 -3.369e-01  1.036e-01  7.332e+02 -3.253  0.001192 **
Model F1: vowel 2                 -1.139e-01  1.193e-01  5.478e+02 -0.954  0.340336
Model F1: vowel 3                 -1.131e-01  1.011e-01  1.029e+03 -1.118  0.263678
Model F1: vowel 4                 -3.959e-02  2.470e-01  9.543e+02 -0.160  0.872688
Model F1: vowel 5                  1.345e+00  4.063e-01  1.330e+03  3.311  0.000955***
Model F1: vowel 6                 -3.340e-01  1.187e-01  1.892e+02 -2.814  0.005408 **
```

Table 7. Linear mixed-effects (lmer) model for F1 across groups. The intercept for vowel is mean of all vowels and the intercept for group is the exposure condition.

```
Formula: F2_post ~ 1 + F2_pre + m_F2 + as.factor(vowel) + group + group * F2_pre +
        group * m_F2 + as.factor(vowel) * m_F2 + (1 + m_F2 + as.factor(vowel) |
        speaker) + (1 | sound)
```

```
                                  Estimate   Std. Error    df       t      Pr(>|t|)
Mean of vowels                    -3.955e-02  4.435e-02  2.710e+02 -0.892  0.373205
Pre-test F2                        2.526e-01  1.937e-02  1.475e+03 13.042  < 2e-16 ***
Model F2                           2.640e-01  7.557e-02  1.251e+02  3.494  0.000659***
Vowel 1                           -4.278e-02  5.341e-02  1.578e+02 -0.801  0.424424
Vowel 2                           -2.666e-01  6.317e-02  1.770e+02 -4.220  3.89e-05***
Vowel 3                            1.055e-01  1.057e-01  1.012e+02  0.998  0.320733
Vowel 4                           -1.374e-01  2.209e-01  3.897e+02 -0.622  0.534186
Vowel 5                            2.808e-01  1.089e-01  1.389e+02  2.577  0.011000 *
Vowel 6                           -5.838e-01  1.141e-01  4.840e+01 -5.119  5.27e-06***
Shadowing group                   -4.462e-03  1.211e-02  5.680e+01 -0.369  0.713855
Pre-test F2:shadowing group       -7.612e-02  1.771e-02  1.186e+03 -4.297  1.87e-05***
Model F2:shadowing group           5.998e-02  2.315e-02  9.190e+01  2.591  0.011138 *
Model F2: vowel 1                 -1.341e-02  1.044e-01  1.568e+02 -0.128  0.897975
Model F2: vowel 2                 -1.739e-01  1.150e-01  4.297e+02 -1.512  0.131227
Model F2: vowel 3                  3.577e-01  7.592e-02  2.133e+02  4.711  4.44e-06***
Model F2: vowel 4                  3.095e-01  1.565e-01  6.748e+02  1.978  0.048388 *
Model F2: vowel 5                 -8.040e-02  2.005e-01  1.717e+02 -0.401  0.688922
Model F2: vowel 6                 -2.532e-01  6.745e-02  1.357e+02 -3.754  0.000257***
```

Table 8. Linear mixed-effects (lmer) model for F2 across groups. The intercept for vowel is mean of all vowels and the intercept for group is the exposure condition.

Overall convergence is found in both F1 and F2, as the main effects of *m_F1* and *m_F2* are significant (see table 7 and 8). Convergence also seemingly varies by vowel categories, as some of the m_F1: vowel and m_F2: vowel interactions show significant differences from the intercept (average of the vowels). This would suggest that there are variations in convergence patterns by vowel category, although the specific effects of any one vowel is not evident as limited by the number of stimuli in the current data. Future studies can include a larger dataset with more even distributions of vowel categories to further explore this effect.

The group conditions revealed an interesting note in F2 that there seems to be slightly more convergence in the shadowing group, although the effect is relatively weak (*model F2: shadowing group*, beta = 0.06, SE = 0.02, t = 2.59, p = 0.01). This could be explained by the strong effect of task on the prediction of pre-test (pre-test *F2: shadowing group*, beta = -0.08, SE = 0.01, t = -4.3, p = 1.87 * 10^{-5}); that is, participants' pre-test F2 is less predictive of post-test F2 in the exposure group, and this effect could be outweighing convergence and thus causing the effects of less convergence in this group. Alternatively, the effects of self-priming as found in previous studies (e.g., Reitter et al. 2006, Nguyen et al. 2012) can have an effect (since the shadowing group involves more self-priming than the exposure group), although the effect is small and thus rather skeptical.

Overall, the results from F1 and F2 show that the Mandarin vowel formants *are* subjected to convergence, as the main effects of the model talker's F1 and F2 are significant respectively, but the extent to which convergence occurs varies with vowel categorical distinctions. Lexical frequency was investigated as a potential factor to cause the variability, as the one vowel (/o/) that showed promising convergent trends across groups were contained in lexical items that are of relatively low frequency, but adding lexical frequency did not improve the regression model. Additionally, the issue of repeated measures may have led to some of the sporadic t-test results (e.g., the positive trend in /ɤ/ F2). Further studies will systematically investigate the factor of tones on top of vowel categories with more stimuli and higher statistical power.

