Phonetic Convergence in Shanghai-accented Mandarin: A Study of Podcast Conversations

Jiale Peng
Student Number: 15418766
Supervisor: Dr. Silke Hamann
Program: General Linguistics
MA Thesis, UvA
July 6 2026

Abstract

Phonetic convergence, the tendency for interlocutors' speech to become more similar during interaction, has been reported in many languages. However, evidence from Mandarin remains limited, particularly at the segmental level. The present study investigated consonantal convergence in natural Mandarin conversations, focusing on the alveolar-retroflex sibilant merger, a feature associated with the Shanghai accent. Two podcast conversations were analyzed, each involving the same Standard Mandarin-speaking host and a different Shanghai-accented guest. The center of gravity (CoG) of the frication portion of the sibilants was measured and compared across early and late phases of interaction. Two exploratory analyses further examined the host's solo speech as a baseline and a later phase of one conversation. Overall, the results provided little evidence for accommodation, and the observed changes were generally small and inconsistent across speakers. Nevertheless, the host showed a stronger alveolar-retroflex distinction in her solo speech than in conversation, and both guests showed a slight increase in their alveolar-retroflex contrast in the late phase, hinting at a possible tendency toward convergence.

1. Introduction

Communication plays a central role in everyday life, through which people exchange information, build social relationships, and express self-identity. Importantly, communication is not a static process. In face-to-face communication, participants may accommodate their behaviors in response to their interlocutor's characteristics and the communicative context (Barón-Birchenall, 2023). Such behavioral accommodation can happen in multiple channels, including language, facial expression, gesture, and other behaviors (Louwerse et al., 2012).

Among these forms of behavioral accommodation, linguistic accommodation has received particular attention, given the central role of language in human communication. Previous research has shown that speakers may adjust various aspects of their language use during interaction, including phonetic features (detailed later), lexical choices (e.g., Brennan & Clark, 1996; Ward & Litman, 2007; Reverdy et al., 2020), syntactic structures (e.g., Branigan et al., 2000; Xu & Reitter, 2016), and broader linguistic or discourse styles (e.g., Thomson et al., 2001; Niederhoffer & Pennebaker, 2002). Among them, phonetic features constitute the most basic aspects of spoken language in face-to-face interaction. This thesis focuses on accommodation at the phonetic level, specifically examining consonantal convergence in natural Mandarin conversations. The following section first reviews the literature on phonetic accommodation and phonetic convergence, and then discusses previous findings on Mandarin.

1.1 Phonetic Accommodation and Phonetic Convergence

Based on Pardo (2006) and Barón-Birchenall (2023), phonetic accommodation concerns speakers' modification of speech acoustics, including both segmental and suprasegmental features. Such accommodation may result in different patterns of phonetic similarity between interlocutors. Within Communication Accommodation Theory (CAT; Giles et al., 1991), these patterns are described as convergence, divergence, and maintenance. In CAT, these patterns are typically interpreted from a sociolinguistic perspective, whereby speakers modify their speech to achieve specific communicative goals. Another influential account, the Interactive Alignment Model (IAM; Pickering & Garrod, 2004), explains convergence as the outcome of largely automatic alignment processes during interaction.

The present study adopts the terms of convergence, divergence, and maintenance but does not assume any particular social or cognitive explanation for these patterns. Instead, following Barón-Birchenall (2023), convergence is described as a symmetric or asymmetric increase in phonetic similarity between interacting individuals. Correspondingly, divergence refers to a decrease in phonetic similarity, whereas maintenance refers to a relatively stable degree of similarity despite any changes. Additionally, the terminology used in this area remains inconsistent, and related terms such as imitation, alignment, entrainment, coordination have also been used to describe similar processes of increasing speech similarity during interaction. In the present study, the terms phonetic accommodation and phonetic convergence were both adopted from Barón-Birchenall (2023).

1.2 Phonetic Convergence in Different Settings

Although convergence, divergence, and maintenance are all possible consequences of phonetic accommodation, more evidence of convergence has been reported in previous research. Among them, much has been conducted in laboratory settings. Early studies often employed speech shadowing paradigms, in which participants repeated words or utterances produced by a model speaker. For example, Goldinger (1998) found that in an immediate shadowing task, shadowers spontaneously imitated the acoustic patterns of the input stimuli. More recent studies have increasingly examined the phenomenon in more interactional contexts. For instance, many studies have examined convergence in task-oriented conversations, including map tasks (Pardo, 2006; Pardo et al., 2010), picture description tasks (Kim et al., 2011), and picture-ordering tasks (Xia et al., 2023). However, as speakers are aware that they are subjects of an experiment, even when the true purpose of the study is not disclosed to them, the speech produced in these contexts may still not fully reflect natural spontaneous interaction.

Some researchers have therefore turned to naturalistic data. One pioneering study, Gregory and Webster (1996), analyzed 25 interviews from an American talk show, the length of which ranged from 20 to 30 min. They measured fundamental frequency (f_0) across three stages of each interview and reported that interview partners converged to each other in the course of the interview. Another notable example, Abel (2012), examined phonetic convergence in 12 interviews from a Canadian English podcast, with a mean length of 20 min. The study compared the first and second halves of each interview to examine whether the host's speech became more similar to that of each guest over the course of the interaction. Two variables were analyzed: the Canadian Raising diphthong [aɪ] and the vowels involved in the *cot-caught* merger. However, the results did not show systematic convergence: changes in the *cot-caught* merger were better explained by phonological environment than by convergence, while the Canadian Raising showed divergence more often than convergence, which the author attributed to the host's maintenance of Canadian identity. It should be noted that this study examined only the host's speech rather than the speech of both interlocutors. The most recent research, Kania and Roselani (2025), also investigated phonetic convergence in six episodes of Indonesian English podcasts. However, the authors only made a categorical distinction and did not investigate how convergence developed over the course of interaction. Overall, phonetic convergence in naturalistic speech remains under investigated.

To date, evidence of phonetic convergence in interactional settings has been widely reported in a number of languages, including English (e.g., Pardo, 2006; Kim et al., 2011), French (Delvaux & Soquet, 2007), German (Schweitzer & Lewandowski, 2014), and Dutch (Wieling et al., 2020).

1.3 Phonetic Convergence across Acoustic Features

When convergence occurs, it can involve both segmental and suprasegmental acoustic features. Suprasegmental convergence has been investigated in speaking rate, pitch, loudness, and other prosodic features. In general, the overall similarity between paired speakers has

been consistently reported at both the conversation level and the turn-by-turn level (e.g., Levitan & Hirschberg, 2011; Levitan et al., 2012, 2015; Weise et al., 2019; Xia et al., 2023). However, results regarding convergence over the course of interaction are mixed. Levitan & Hirschberg (2011) reported significant convergence in intensity, shimmer, and noise-to-harmonics ratio when comparing the first and second halves of interactions in American English. In contrast, Xia et al. (2023) found no significant convergence at either the conversation-level or the turn-by-turn level in Mandarin. Notably, although Weise et al. (2019) found both convergence and divergence in specific suprasegmental features in English conversations, they pointed to the considerable variation existed between individuals and suggested that synchrony was somewhat more prevalent.

Regarding segmental features, for vowels, reported results have also been mixed. As summarized by Pardo et al. (2013), Pardo (2010) found convergence on /i/ and /u/ but divergence on /æ/ and /a/ in English, while Babel (2012) found the opposite, and Pardo et al. (2010) found neither. Compared to vowels, consonants have received less attention, and much of the existing work has relied on shadowing tasks. Early studies on voice onset time (VOT) found that exposure to extended VOTs led participants to produce longer VOTs (Shockley et al., 2004; Nielsen, 2011). More recently, Wagner et al. (2021) examined whether native Dutch speakers converged to a non-native model speaker on [t] and [s] in a repetition task, combining acoustic analysis and perceptual judgements. For [t], significant convergence was found for stop closure duration but not for stop aspiration duration; for [s], significant convergence was found for center of gravity (CoG) but significant divergence was found for fricative duration. Additionally, these acoustic changes did not contribute equally to the perceptual judgement of similarity.