4.3 VOT

Mandarin makes contrastive distinctions in stop consonants across two dimensions: aspiration and place of articulation. The phonemic inventory of word-initial stop consonants thus include /p/ (*pinyin 'b')*, /t/ (*pinyin 'd')*, /k/ (*pinyin 'g')*, /p^h/ (*pinyin 'p')*, /t^h/ (*pinyin 't')*, and /k^h/ (*pinyin 'k')*. In the current data, the baseline differences for all of these consonants were significant when grouped by aspiration, as shown in table 9. Table 10 shows the breakdown of the difference-of-differences by consonant. No significant indications of convergence were found in the current data.

VOT baseline						
	shadowing			**exposure**		
Consonant	**Participant's Pre-test avg.**	**Model talker's avg.**	**p-value**	**Participant's Pre-test avg.**	**Model talker's avg.**	**p-value**
aspirated	0.132	0.1454	6.40E-12***	0.1332	0.1452	8.22E-10***
unaspirated	0.0274	0.024	0.0019**	0.0284	0.0244	0.0005***
VOT test						
	shadowing			**exposure**		
Consonant	**Pre-test avg. diff.**	**Post-test avg. diff.**	**p-value**	**Pre-test avg. diff.**	**Post-test avg. diff.**	**p-value**
aspirated	0.03	0.0304	0.8764	0.0292	0.0269	0.2633
unaspirated	0.0108	0.0103	0.5778	0.0109	0.01	0.4473

Table 9. Average speaker-model talker difference (z-normalized seconds) and p-value of the baseline and t-test of the difference of difference in VOT.

	Shadowing			Exposure		
Consonant	**Pre-test avg. diff.**	**Post-test avg. diff.**	**p-value**	**Pre-test avg. diff.**	**Post-test avg. diff.**	**p-value**
p^h	0.0267	0.0274	0.8247	0.0278	0.0232	0.1446
t^h	0.0315	0.0313	0.9544	0.0287	0.0312	0.5133
k^h	0.0337	0.0343	0.8643	0.0323	0.0273	0.2346
p	0.01	0.0098	0.8384	0.0099	0.0096	0.8346
t	0.0068	0.0059	0.3105	0.0068	0.0063	0.5958
k	0.0211	0.0205	0.8279	0.0215	0.0187	0.4794

Table 10. Average speaker-model talker difference (z-normalized seconds) pre-test, post-test, and p-value of the t-test of the difference of difference in VOT broken down by each stop consonant.

The regression analysis confirms what the t-tests reveal, namely that convergence is not significant in the current VOT data as the main effect of model talker's VOT (denoted as duration_m) is not significant (see table 11). This remained true when the regression is run separately by aspiration, with the particular stop (/p/, /t/, /k/, /p^h/, /t^h/, and /k^h/) being factors. Word frequency also did not significantly contribute in post-hoc analysis.

```
Formula: post ~ 1 + pre + duration_m + as.factor(aspiration) + group +
    group * pre + group * duration_m + as.factor(aspiration) * duration_m +
    (1 + duration_m + as.factor(aspiration) | speaker) + (1 | sound)


                                  Estimate  Std. Error   df       t     Pr(>|t|)
Exposure group                    0.181729   0.052480   20.4    3.463   0.0024 **
Pre-test VOT                      0.144955   0.103437   626.9   1.401   0.1616
Model VOT                         1.134782   0.656955   20.6    1.727   0.0991 .
Aspirated                         0.074280   0.052140   19.9    1.425   0.1697
Shadowing group                   0.003494   0.008153   35.3    0.429   0.6708
Pre-test VOT: shadowing group    -0.098530   0.091006   623.4  -1.083   0.2794
Model VOT: shadowing group        0.053331   0.083620   379.4   0.638   0.5240
Model VOT: aspirated              0.037460   0.643499   19.0    0.058   0.9542
```

Table 11. Linear mixed-effects (lmer) model for VOT across groups. The intercept for aspiration is the aspirated condition and the intercept for group is the exposure group.

It is perhaps worth noticing from the t-tests that the aspirated consonants in the exposure group show the most promising likelihood for convergence. As VOT measure is similar to speech rate, with the exposure condition, the model talker's pattern may be easier to be picked up by participants since they are exposed to 20 words at a time rather than in the one-word-at-a-time shadowing condition, however this speculation needs to be supported by more evidence and further investigation.

A possible explanation for the lack of convergence found in VOT is that although the model talker VOTs are lengthened overall, the effects of repetition may have masked the effects of convergence. That is, as speakers go through more repetitions of the same task procedures, they may speed up their production or reduce the level of details in their production. As other studies have found VOT convergence in languages that use aspiration contrastively (e.g., Mitterer & Ernestus 2008, Podlipský & Šimáčková 2015), it is unlikely that the lack of convergence in VOT in the current study has a language-specific cause. However, further investigation is first needed to establish that the model talker's aspiration is indeed longer than average at least within the participants in this study, and also confirm that the participants are indeed subjecting to repetition effects.