Beyond imitation tasks, few studies have investigated consonantal convergence in interaction. Cao (2019), using a map task, investigated whether non-native English speakers from Hong Kong converged to native English speakers on /z/ and /θ/ across pre-interaction, interaction, and post-interaction phases, reporting significant convergence for /z/ but not /θ/. However, convergence was defined as an increase in correct phoneme realizations rather than as a change in acoustic properties. Notably, the most recent study, Kania and Roselani (2025), investigated consonantal convergence on /t/ and /ɹ/ in Indonesian English podcast conversations. However, as mentioned earlier in Section 1.2, the study adopted a similar categorical approach and only reported a positive correlation between the two speakers' frequencies of incorrect phoneme realizations, without tracking how they developed over the course of the interaction.

Taken together, these findings suggest that convergence is not always guaranteed and can be selective in different features, and not all acoustic changes equally affect the perception of similarity. As for a gap in the research field, no study seems to have examined consonantal convergence in terms of acoustic change over the course of natural conversations, which the present study seeks to address.

1.4 Phonetic Convergence in Accent and Dialect Context

Phonetic convergence can also occur in interaction between speakers of different dialects or accents. More broadly, conversation-level phonetic convergence has been argued to contribute to population-level accent change and dialect formation (Pardo, 2006), which has motivated much of previous research to take a longitudinal approach. For instance, Evans and Iverson (2007) examined accent change among university students over a two-year period.

Fewer studies have focused on the immediate effects of interacting with a speaker of a different accent or dialect. Among existing work, some studies focused on native speakers' interaction with non-native speakers (e.g., Kim et al., 2011; Lewandowski & Nygaard, 2018; Wagner et al., 2021), while other studies investigated native speakers of the same language but with different accents or dialects, such as Liège French vs. Brussels French (Delvaux & Soquet, 2007) and Southern American English vs. non-Southern American English (Clopper & Dossey, 2020). However, short-term accent convergence in Mandarin has not been investigated, though the language exhibits rich regional variation and can serve as a promising context to examine this phenomenon. The present study therefore aims to fill this gap.

1.5 Factors Influencing Phonetic Convergence

Although phonetic convergence is often subtle and varies across speakers and settings in phonetic research (Pardo et al., 2018), its degree and direction has been argued to be influenced by a range of linguistic and social factors.

With respect to linguistic factors, acoustic distance between interlocutors has been one focus of discussion. Some researchers argue that greater acoustic distance may lead to greater convergence, as it provides more phonetic space for convergence to happen (e.g., Babel, 2012; Walker & Campbell-Kibler, 2015) and is more salient to interlocutors (Cao, 2019; Paquette-Smith, 2022), provided that the target still falls within the speaker's phonetic repertoire (Babel, 2009). In contrast, Kim et al. (2011) found that speakers sharing the same dialect of the same language converge more than speakers sharing different dialects or languages, suggesting that closer language distance facilitates convergence. Tobin (2022) offers a reconciling account, proposing that there may be an optimal acoustic distance that maximizes convergence, neither too similar nor too different. In addition to language distance, language proficiency and familiarity has also been argued to play a part. Ulbrich (2024) found that among Spanish learners of German, high-proficiency speakers converged more to native German speakers than low-proficiency speakers did, indicating that a sufficient level of linguistic competence is needed to perceive and accommodate to another speaker's features. Ross et al. (2021) also found that the familiarity with a dialect increased the magnitude of convergence.

Social factors have also been shown to participate in the process and often interact with linguistic factors, which complicates the picture. Regarding social status, lower-status speakers tend to converge more toward those who are believed to have higher status, while higher-status speakers may converge less or diverge, as a way of maintaining their identity

(e.g., Gregory & Webster, 1996; Pardo, 2006; Abel, 2012; Muir et al., 2017; Dippong, 2020). The role of gender has also been investigated, although mixed results have been reported. Namy et al. (2002) reported that female speakers converged more than male speakers, while Pardo (2006) reported the opposite. A more recent larger-scale study by Pardo et al. (2016), examined 92 speakers in same-sex and opposite-sex pairs and reported no reliable main effect of sex pairing. Beyond social status and gender, social biases toward certain groups or their associated linguistic features have also been shown to hinder convergence (e.g., Babel, 2010; Clopper & Dossey, 2020). However, some researchers also questioned how powerful social factors could be. For instance, Lewandowski and Nygaard (2018) found that native American English speakers converged more toward Spanish-accented speakers than toward native speakers, though native speakers hold a relatively negative attitude towards Spanish accent. This suggests that the influence of social factors on convergence may not be overriding.

To conclude, beyond various research on phonetic convergence, it is still a complicated phenomenon and needs to be further unfolded, especially in natural interactional settings with numerous factors operating together. Nevertheless, natural conversations are still an attractive and valuable context to explore how phonetic accommodation and convergence works in real interaction, which motivated the present study.

1.6 Standard Mandarin and Shanghai-accented Mandarin

The language of interest in the present study is Mandarin. Before proceeding, a few key terms should be clarified. Standard Mandarin is the standardized variety of Chinese used officially in China. When spoken in different regions, it interacts with local dialects and acquires their linguistic features, giving rise to regional varieties. The phonetic aspect of such regional variation is commonly referred to as accent. Shanghainese, or Shanghai dialect, is a regional variety of the Wu dialect group spoken in the Yangtze River Delta in southeastern China. Shanghai-accented Mandarin (hereafter SHM) is the regional variety that results from the interaction of Shanghai dialect and Standard Mandarin.

Given the crucial role of Shanghai, a number of studies have investigated SHM. Among them, two consonantal features have been consistently documented. First, SHM speakers somewhat merge the alveolar sibilants [ts, tsʰ, s] with their retroflex counterparts [ʈʂ, ʈʂʰ, ʂ], as retroflex sibilants are absent in Shanghainese. Yao (1988) described two specific patterns that may occur in SHM speakers' realization of these sounds. One is an incomplete merger, in which the produced sounds fall between alveolar and retroflex. The other is phoneme substitution. Because SHM speakers may have to memorize the pronunciation patterns of each words involving these sounds, they may sometimes overcorrect and produce a wrong phoneme. Importantly, such substitution is asymmetric: retroflex sounds are more often mispronounced as alveolar than alveolar sounds are mispronounced as retroflex (Li & Wang, 2003). Second, SHM speakers tend to merge the nasal codas [n] and [ŋ], as they are not distinguished in Shanghainese. According to Ye (2012), this merging tendency is particularly apparent before the vowels [i] and [ə], and the tongue position tends to be more fronted, approximating [n].

The vowel system of SHM has also been examined systematically (e.g., Li & Wang, 2003; Yu, 2004; Yu et al., 2008), but the reported findings are mixed. Li and Wang (2003) reported that, compared to Beijing Mandarin, [i] was more fronted and close in SHM while [u] showed no difference. Yu (2014), however, found little difference in [i] once gender was controlled for, and reported difference in [u] instead. More generally, Yu (2014) argued that SHM has a more dispersed vowel space, rather than showing large shifts in individual vowels.

Taken together, two consonantal features distinguish SHM from Standard Mandarin: a reduced contrast between alveolar and retroflex sibilants, and a reduced contrast between the nasal codas [n] and [ŋ]. Since nasal codas may be difficult to detect in natural fast speech, and vowel features have been inconsistently documented, the present study therefore focuses on the alveolar-retroflex contrast as an acoustic cue of Shanghai accent.

1.7 Phonetic Convergence in Mandarin

Compared to English and other European languages, phonetic convergence in Mandarin has received far less attention, and research on this topic began relatively late. The earliest study, Xia et al. (2014), examined task-oriented Mandarin conversations lasting approximately six minutes on average. The authors divided a conversation into two halves and compared three prosodic features (speech rate, intensity, and pitch) between the two halves. Results showed that, although speakers were more similar to their partner than to a non-partner of the same gender and conversational role in intensity and pitch, no convergence was found over the course of interaction. Notably, speakers did show local convergence at the turn level on intensity and f_0 features, but this did not lead to global convergence at the conversation level. Using the same corpus, Xia et al. (2023) and Xia (2024) reported similar findings. It should be noted that the available literature has been based on a single conversational corpus, which has been analyzed by a small group of researchers repeatedly. As a result, little is known about whether the reported patterns generalize to other Mandarin conversational settings.

Notably, two more recent studies, Sun and Ding (2023, 2024), investigated this phenomenon in a Mandarin talk show in which a female host interacted with different guests, also measuring speech rate, intensity, and pitch. Compared to the studies by Xia and colleagues, these conversations were more natural and substantially longer, with an average duration of 22 minutes. However, both studies focused primarily on convergence at the turn-by-turn level rather than convergence over the course of the conversation. At the turn level, interlocutors showed more divergence than convergence across all three measures. Additionally, the host showed more convergence toward guests than the reverse, which the authors interpreted as a conversational strategy to put guests at ease.

In summary, no evidence of convergence over the course of interaction has been reported in Mandarin for suprasegmental features. Moreover, no study has examined segmental convergence in Mandarin, and no study has focused on speakers with different regional accents, despite the rich regional variation Mandarin exhibits.

1.8 Research Questions and Hypotheses

To address these gaps, the present study investigates phonetic convergence in spontaneous Mandarin podcast conversations. Two conversations are analyzed in which a female host, who speaks Mandarin without an obvious regional accent, interacts separately with two male guests who speak Mandarin with a Shanghai accent. The analysis focuses on the alveolar sibilants /s/ and /ts/ (only the frication portion), and the retroflex sibilant /ʂ/. By examining the acoustic realization of these segments in the early and late phases of each conversation, the study explores whether two interlocutors become more similar on the production of alveolar and retroflex sibilants as the conversation progresses.

Specifically, the study aims to answer two research questions and test two corresponding hypotheses in each conversation:

> *(RQ1): Do two interlocutors become more similar as the conversation goes on?*
> *(H1): The two interlocutors are expected to exhibit phonetic convergence over the course of interaction, such that their realization of /s, ts/ and /ʂ/ will be more similar in the late phase than in the early phase of the conversation.*
>
> *(RQ2): If convergence occurs, is it symmetric or asymmetric between the two interlocutors?*
> *(H2): Convergence is expected to be asymmetric, which means the host would contribute more to the convergence. This prediction is based on two reasons. First, from a linguistic perspective, the guests' accent is more salient than that of the host, making the host more likely to adjust her speech than the guests. Second, from a social perspective, the host is expected to adapt more, as her role is to engage with the guests and keep the conversation going smoothly; the guests' higher social prestige may further encourage this.*

In addition, the study also conducts two exploratory analyses:

> (EA1) A recording of the host's solo speech is analysed and compared with her speech in the two conversations, which provides a rough baseline for the host's speech without interaction;
>
> (EA2) As Conversation 2 is nearly twice as long as Conversation 1, an additional phase is extracted from the end of the conversation, in addition to the early and late phases included in the main analysis. The main purpose is to examine whether convergence, if found any, still continues to develop over a longer interaction.

The remainder of this thesis is organized as follows. Section 2 describes the dataset and methodology. Section 3 presents the results, which are discussed in Section 4. Finally, Section 5 concludes the thesis.

2. Methods

2.1 Corpus

The data in this study were drawn from a personal podcast program hosted by a young female writer, Fangzhou Jiang. In the program, Jiang shares her thoughts on literature and daily life, either in conversation with invited guests or in solo talk episodes. The program provides a suitable source of relatively natural speech, as the conversations are spontaneous rather than elicited in laboratory settings.

Two conversation episodes from this program were selected for analysis. Both were released in the same year and centered on similar literary topics, which ensures comparability. Conversation 1 is between Jiang and Zidong Xu (Guest 1), while Conversation 2 is between Jiang and Danqing Chen (Guest 2). Information of the three speakers will be detailed in 2.2. In addition to the two conversations, one solo talk episode was also included to establish a rough baseline for the host's speech without interaction. Although it was not recorded exactly before the two conversations, this episode was from the same program and released in the same year, and was likely to have been produced with similar equipment in a similar environment, making it a comparable baseline. Additionally, as solo talk on literary topics is more likely to induce reading from a prepared script, an episode on an everyday topic was selected instead, which was more likely to reflect natural spontaneous speech.

As podcast episodes are edited recordings, the actual duration of each interaction was unknown and may have been longer than the released version. Nevertheless, the sequential order of the content was assumed to have been preserved. Detailed information of the selected recordings can be seen below in Table 1.

Speech Setting	Speakers	Duration	Content	Release Date	Link
Conversation 1	Jiang & Xu	1h 41 min	Chinese literature, focusing on a famous Chinese writer Ailin Zhang	Nov.14, 2025	https://www.xiaoyuzhoufm.com/episode/6915b499cbba038b4261fd8c
Conversation 2	Jiang & Chen	3h 9 min	Literature, focusing on Tolstoy	Sep.17, 2025	https://www.xiaoyuzhoufm.com/episode/6878c298a12f9ff06aa2b51a
Host's solo talk	Jiang	50 min	Daily topic, talking about how to overcome procrastination	Apr.10, 2025	https://www.xiaoyuzhoufm.com/episode/67f79bc159699d74dcd9025d

Table 1. Speech information

2.2 Speakers and Their Accents

The three speakers differ in age and accent background. While the two guests' age is close to each other, the host is much younger than them. As a general trend in China, younger generations tend to speak Mandarin in a more standard way, as they have grown up during a period of intensive promotion of Standard Mandarin. This appears to also be the case for the host. As confirmed by the author[1] and three language consultants[2], her speech showed no obvious regional accent and was therefore considered as Standard Mandarin.

For the two guests, native Shanghainese speakers helped verify the two guests' accents, as the author is not a native of Shanghai. For Guest 2, a clear Shanghai accent was confirmed by five consultants. This judgement was further supported by his occasional use of Shanghainese expressions during the conversation. For Guest 1, the verification process was more complex. As he has lived and worked in Hong Kong for several years and speaks Cantonese as well, three additional consultants were added, including one Cantonese speaker who confirmed that no traces of Cantonese accent were present in his speech. Additionally, some consultants were uncertain whether Guest 1 could be specifically identified as a speaker from Shanghai or a neighboring city, but they agreed that his accent at least belonged to the Yangtze River Delta region. Taken together with his Shanghai upbringing, Guest 1's accent was therefore treated as a Shanghai accent in the present study. It should be noted that, some consultants reported that the alveolar-retroflex merger was one of the main cues they used to identify the accent. Since this feature is shared by Mandarin speakers from Shanghai and other parts of the Yangtze River Delta region, it may also explain the uncertainty about Guest 1's specific origin. As the present study focuses on this shared feature, it did not affect the analysis. Detailed speaker information can be seen below in Table 2.

Role	Name	Occupation	Born and Raised in	Accent	Gender	Age in Podcast
Host	Jiang	Writer	Hubei Province (middle region of China)	None	Female	36
Guest 1	Xu	Scholar in Chinese literature	Shanghai (southeastern region)	Shanghai	Male	71
Guest 2	Chen	Painter, writer	Shanghai (southeastern region)	Shanghai	Male	72

Table 2. Speaker information

2.3 Target Variable and Acoustic Measurement

Targeting the alveolar-retroflex contrast in Mandarin, the study initially selected the alveolar fricative /s/ and its retroflex counterpart /ʂ/. Due to the limited number of /s/ tokens in the data, however, the alveolar affricate /ts/ was also included in the alveolar category. These two

[1] The author is a native speaker of Mandarin, from the northwestern region of China.

[2] All three language consultants are native speakers of Mandarin, one from the same province of the host, one from the north region, and one from Shanghai.

segments share the same place of articulation and differ only in manner of articulation, with the affricate /ts/ consisting of a stop closure followed by frication (Ashby, 2013). Therefore, only the frication part of /ts/ is taken for the acoustic analysis. The alveolar category then comprises /s/ and /ts/, and the retroflex category comprises /ʂ/.