4.4 Vowel duration

A significant evidence of convergence in vowel duration was found in the shadowing group, while the exposure group did not reveal significant convergence in the t-tests (see table

12 and figure 4). The baseline t-test comparisons showed that participants in the shadowing group may have started with enough differences so that there is room for convergence but the ones in the exposure group may have not.

Vowel duration baseline						
	shadowing			**exposure**		
	Participant's Pre-test avg.	**Model talker's avg.**	**p-value**	**Participant's Pre-test avg.**	**Model talker's avg.**	**p-value**
baseline	0.2786	0.2929	.0007***	0.2918	0.2948	0.4746
Vowel duration test						
	shadowing			**exposure**		
	Pre-test avg. diff.	**Post-test avg. diff.**	**p-value**	**Pre-test avg. diff.**	**Post-test avg. diff.**	**p-value**
test	0.077	0.0684	.0018**	0.6765	0.6755	0.9713

Table 12. Average speaker-model talker difference (z-normalized seconds) and p-value of the baseline and t-test of the difference of difference in vowel duration.

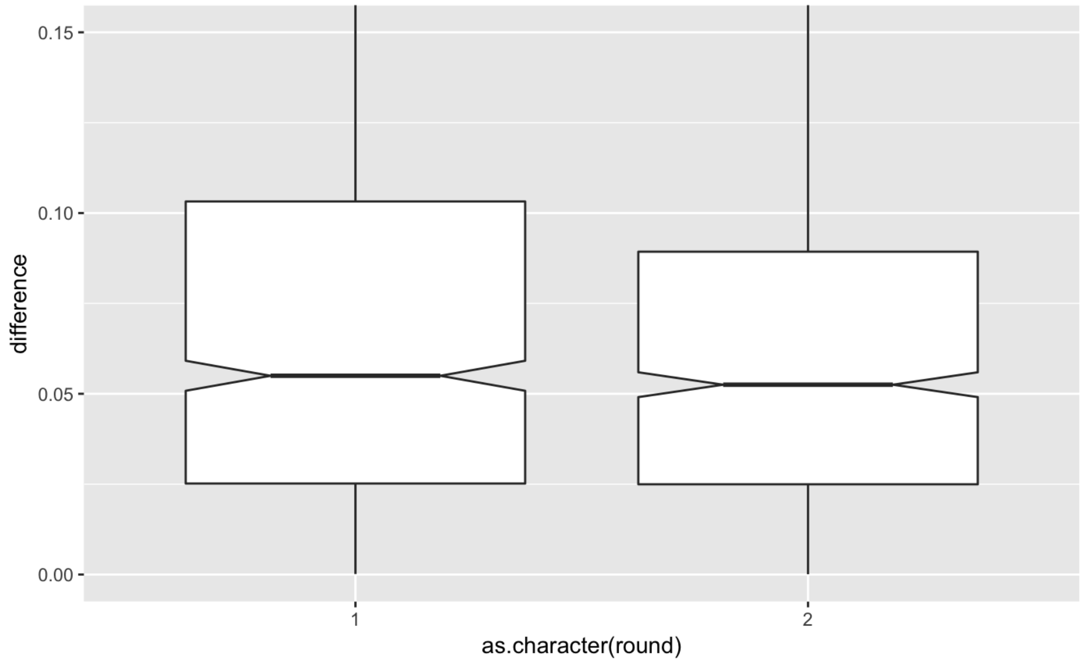


Figure 4. Speaker-model talker difference pre-test vs. post-test (y-axis) in vowel duration in the shadowing group

Table 13 shows the linear regression analysis, where the existence of convergence in vowel duration is evident in that the main effects of the model talker’s vowel duration is significant. The same results held when variables of tone factor or vowel factor were added (either one but not simultaneously) in the sense that all interaction terms with tones or vowels were insignificant, suggesting that tonal or vowel categories did not significantly impact convergence in vowel duration.

```
Formula: post ~ 1 + pre + duration_m + group + group * pre + group * duration_m +
    (1 + duration_m | speaker) + (1 | sound)


                                  Estimate   Std. Error    df       t      Pr(>|t|)
Exposure group                    8.728e-02  1.920e-02  6.050e+01  4.546  2.68e-05 ***
Pre-test duration                 3.075e-01  3.163e-02  1.851e+03  9.722   < 2e-16 ***
Model duration                    3.944e-01  7.545e-02  4.910e+01  5.227  3.52e-06 ***
Shadowing group                  -1.401e-02  2.252e-02  3.700e+01 -0.622   0.538
Pre-test: shadowing group         5.624e-02  4.189e-02  1.785e+03  1.342   0.180
Model duration: shadowing group   7.180e-04  9.383e-02  3.470e+01  0.008   0.994
```

Table 13. Linear mixed-effects (lmer) model for vowel duration across groups. The intercept for group is the exposure group.