Among the available acoustic measures, center of gravity (CoG) has been widely used to distinguish alveolar and retroflex sibilants (e.g., Hamann & Avelino, 2007; Wagner et al., 2021). Typically, alveolar sibilants exhibit higher CoG values because they involve a shorter articulatory tract (Gordon et al., 2002). Accordingly, CoG was selected as the acoustic measure in the present study. In addition, as lip-rounding may lower the CoG of sibilants (Smith et al., 2019), the rounding status of the following vowel was also considered in token selection and statistical analysis.

2.4 Token Selection

Since /ʂ/ tokens occur much more frequently than /s, ts/ tokens in natural speech, approximately a duration of 15 min was sufficient to collect enough alveolar tokens for analysis. Different portions of the recordings were extracted as follows:

(A) For the main analysis:
As Conversation 1 lasted 101 min and Conversation 2 lasted 190 min, only the first 101 min of Conversation 2 were used for comparison. From each conversation, approximately the first and last 15 min were extracted to represent the early and late phases of interaction. It should be noted that, in Conversation 2, the first 5 min consisted of the host's introduction of the guest, which was recorded separately from the conversation and was therefore excluded; the interaction began from around the 5-min point. As the portion near the 101-min point of Conversation 2 did not yield enough tokens, a slightly earlier portion was selected to represent the late phase.

(B) For the two exploratory analyses:
For the first exploratory analysis, a stable period of approximately 15 min from the middle of the host's solo talk was extracted to provide a rough baseline for her speech without interaction.

For the second exploratory analysis, approximately the last 15 min of Conversation 2 was extracted and marked as the extended phase, allowing for an examination of whether any convergence pattern continues to develop over a longer duration.

Due to differences in speaker turn-taking, token distribution, and background music interruptions, the exact durations of the selected phases were not identical across recordings. This was not considered problematic in the study, as the selected portions were intended to represent relative stages of interaction in edited recordings, rather than strictly time-matched windows. Detailed information of each phase can be seen in Table 3.

Speech Setting	Phase of Interaction	Window
Conversation 1 (Duration: 01:41:06)	**Early**	00:00:00 - 00:16:01 (16min)
	Late	01:23:58 - 01:39:31 (15.5min)
Conversation 2 (Duration: 03:09:48)	**Early**	00:04:57 - 00:23:37 (18min)
	Late	01:14:52 - 01:30:12 (15.5min)
	Extended	02:50:36 - 03:05:53 (15min)
Host's solo talk (Duration: 00:50:18)	/	00:23:40 - 00:38:17min (15.5min)

Table 3. Phase information (phases for the main analysis in bold)

Within each phase, the selected portion was divided into three consecutive sections to ensure that the more frequently occurring /ʂ/ tokens are distributed across the full phase rather than being concentrated in the first few minutes. Within each section, the first eight usable tokens per category per speaker were selected, with a target balance of four tokens followed by rounded vowels /o, u/ and four tokens followed by unrounded vowels /a, i, ɤ, ə/. In very few cases, fewer than eight tokens were available, therefore all usable tokens were included. Tokens were excluded if they involved laughter, background music, overlap with another speaker, or were produced in Shanghainese or English. Tokens produced at a fast speech rate were retained if the frication remained perceptually identifiable and their fricative characteristics were visible in the spectrogram; tokens that meet neither criterion were excluded. At last, a total of 517 tokens were collected. Detailed information can be seen below in Table 4.

Setting	Phase of Interaction	Speaker	Number of Tokens				Total
			Alveolar		Retroflex		
			Rounded	Unrounded	Rounded	Unrounded	
Conversation 1	**Early**	Host	12	12	12	12	48
		Guest 1	11	12	12	12	47
	Late	Host	12	12	10	12	46
		Guest 1	12	11	11	12	46
Conversation 2	**Early**	Host	7	12	12	12	43
		Guest 2	12	12	12	12	48
	Late	Host	11	12	12	12	47
		Guest 2	12	12	12	12	48
	Extended	Host	12	12	12	12	48

Setting	Phase of Interaction	Speaker	Number of Tokens				Total
			Alveolar		Retroflex		
			Rounded	Unrounded	Rounded	Unrounded	
	Extended	Guest 2	12	12	12	12	48
Host's solo talk	/	Host	12	12	12	12	48
Total	/		125	131	129	132	517

Table 4. Token information (phases for the main analysis in bold)

2.5 Segmentation

All tokens were segmented and manually annotated in *Praat* (Boersma & Weenink, 2025). Only the frication portion of each target segment was included. For affricates, the stop release was excluded. In addition, as natural speech is often fast and produces relatively short frication, segment boundaries were placed to avoid transitional portions with adjacent segments, using voicing for exclusion. By doing so, the length of the stable frication interval available for analysis was improved, which helped yield a more stable measure. An annotation example is provided in Figure 1.

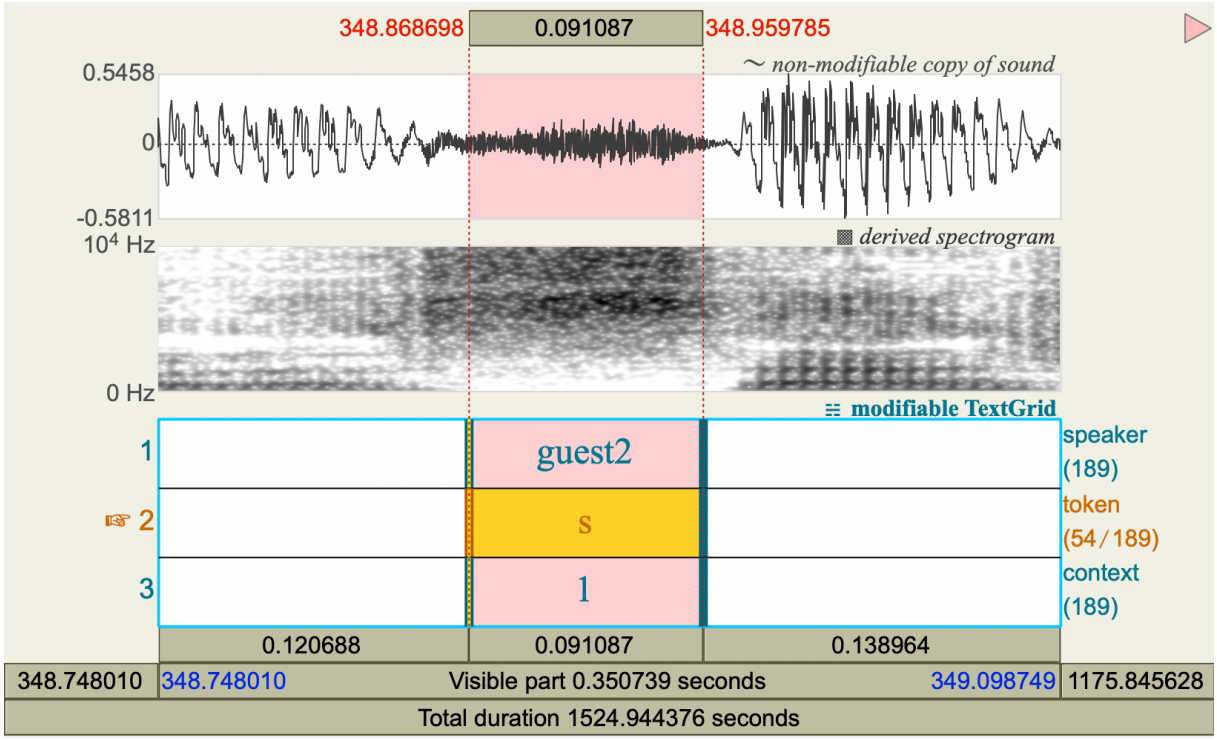


Figure 1. Annotation example ("1" in the context tier means "unrounded")

After annotation, CoG measurements were extracted automatically using a script in *Praat* (see Appendix 1). For each segmented token, the script generated a spectrum from the labelled interval and calculated its CoG. Around 15 outliers were identified and manually checked in *Praat*. Among them, two were found to be data-entry errors and one was found to be a labelling error, all of which were corrected accordingly. Two extreme outliers were removed following visual inspection of the data distribution (see Appendix 2 for distribution involving these two tokens). The remaining outliers were retained as they reflected natural variation in speech. In the end, a total of 515 tokens were used for statistical analysis.

2.6 Statistical Analysis

All statistical analyses were conducted in *R* (version 4.5.2; R Core Team, 2025). The acoustic measure analysed was centre of gravity (CoG).

(A) For the main analysis:
The main analysis aimed to answer the two research questions. For each conversation, a linear model was fitted with *CoG* as the dependent variable. The model included *Phase* (Early vs. Late), *Speaker* (Host vs. Guest), *Place of articulation* (Alveolar vs. Retroflex), and *Vowel context* (Rounded vs. Unrounded) as fixed effects, together with all interaction terms:

*a. CoG ~ Phase * Speaker * Place * Context*

The four-way interaction *Phase * Speaker * Place * Context* answered RQ1: whether the relationship between *Speaker*, *Place*, and *Context* differed between the two phases.

However, as the coefficients of a four-way interaction were too complex to interpret, the linear model was used only to determine whether accommodation occurred. If accommodation were found, its direction (either convergence or divergence) was examined by calculating the distance between the two interlocutors' mean CoG values and comparing this distance under each condition between the early and late phases. A decrease in distance was interpreted as convergence, and an increase as divergence. This was operationalized as follows:

DistanceEarly = MeanCoGHost − MeanCoGGuest
DistanceLate = MeanCoGHost − MeanCoGGuest
DistanceLate − DistanceEarly = ?

To address RQ2 (whether the convergence is symmetric or asymmetric between the two interlocutors), the magnitude of change of each speaker under each condition was calculated and compared to each other. This was operationalized as follows:

ChangeHost = MeanCoGLate − MeanCoGEarly
ChangeGuest = MeanCoGLate − MeanCoGEarly
ChangeHost vs. ChangeGuest ?

It should be noted that *Distance* and *Change* were calculated and compared separately for each *Place* * *Context* condition: alveolar-rounded, alveolar-unrounded, retroflex-rounded, and retroflex-unrounded.

(B) For the two exploratory analyses:

(1) Using the host's speech in the early phases of the two conversations and in the solo talk, this exploratory analysis examined whether interaction affected her speech. A linear model was fitted with *CoG* as the dependent variable. The model included *Speech Setting* (Conversation 1 vs. Conversation 2 vs. Solo), *Place of articulation* (Alveolar vs. Retroflex), and *Vowel context* (Rounded vs. Unrounded) as fixed effects, together with all interaction terms:

*b. CoG ~ Speech * Place * Context*

The three-way interaction *Speech * Place * Context* answered the question: whether the relationship between *Place* and *Context* differed among the three speech settings.

(2) Using all three phases of Conversation 2, this exploratory analysis examined whether the possible convergence continued to increase. A linear model was fitted with *CoG* as the dependent variable. The model included *Phase* (Early vs. Late vs. Extended), *Speaker* (Host vs. Guest), *Place of articulation* (Alveolar vs. Retroflex), and *Vowel context* (Rounded vs. Unrounded) as fixed effects, together with all interaction terms:

*c. CoG ~ Phase * Speaker * Place * Context*

The four-way interaction *Phase * Speaker * Place * Context* answered the question: whether the relationship between *Speaker*, *Place*, and *Context* differed among the three phases.

3. Results

The results are presented in four parts: (1) overview of the dataset; (2) a sanity check of the three speakers' alveolar and retroflex productions as well as their alveolar-retroflex contrast; (3) results of the main analysis; (4) results of the two exploratory analyses.

3.1 Overview of Dataset

First of all, the distribution of CoG values for all extracted tokens is shown in Figure 2, split by *Speaker* (Guest1, Guest2, and Host) and *Vowel Context* (Rounded vs. Unrounded). Early and late phases include data from both conversations, the extended phase includes Conversation 2 only, and solo contains data from the host's solo talk.

Overall, alveolar tokens showed higher CoG values and exhibited greater variation than retroflex tokens. Considerable between-speaker variation was also observed. Guest 2's

productions were the most stable, with tokens closely clustered together. Guest 1's alveolar and retroflex productions were relatively similar to each other, and his retroflex tokens showed more variation compared to those of Guest 2 and the host. The host showed the most variation among the three speakers, particularly in her alveolar tokens.

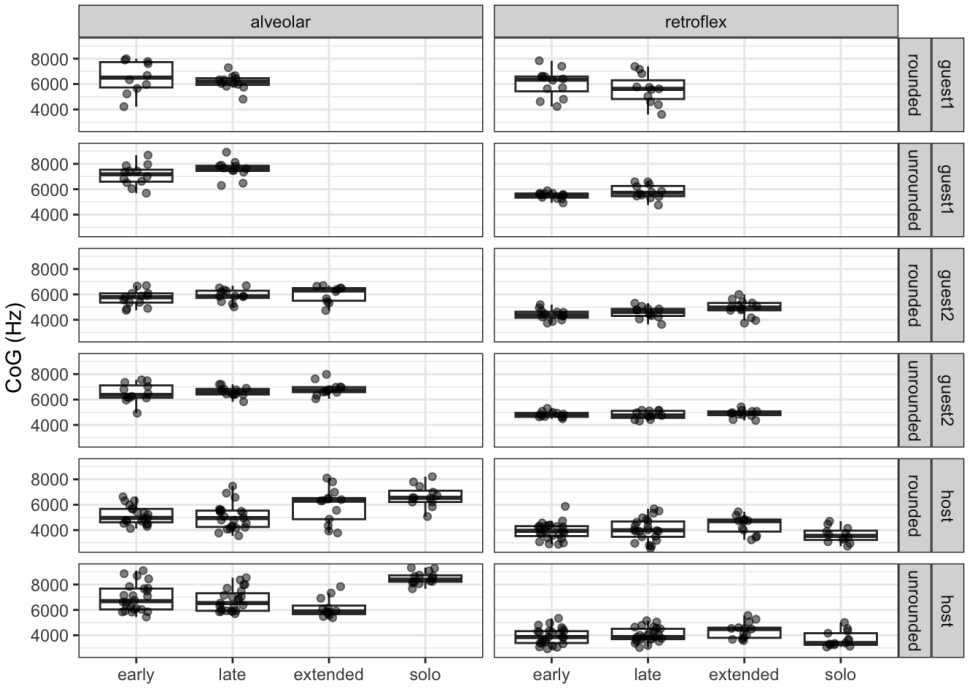


Figure 2. CoG distribution
(Boxes represent the interquartile range (IQR), horizontal lines inside indicate medians, whiskers extend to 1.5 × IQR, and points represent individual tokens.)

3.2 Sanity Check

Before examining accommodation, the speech of the three speakers was compared using data from the early phase of each conversation, which may reflect their speech with relatively less interaction effects. The two Shanghai speakers' mean CoG values and their alveolar-retroflex contrast are shown in Figures 3a and 4a. Since the host participated in both conversations, her speech from both was included, with mean CoG and contrast shown in Figures 3b and 4b.

As mentioned in Section 3.1, here we can see more clearly that alveolar tokens overall exhibited higher CoG values than retroflex tokens. Particularly, when followed by rounded vowels, the sibilants showed lower CoG values, which is consistent with previous literature (e.g., Gordon et al., 2002).

Between the two Shanghai guests, Guest 1 exhibited higher CoG values than Guest 2 overall, suggesting a systematic difference between the two guests. Regarding the alveolar-retroflex contrast (alveolar minus retroflex), the two guests showed a similar value in the unrounded context. In the rounded context, however, Guest 1 had a much smaller contrast, approximately one third of that of Guest 2, suggesting that the degree of merger varied between the two

Shanghai guests in this context. For the host, her speech did not differ much between the two conversations. Compared to the two guests, her contrast in the rounded context was comparable to that of Guest 2, while in the unrounded context it was almost twice that of both guests. This suggests that her alveolar-retroflex distinction was stronger than that of the two Shanghai guests, at least in the unrounded context.