The result confirms previous studies that vowel duration is one of the dimensions subjected to phonetic convergence. The exposure group did not show statistical significance in the t-test, but in the regression the groups did not differ. This could be a result of the t-test having less controlling power for effects from individual speakers than the regression analysis, such that any particular speaker’s performance may have a stronger effect on the average. It was also mentioned that the baseline differences in the exposure group was not significant (see table 12), meaning that the group average may not have started off different enough from the model talker's average and therefore did not have enough room to converge.[3] It is also possible that the participants in the shadowing group converged more due to the fact that durational measures such as vowel duration and VOT are often subjected to convergence of speech rate, which is found more often in studies of social settings than non-social (e.g., Cohen Priva et al. 2017; Schweitzer & Walsh 2016; De Looze et al. 2011). The shadowing condition would allow for a more direct and immediate priming of speech rate as the participants repeat each word after the model talker, while the exposure condition would have less of such an effect as the participants utter 20 words at their own pace each time. Further investigation revealed that for the shadowing group, participants were already converging to vowel duration at the first block of shadowing, and the convergence was more significant at both the first and the last (third) block than at post-test (see table 14). This finding is consistent with previous studies that have found a rapid decay of the effects of priming (e.g., Reitter et al. 2006, Bernolet et al, 2016). In the shadowing condition of the current study, participants were immediately primed during test blocks but lost the priming effect to an extent in the post-test reading of the words, although the effects maintained to the extent that convergence to the model talker still significantly exists.

Vowel duration			
shadowing condition	**Pre-test avg. diff.**	**1st block avg. diff.**	**p-value**
pre-test vs. 1st block	0.077	0.0545	<0.0001***

[3] An individual-difference analysis was also conducted with regards to vowel duration and VOT; that is, whether or not there are individual differences in whether the participants each subject to convergence and whether or not those who exhibited convergence in VOT would also converge in vowel duration, or vice versa. Current results were inconclusive, but subsequent analysis will check for individual differences across all variables in the study.

Vowel duration			
shadowing condition	**Pre-test avg. diff.**	**3rd block avg. diff.**	**p-value**
pre-test vs. 3rd block	0.077	0.0533	<0.0001***

Table 14. Average speaker-model talker difference (z-normalized seconds) and p-value of the t-tests of the difference of difference in vowel duration, in first and last block of the shadowing group.

5 Discussion

The results from the current study show across- and within-category evidence of convergence in F0-maximum, F1, F2, and vowel duration. Convergence was not found across categories in F0-minimum and VOT, and only appeared in a few categories when broken down by categories in these two measures. Various categorical differences in convergence patterns also emerged, namely tonal categories affected convergence in F0-maximum and vowel categories affected convergence in F1 and F2. Task differences did not show significant effects on convergence except for a marginal effect in F2.

Collectively, these results answer our hypothesis in the following ways:

(1a) convergence occurs in Mandarin as it does in other languages, reflected in the decrease in participant-model talked distance from pre-test to post-test in multiple measures. Not all measures exhibited convergence but it was expected that the degrees of convergence would vary, as the amount of exposure to each measurement was different. The variance was found but cannot be fully accounted solely by differences in the amount of exposure received, for measures such as formants did show differences by each vowel (hence different amounts of exposure) but others such as F0-maximum and minimum have the same counts of exposure but revealed different results. However, F0-minimum is not one of the most often-used measures in convergence studies, potentially for reasons of 1) participants already at lower range of their F0 and thus have less room to converge downward than upward (as outlined in previous section on results); 2) a difference in attention such that higher tones are paid more attention to than the lower ones (for reasons of saliency of the tone phonological contrast but also general intonational patterns as well); and 3) creakiness or other co-articulation patterns that tend to occur with low-F0 segments may induce additional measurement difficulties and complications. In the current study, heavy creakiness is often found in places where F0 reaches minimum within segments in both the model talker and the participants' speech, especially in cases of tone 3 and 4 (see figure 5). While hand corrections may resolve the issue, and future analysis of the current data will attempt to do so systematically, it is nonetheless the case that the pitch minimum in these segments can be trickier to analyze than the maximum. Future analysis may also investigate whether the amount or length of creakiness can be a factor of convergence, although no previous studies have revealed similar analysis to the best of the author's knowledge so far.

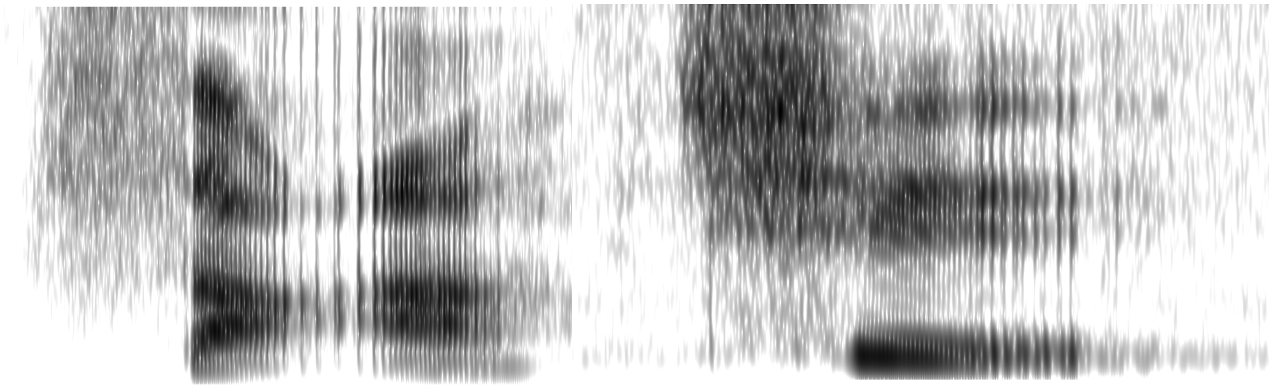


Figure 5. Example of segments that show creakiness at F0-mimimum from two different speakers; the words are /kʰa3/ (left) and /tɕʰy4/ (right).