In summary, compared to the host, the two Shanghai guests showed a smaller alveolar-retroflex contrast as expected for a Shanghai accent, particularly when sibilants were followed by unrounded vowels. The degree of merger also varied between the two guests, especially in the rounded context where Guest 1 showed a much smaller contrast.

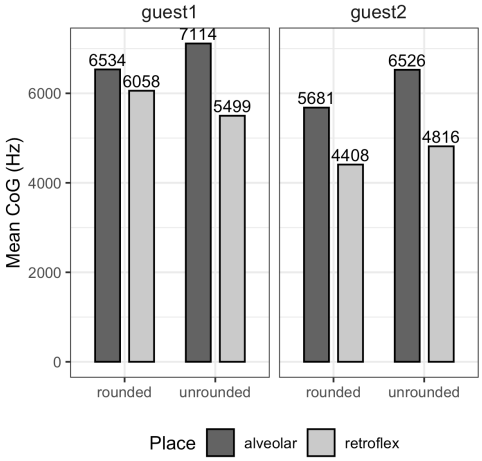


Figure 3a. Initial mean CoG (guests)

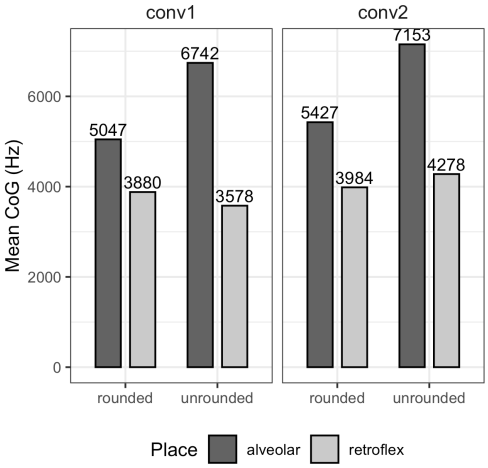


Figure 3b. Initial mean CoG (host)

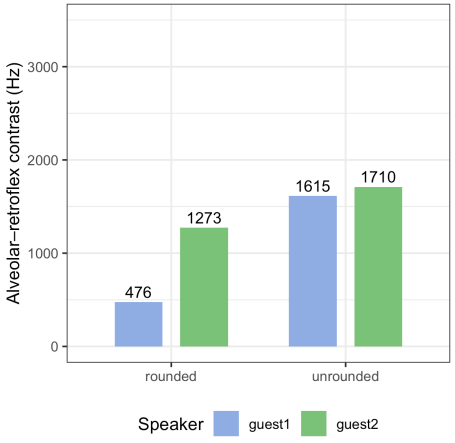


Figure 4a. Initial alveolar-retroflex contrast (guests)

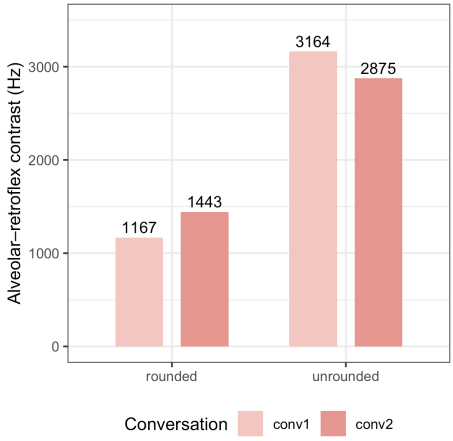


Figure 4b. Initial alveolar-retroflex contrast (host)

3.3 Results of the Main Analysis

3.3.1 Conversation 1: Host & Guest 1

*a. CoG ~ Phase (Early vs. Late) * Speaker (Host vs. Guest) * Place (Alveolar vs. Retroflex) * Context (Rounded vs. Unrounded)*

The linear model *a* was run for Conversation 1, and the significant and marginally significant effects are listed in Table 5 (see Appendix 3 for complete results). The main effect of *Speaker* was significant, with Guest 1 showing a higher average CoG than Host. The main effect of *Place* was also significant, with alveolar sibilants exhibiting higher CoG values than retroflex sibilants. In addition, the main effect of *Context* was significant as well, with unrounded vowels yielding higher CoG values in the preceding sibilant. These findings were all expected, which further confirmed the sanity check reported in Section 3.2.

The interaction of *Speaker * Place* was also significant, indicating that the difference between the two speakers was larger in retroflex sibilants. Besides, the interaction of *Place * Context* was also significant, suggesting that the effect of lip-rounding differed between the two sibilant categories, with a stronger lowering effect on alveolar sibilants than on retroflex sibilants.

However, no significant effects involving *Phase* were found, thus providing no evidence of accommodation. Notably, the three-way interaction of *Phase * Speaker * Context* was approaching significance, but since this effect was averaged across these two sibilant categories, it was not helpful to address the main research question.

Effect	Estimate	95% CI	*t* value	*p* value
speaker-Host+Guest	1504.13	[1269.49, 1738.77]	12.65	**<.001**
place-Retro+Alveo	1631.63	[1396.99, 1866.27]	13.73	**<.001**
context-Round+Unrou	543.05	[308.41, 777.69]	4.57	**<.001**
speaker-Host+Guest:place-Retro+Alveo	-1045.07	[-1514.35, -575.79]	-4.40	**<.001**
place-Retro+Alveo:context-Round+Unrou	1372.96	[903.68, 1842.24]	5.78	**<.001**
phase-Early+Late:speaker-Host+Guest:context-Round+Unrou	809.62	[-128.94, 1748.18]	1.70	.0904

Table 5. Significant (in bold) and marginally significant results of conv1

To further examine the pattern, the mean CoG values of the two speakers across different conditions in the two phases are shown in Figure 5. For retroflex sibilants, rather than converging or diverging, the two speakers showed small but parallel changes. In the rounded context, both speakers showed slightly lower CoG values in the late phase, whereas in the unrounded context both showed slightly higher values. For alveolar sibilants, different patterns were observed in the two contexts. In the rounded context, a slight convergence was observed. In the unrounded context, the two speakers were initially close but showed slight divergence in the late phase, with roughly equal contribution from both speakers, resulting in a distance of approximately 1000 Hz. Overall, the magnitude of changes was small, which was in line with the non-significant results reported above.

For the change in their alveolar-retroflex contrast, as shown in Figure 6, the guest's contrast showed a slight increase in both contexts, though the magnitude was small. However, the host's contrast did not change consistently across the two contexts: in the rounded context, it became larger; in the unrounded context, it became smaller.

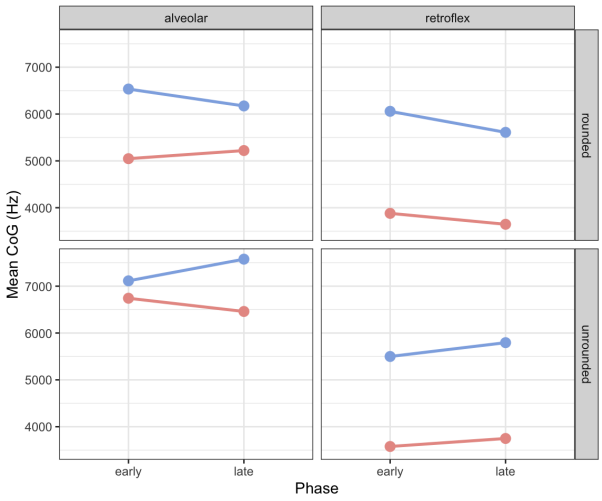


Figure 5. Change in mean CoG

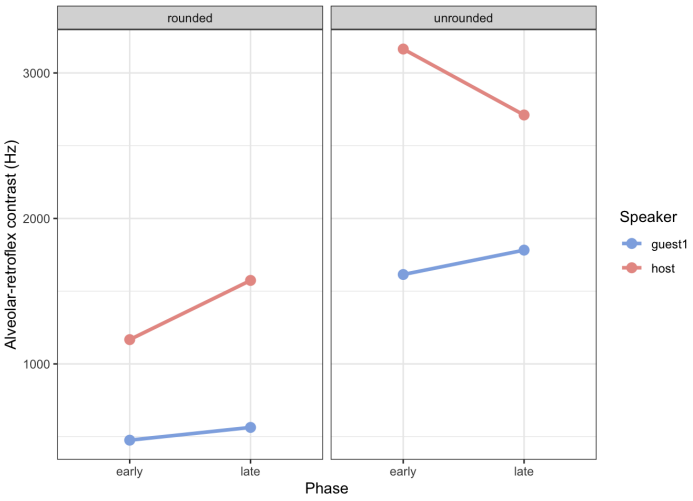


Figure 6. Change in alveolar-retroflex contrast

3.3.2 Conversation 2: Host & Guest 2

*a. CoG ~ Phase (Early vs. Late) * Speaker (Host vs. Guest) * Place (Alveolar vs. Retroflex) * Context (Rounded vs. Unrounded)*

The same linear model *a* was run for Conversation 2. Similar to Conversation 1, the main effects of *Speaker*, *Place*, and *Context*, and the *Place * Context* interaction were again significant.