(1b) convergence is facilitated by attention to detail, and the amount of details to pay attention to is affected by the existence of phonologically contrastive characteristics. It was found in the current data that measurements such as F0 exhibit different convergence patterns when broken down by tonal categories, as tone creates a phonological contrast that speakers are forced to pay attention to. Variations in vowel categories may also be explained by the particular phonological contrasts that speakers must focus on. The current results echo the arguments from Nielson (2011) in that participants may be resistant to imitation of details that may blur the contrastive category boundaries and may also pay less attention to the exact changes in the details once the contrast is established. Podlipský & Šimáčková (2015) further argue that the exact changes in the details are imitated only when they are salient to speakers, by offering evidence that measures such as prevoicing and vowel duration are imitated when they are extended but not when they are shortened. Notably, the previous studies in this line of research focused on duration measures that may simply suggest convergence in speech rate (except for the D'Imperio et al. (2014) study of intonation at the phrase level); the current study, however, by demonstrating similar patterns in F0 and formant measures, extends the line of analysis into phonetic convergence of non-durational measures. Thus, the current results urge for a more comprehensive comparison across phonological contrasts within a language and investigate the potential differences they may induce on convergence patterns. Future directions of the current study will attempt to do so by expanding the data collection and building better contrast comparisons within each measure, taking into account contrasts such as vowel categories and tonal categories at the same time.

(2) both the shadowing and the exposure paradigm revealed patterns of convergence and only differed in the measurement of F2. This result echoes previous studies that have found convergence effects using either or both mechanisms (e.g, Bulatov 2009, Abrego-Collier et al. 2011, Kim 2012, Babel 2012, etc.). For F2, based on task having effects on differences in pre- and post- production, convergence is affected so that the model talker*group interaction became significant. Self-priming could be contributing to this effect although it was not consistently found across measures.

6 Possible factors and future directions

Collectively, the current data suggest that cross-linguistic differences exist in studying convergence especially as the phonologically contrastive categories differ. Specifically, breaking things down by tonal categories made a difference: convergence patterns appeared stronger when observed *within* tonal categories than across them, in F0 and formant measurements. This argument may extend to the individual differences in speakers' sensitivity to convergence in particular contrasts with different language backgrounds. A study by Kim et al. (2011) on the relationship between interlocutor distance and degree of convergence found that people who speak the same dialect of a language tend to converge more than pairs who spoke different dialects or one is a native speaker and one is not. It may be possible that the results obtained from the current study is also subjected to this effect, since some of the participants have a different native dialect from the model talker. The participant survey indicated several participants were exposed to both standard Mandarin and another dialect of Mandarin at an early age, although it is unclear which one would be the native dialect for them. Future investigations should better categorize each speaker's language background and potentially consider them as factors in the regression analysis.

Although the current study found converging patterns in most of the variables that were previously found, the current data does not seem to exhibit as much of an effect as some of the previous studies have shown, such as in VOT. However, convergence effects are somewhat small in general; trends towards convergence often exist but fail to reach statistical significance. That is not to say that convergence was not found, but rather that the effects are subtle and potentially subjected to variations by other factors, such as setting, task, speaker, etc. To find stronger effects of convergence, previous studies such as Kim et al. (2012) and Nielson et al. (2011) implemented a much larger set of stimuli and also included more rounds of shadowing and repetition for each participant than in the current study. In these studies, it has also been found that the effects of convergence increase as more exposure to the stimuli occurs. The amount of exposure in the current study may thus be at a lesser level, although it still resulted in significant evidence of convergence across several measures.

Another potential concern is that Mandarin speakers could be subjected to effects of having separate representation when reading from when speaking. As mentioned in the background, the heavy amount of schooling and strong advocacy for "standard" pronunciations in schools can result in a default activation of the "standard" pronunciations for the native speakers, and this effect can begin immediately when cued to read in Mandarin. In other words, participants may have a "standardized" reading register that does not entirely align with their speaking register. However, an analysis of the findings from comparing the first and third round of the shadowing task condition (purely auditory) revealed no significant differences from the comparison of pre-test and post-test (purely reading), which would suggest that the effect of separate representation (if exists) is not found in the current data.

Finally, it is worth mentioning that the current study is based in non-social settings and does not investigate any social aspects of the measures that affect convergence. Previous studies have argued that social settings facilitate phonetic-level convergence; for example, studies have demonstrated that measurements of F0 correlates with speaker gender, age, perception of personality and occupation, etc. (Pittam, 1994). Babel (2009) showed increasing convergence when only a picture for the model talker is added, suggesting a social component can nonetheless impact the automatic perception-production link when it comes to convergence. However, the goal of the current study was to primarily establish a comparison between the phonetic convergence patterns in a tonal language and those previously found in other languages; previous studies that focus on the automatic components of convergence have commonly restricted the variables of discussion to the phonological-level and below to ensure a more concentrated analysis. Subsequent studies can investigate whether the minimally social components of the shadowing and exposure task designs can affect the convergence patterns, such as potential influences from the researcher’s pronunciations as the experiment is explained and implemented, differences across speaker gender, etc. A recording can be made of the researcher’s utterance and analyzed in comparison to the participants’ data should this effect become a concern.

7 Conclusion

Applying classic task designs to a new language in question with several different phonetic measures, the current study confirms that tonal languages are subjected to phonetic-level imitation effects similarly found in other languages. Convergence patterns were evident in F0-maximum, F1, F2, and vowel duration. In addition, the current study demonstrated that the existence of phonologically contrastive categories could induce differences in convergence effects when separating the analysis within categories. This is hypothesized to be facilitated by effects of selective tuning of attention and the perceptual saliency of the details in these contrasts. Finally, the study investigated potential differences from test conditions, and concluded that both the shadowing and the exposure paradigm were effective in revealing convergence effects in general. Differences were found in durational measures that point to effects of the cumulative but rapidly-decaying nature of priming.

Subsequent analysis will explore a few of the remaining question of the study, namely building a dataset to better compare the effects of vowel-tone category interactions, and conducing by-token analysis of vowel formant measurements to further investigate whether effects of convergence appear at a lexical or a phonemic level.

Works Cited

Abel, J., & Babel, M. (2017). Cognitive load reduces perceived linguistic convergence between dyads. *Language and Speech*, *60*(3), 479-502.

Abrego-Collier, C., Grove, J., Sonderegger, M., & Yu, A. C. (2011). Effects of speaker evaluation on phonetic convergence. In *Proceedings of the 17th International Congress of the Phonetic Sciences* (pp. 192-195).

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, *21*(12), 1903-1909.

Babel, M. (2009). *Phonetic and social selectivity in speech accommodation.* Doctoral dissertation, University of California, Berkeley.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40*(1), 177-189. <doi:10.1016/j.wocn.2011.09.001>.

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech, 55*(2), 231-248. doi:10.1177/0023830911417695

Balukas, C., & Koops, C. (2015). Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism, 19*(4), 423-443. doi:10.1177/1367006913516035

Bates, D., Maechler, M., Bolker, B, Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.<doi:10.18637/jss.v067.i01>.

Bernolet, S., Collina, S., & Hartsuiker, R. J. (2016). The persistence of syntactic priming revisited. *Journal of Memory and Language*, *91*, 99-116.

Boersma, Paul & Weenink, David (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.40, retrieved 11 May 2018 from http://www.praat.org/.

Bulatov, D. (2009). The effect of fundamental frequency on phonetic convergence. *UC Berkeley Phonology Lab Annual Reports*, 5(5), 404-434.

Chao, Y.R. (1930). A system of “tone letters”. *Le Maître Phonétique* 45, 24-27.

Cibelli, E. (2009). *Phonetic convergence during conversational interaction in bilingual speakers* (Doctoral dissertation, Wellesley College).

Cohen Priva, U., Edelist, L., & Gleason, E. (2017). Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *The Journal of the Acoustical Society of America*, *141*(5), 2989-2996.

Collins, R. (1998). Back to the future: digital television and convergence in the United Kingdom. *Telecommunications Policy*, *22*(4), 383-396.

Christensen, M. B. (1994). *Variation in spoken and written Mandarin narrative discourse* (Doctoral dissertation, The Ohio State University).

Decety, J., Chaminade, T., Grezes, J., & Meltzoff, A. N. (2002). A PET exploration of the neural mechanisms involved in reciprocal imitation. *Neuroimage*, *15*(1), 265-272.

De Looze, C., Oertel, C., Rauzy, S., & Campbell, N. (2011). Measuring dynamics of mimicry by means of prosodic cues in conversational speech. In *International Conference on Phonetic Sciences (ICPhS),* Hong Kong. 1294-1297.

D'Imperio, Mariapaola, Rossana Cavone, and Caterina Petrone. "Phonetic and Phonological Imitation of Intonation in Two Varieties of Italian." *Frontiers in Psychology* 5 (2014): 1226. *PMC*. Web. 22 Aug. 2018.

Dossey, E. E. (2012). Spontaneous phonetic imitation across regional dialects. (2012). *Linguistics Honors Projects*. 8. Retrieved from http://digitalcommons.macalester.edu/ling_honors/8.

Edlund, J., Heldner, M., & Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. In *Proceedings of Interspeech*. 2779–2782.