However, several differences from Conversation 1 were observed, as shown in Table 6. First, the *Speaker * Place* interaction was only marginally significant, in contrast to Conversation 1 where it was significant. This suggests that, unlike Host and Guest 1 who differed more in their retroflex productions, Host and Guest 2 showed a more comparable difference across the two sibilant categories. Second, the *Speaker * Context* interaction and the *Speaker * Place * Context* interaction both reached significance in Conversation 2. These results may suggest that, in Conversation 2, the effect of vowel context varied not only across sibilant categories but also across speakers. That is to say, the effect of lip-rounding on CoG was more speaker-specific than in Conversation 1.

Again, the main effect of *Phase* and all interactions involving it were non-significant, providing no evidence of accommodation. As shown in Figure 7, the initial distance between the two speakers was relatively small. In the alveolar-rounded condition, however, they diverged from their initially close values, with both contributing relatively equally to the change. Furthermore, as the host's alveolar mean CoG decreased and the retroflex mean CoG

increased in this condition, her contrast became smaller accordingly, as shown in Figure 8. The guest's sibilant productions, in contrast, remained relatively stable in both mean CoG and the contrast, though a slight but consistent increase in his contrast was observed in both contexts.

Effect	Estimate	95% CI	*t* value	*p* value
speaker-Host+Guest	232.74	[32.13, 433.36]	2.29	**.0232**
place-Retro+Alveo	1719.88	[1519.26, 1920.49]	16.92	**<.001**
context-Round+Unrou	792.91	[592.29, 993.52]	7.80	**<.001**
speaker-Host+Guest:place-Retro+Alveo	-356.25	[-757.48, 44.98]	-1.75	.0815
speaker-Host+Guest:context-Round+Unrou	-492.32	[-893.55, -91.09]	-2.42	**.0165**
place-Retro+Alveo:context-Round+Unrou	1168.04	[766.81, 1569.27]	5.75	**<.001**
speaker-Host+Guest:place-Retro+Alveo:context-Round+Unrou	-1432.76	[-2235.22, -630.31]	-3.52	**<.001**

Table 6. Significant (in bold) and marginally significant results of conv2

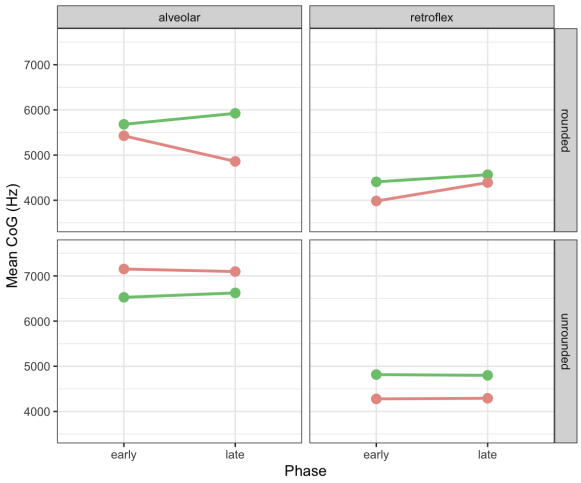


Figure 7. Change in mean CoG

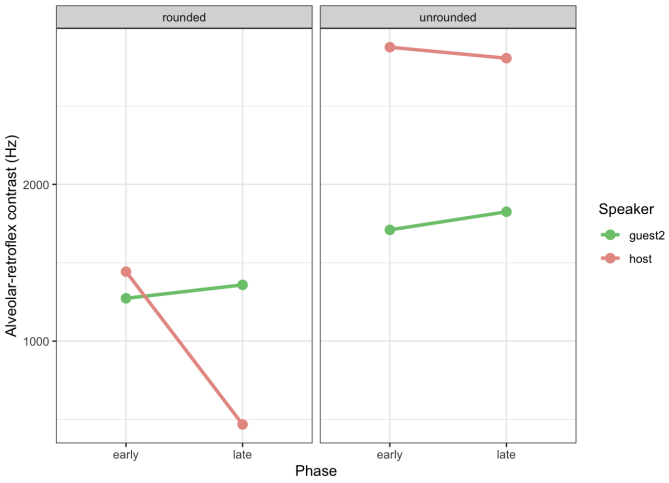


Figure 8. Change in alveolar-retroflex contrast

3.4 Results of the Two Exploratory Analyses

3.4.1 Exploratory Analysis 1: Host solo vs. Host in conversations

*b. CoG ~ Speech (Conv1 vs. Conv2 vs. Solo) * Place (Alveolar vs. Retroflex) * Context (Rounded vs. Unrounded)*

Speech was coded using two orthogonal contrasts: (1) Conversation 1 and Conversation 2 combined as Conversation, vs. Solo (1/3, 1/3, -2/3); (2) Conversation 1 vs. Conversation 2 (-1/2, 1/2, 0).

For Exploratory Analysis 1, the linear model *b* was run to compare the host's speech in the two conversations as well as in the solo talk. To better understand the results in Table 7, the coding of the three-level variable *Speech* (setting) is provided above.

In Table 7, the significant effects of *Place*, *Context*, and the *Place* * *Context* interaction were all expected, as we have seen earlier. Interestingly, the host's solo speech differed significantly from her speech in the two conversations, with overall higher CoG values in the solo speech. The significant *Speech* * *Place* interaction further showed that her alveolar-retroflex contrast was on average larger in the solo talk, suggesting that she maintained a stronger distinction when speaking alone. This can be seen more clearly in Figures 9 and 10.

Additionally, the host's speech also differed between the two conversations, with slightly higher CoG values in Conversation 2 on average. As shown in Figure 9, the difference between the host's speech in the two conversations was greatest in the retroflex-unrounded condition, with a difference of approximately 700 Hz.