Evans, B., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America, 121*(6), 3814-3826. doi:10.1121/1.2722209

Fowler, C. A., Brown, J. M., Sabadini, L., and Weihing, J. (2003). Rapid access to speech gestures in perception: evidence from choice and simple response time tasks. *J. Mem. Lang*. 49, 396–413. doi: 10.1016/S0749-596X(03)00072-X

Giles, H. (1973). Accent mobility: a model and some data. Anthropological Linguistics 15:87–105.

Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in society*, *2*(2), 177-192.

Giles, H., and Coupland, N. (1991). *Language: Contexts and consequences.* Milton Keynes: Open University Press.

Giles, H., Coupland, N. and Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In *Contexts of Accommodation*, ed. 1–68.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105:251–279.

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review (2004) 11: 716*. <https://doi.org/10.3758/BF03196625>.

Gregory, S. W., & Hoyt, B. R. (1982). Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *Journal of Psycholinguistic Research*, *11*(1), 35-46.

Gregory, S. W., Webster S., & Huang F. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language and Communication* 13:195–217.

Heath, J. (2017). How automatic is convergence? Evidence from working memory. *Proceedings of the Linguistic Society of America*, *2*, 35-1.

Kaschak, M. P., Kutta, T. J., & Jones, J. L. (2011). Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic Bulletin & Review*, *18*(6), 1133-1139.

Khattab, G. (2013). Phonetic convergence and divergence strategies in English-Arabic bilingual children. *Linguistics*, *51*(2), 439-472.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 125-156.

Kim, M. (2012). *Phonetic accommodation after auditory exposure to native and nonnative speech.* (Doctoral dissertation, Northwestern University). Available from Dissertations & Theses @ CIC Institutions.

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168-173.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology, 32*(5), 790-804. <10.1037//0022-3514.32.5.790>.

Nguyen, N., Dufour, S., & Brunellière, A. (2012). Does imitation facilitate word recognition in a non-native regional accent? *Frontiers in Psychology*, *3*, 480.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*(2), 132-142. <doi:10.1016/j.wocn.2010.12.007>.

Nixon, J. S., Chen, Y. & Schiller, N. (2014). Multi-level processing of phonetic variants in speech production and visual word processing: evidence from Mandarin lexical tones. *Language, Cognition and Neuroscience*. 10.1080/23273798.2014.942326.

Oben, B., & Brône, G. (2016). Explaining interactive alignment: A multimodal and multifactorial account. *Journal of Pragmatics*, *104*, 32-51.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119: 2382–2393.

Pardo, J. S. (2010). Expressing oneself in conversational interaction. *Expressing oneself/expressing one's self: Communication, cognition, language, and identity*, 183-196.

Pardo, J. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, *4*, 559.

Pardo, J. S., Jay, I. C., Hoshino, R., Hasbun, S. M., Sowemimo-Coker, C., & Krauss, R. M. (2013). Influence of role-switching on phonetic convergence in conversation. *Discourse Processes, 50*(4), 276-300. <doi:10.1080/0163853X.2013.778168>.

Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., & Ward, M. (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics*, *69*, 1-11.

Peirce, JW (2007) PsychoPy - Psychophysics software in Python. J Neurosci Methods, 162(1-2):8-13.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. *Typological Studies in Language*, *45*, 137-158.

Pittam, Jeffery. *Voice in Social Interaction: An Interdisciplinary Approach.* Thousand Oaks: SAGE Publications, 1994.

Podlipský, V. J., & Šimáčková, Š. (2015) Phonetic Imitation is Not Conditioned by Preservation of Phonological Contrast but by Perceptual Salience. *International Congress of Phonetic Sciences (ICPhS) 2015.*

Reitter, D., Keller, F., & Moore, J. D. (2006). Computational modelling of structural priming in dialogue. In *Proceedings of the human language technology conference of the NAACL, companion volume: Short papers* (pp. 121-124). Association for Computational Linguistics.

Russell, N. K., & Stathopoulos, E. (1988). Lung volume changes in children and adults during speech production. *Journal of Speech, Language, and Hearing Research*, *31*(2), 146-155.

Sanker, C. (2015). Comparison of phonetic convergence in multiple measures. *Cornell Working Papers in Phonetics and Phonology,* pp. 60-75.

Schweitzer, A., & Walsh, M. (2016). Exemplar Dynamics in Phonetic Convergence of Speech Rate. *INTERSPEECH.*

Shockley K., Sabadini L., Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics* 66: 422–429.

Simonet, M. (2011). Intonational convergence in language contact: Utterance-final F0 contours in Catalan–Spanish early bilinguals. *Journal of the International Phonetic Association, 41*(2), 157-184. <doi:10.1017/S0025100311000120>.

Sonderegger, M. Bane, M. & Graff, P. (2017). The medium-term dynamics of accents on reality television*. Language, 93(3)*, 598-640. <doi:10.1353/lan.2017.0038>.

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research, 11*(2), 139-169. 10.1111/j.1468-2958.1984. tb00043.x

Xia, Z., Levitan, R., Hirschberg, J.B. (2014). Prosodic Entrainment in Mandarin and English: A Cross-Linguistic Comparison, Columbia University Academic Commons, <https://doi.org/10.7916/D8F47M84>.

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PloS one*, *8*(9), e74746.

Zemlin, W. R. (1998). *Speech and hearing science: anatomy and physiology.* 4th ed. Boston: Allyn and Bacon.

Appendix 1 – list of stimuli

text	Pinyin	English	text	Pinyin	English
八	ba1	eight	怕	pa4	afraid
爸	ba4	dad	皮	pi2	skin
笔	bi3	pen	坡	po1	slope
剥	bo1	peel	谱	pu3	(music) sheets
步	bu4	step	气	qi4	gas
测	ce4	measure	取	qu3	take
茶	cha2	tea	去	qu4	go
车	che1	car	惹	re3	bother
扯	che3	pull	热	re4	hot
吃	chi1	eat	洒	sa3	pour
出	chu1	exit	蛇	she2	snake
大	da4	big	事	shi4	thing
德	de2	virtue	书	shu1	book
帝	di4	emperor	塔	ta3	tower
读	du2	read	特	te4	extremely
法	fa3	law	踢	ti1	kick
父	fu4	father	兔	tu4	rabbit
歌	ge1	song	武	wu3	martial arts
葛	ge3	common surname	洗	xi3	wash
呵	he1	a short laugh	徐	xu2	slowly; common surname
寄	ji4	send	鸭	ya1	duck
句	ju4	sentence	衣	yi1	clothes
卡	ka3	card	姨	yi2	aunt
哭	ku1	cry	鱼	yu2	fish
苦	ku3	bitter	雨	yu3	rain
辣	la4	spicy	杂	za2	mess
力	li4	force (n.)	择	ze2	select
鹿	lu4	deer	紫	zi3	purple
爬	pa2	climb	租	zu1	rent

Appendix 2—participant questionnaire

Language background questionnaire for Mandarin speakers

Date: ________________

■■

I. Personal Data

(This information will be kept confidential)

Age: ________________ Sex: ________________

Education (your current or most recent educational level, even you have not finished the degree): (circle one)

elementary school middle school high school college (undergraduate)

college (masters or equivalent) college (PhD, JD, MD, or equivalent)

Others:________________

City & country of origin (e.g., Beijing, China): ______________________________________

City of current residence: ________________________________

****

II. Family History

Where are your parents/caregivers from? Please list for all members.

____

What languages do your parents/caregivers speak? Please list for all members.

____

****

III. Your Linguistic History

Please list all of the languages AND/OR dialects that you can speak *fluently*: (e.g., Mandarin, English, Cantonese, Nanjing dialect, etc.)

____

At what age did you first begin to learn standard Mandarin? ____________ (birth – current age)

What language(s) did your parents/caregivers use mostly when speaking to you?

Mandarin   Other dialect/language (please specify) _________   Mixed

What language(s) did you use mostly when speaking to your parents/caregivers?

Mandarin   Other dialect/language (please specify) _________   Mixed

Was there anyone else frequently living with you (i.e., siblings, grandparents, other caregivers, etc.)? If yes, what language(s) did you use when speaking with them, and what language(s) did they use when they were speaking with you? (skip if does not apply, and if multiple answers, please elaborate)

Mandarin   Other dialect/language (please specify) _________   Mixed

Did you attend any kindergarten / daycare before age 5? If yes, what language(s) were you spoken to there? (skip if no)

Mandarin   Other dialect/language (please specify) _________   Mixed

What language(s) did you most frequently use at grade school?

Mandarin   Other dialect/language (please specify) _________   Mixed

What language(s) did you most frequently use in college? (skip if doesn't apply)

Mandarin   Other dialect/language (please specify) _________   Mixed

What language(s) do you typically use with your closest friends?

Mandarin   Other dialect/language (please specify) _________   Mixed

If you have lived or travelled in *any part of China other than your origin for 1 year or longer*, please indicate the city or town, your length of stay, the language or dialect you most often used, and the frequency of your use of the language for each place.

City / Town	Length of stay[a] [month(s)]	Language / Dialect	Frequency of use[b]
			1 2 3 4 5 6 7
			1 2 3 4 5 6 7
			1 2 3 4 5 6 7

			1 2 3 4 5 6 7

a. You may have been to the city or town on multiple occasions, each for a different length of time. Add all the trips together.
b. Please rate according to the following scale (circle the number in the table)

Never	*Rarely*	*Sometimes*	*Regularly*	*Often*	*Usually*	*Always*
1	2	3	4	5	6	7

****

IV. Your linguistic proficiency

Please rate your current overall language ability in MANDARIN: _____________

1 = understand but cannot speak
2 = understand and can speak with great difficulty
3 = understand and speak but with some difficulty
4 = understand and speak comfortably, with little difficulty
5 = understand and speak fluently like a native speaker

Please rate your language learning skill. In other words, how good do you feel you are at learning new languages, relative to your friends or other people you know? (circle one)

Very poor	*Poor*	*Limited*	*Average*	*Good*	*Very Good*	*Excellent*
1	2	3	4	5	6	7