Effect	**Estimate**	**95% CI**	***t* value**	***p* value**
speechConversation_vs_Solo	-612.21	[-870.82, -353.61]	-4.68	**<.001**
speechConv1_vs_Conv2	398.60	[85.87, 711.33]	2.52	**.0129**
place-Retro+Alveo	2751.17	[2501.52, 3000.81]	21.80	**<.001**
context-Round+Unrou	886.19	[636.55, 1135.84]	7.02	**<.001**
speechConversation_vs_Solo:place-Retro+Alveo	-1765.96	[-2283.16, -1248.75]	-6.76	**<.001**
place-Retro+Alveo:context-Round+Unrou	1725.52	[1226.23, 2224.81]	6.84	**<.001**

Table 7. Significant results of exploratory analysis 1

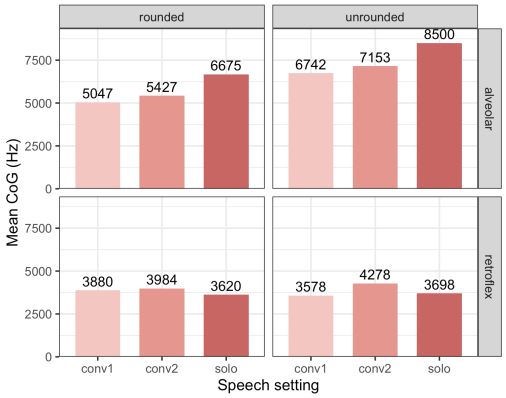


Figure 9. Host's mean CoG in three speech settings

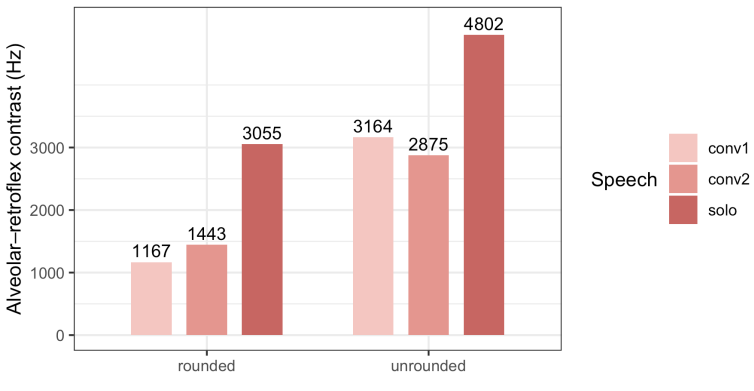


Figure 10. Host's contrast in three speech settings

3.4.2 Exploratory Analysis 2: Three phases of Conversation 2

*c. CoG ~ Phase (Early vs. Late vs. Extended) * Speaker (Host vs. Guest) * Place (Alveolar vs. Retroflex) * Context (Rounded vs. Unrounded)*

Phase was coded using two orthogonal contrasts: (1) Late and Extended combined as Later, vs. Early (1/3, 1/3, -2/3); (2) Late vs. Extended (-1/2, 1/2, 0).

For Exploratory Analysis 2, an additional phase of Conversation 2 was included to examine whether any accommodation pattern continued to develop over a longer interaction. The early, late, and extended phases represented three stages of Conversation 2, corresponding to the beginning, middle, and near-end of the interaction. For simplicity, they are referred to as the first, second, and third phase of Conversation 2 in the following paragraphs. The linear model *c* was run for the three phases, and the coding of the three-level variable *Phase* is provided above.

The results, shown in Table 8, were more complicated than those reported earlier. The main effects of *Speaker*, *Place*, and *Context* were all significant, as expected. The *Place* * *Context* interaction and the *Speaker* * *Place* * *Context* interaction also reached significance, consistent with the results reported for the first two phases of Conversation 2.

With the third phase included, several marginally significant and significant interactions involving *Phase* emerged. Most importantly, the four-way interaction of *Phase* * *Speaker* * *Place* * *Context* was significant, providing evidence for accommodation. Figure 11 visualized the accommodation pattern. Across the three phases, the guest remained relatively stable, with

Effect	Estimate	95% CI	*t* value	*p* value
speaker-Host+Guest	301.10	[131.21, 470.99]	3.49	**< .001**
place-Retro+Alveo	1680.54	[1510.65, 1850.43]	19.48	**< .001**
context-Round+Unrou	614.19	[444.30, 784.08]	7.12	**< .001**
phaseEarly_vs_Later:context-Round+Unrou	-306.47	[-670.53, 57.59]	-1.66	.0986
phaseLate_vs_Extended:context-Round+Unrou	-510.56	[-922.43, -98.70]	-2.44	**.0153**
place-Retro+Alveo:context-Round+Unrou	961.49	[621.71, 1301.27]	5.57	**< .001**
phaseLate_vs_Extended:speaker-Host+Guest:context-Round+Unrou	911.83	[88.10, 1735.56]	2.18	**.0302**
phaseLate_vs_Extended:place-Retro+Alveo:context-Round+Unrou	-853.37	[-1677.10, -29.63]	-2.04	**.0424**
speaker-Host+Guest:place-Retro+Alveo:context-Round+Unrou	-750.23	[-1429.79, -70.66]	-2.17	**.0306**
phaseLate_vs_Extended:speaker-Host+Guest:place-Retro+Alveo:context-Round+Unrou	2485.18	[837.72, 4132.64]	2.97	**.0033**

Table 8. Significant (in bold) and marginally significant results of exploratory analysis 2

almost no change in the retroflex-unrounded condition and only very slight increases in the other three conditions. In contrast, the host showed larger but less consistent changes, particularly in her alveolar productions, whereas her retroflex productions remained relatively stable. Notably, in the retroflex-unrounded condition where the guest was the most stable, the host also showed little change. The resulting changes in their contrasts are shown in Figure 12. By the third phase, the host's contrast in the rounded context had returned to a level similar to its initial value, which was also close to the guest's; in the unrounded context, it decreased and became more closer to the guest's.

Overall, although accommodation evidence emerged with the inclusion of the third phase, it was largely driven by the host's changes, which were inconsistent rather than reflecting a clear pattern of convergence or divergence. Nevertheless, by the end of the conversation, the two speakers' contrasts in both contexts had become more similar to each other overall.

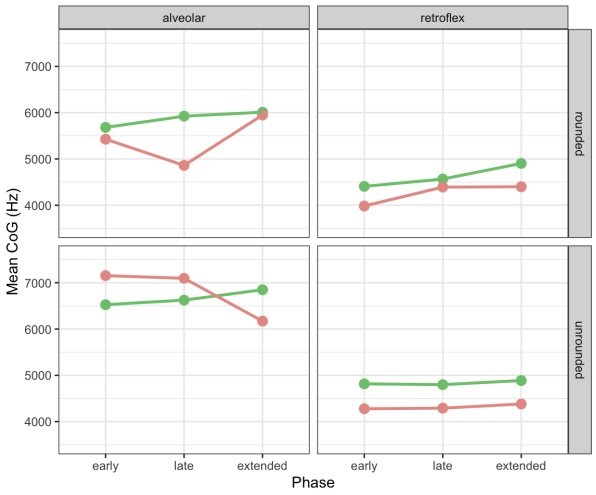


Figure 11. Change in mean CoG

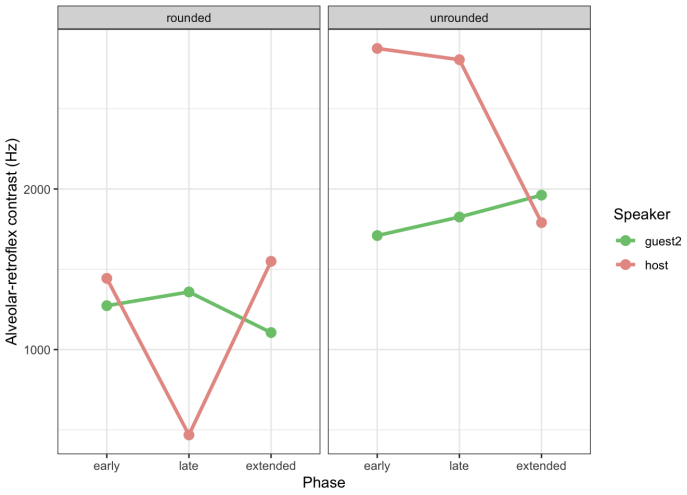


Figure 12. Change in alveolar-retroflex contrast

4. Discussion

This section first discusses the findings of the present study in relation to the two research questions, then proposes several possible explanations for the observed patterns, and finally provides some suggestions for future research.

4.1 Main Finding: Little Evidence for Accommodation

(RQ1): Do two interlocutors become more similar as the conversation goes on?
(H1): The two interlocutors are expected to exhibit phonetic convergence over the course of interaction, such that their realization of /s, ts/ and /ʂ/ will be more similar in the late phase than in the early phase of the conversation.

For the first research question, the results did not support the hypothesis: neither conversation showed statistically significant evidence for convergence. Although some feature-specific convergence could be observed, occasional divergence and synchrony were also present, and the overall magnitude of the changes was small. Taken together, these results suggest that accommodation in Mandarin sibilant production was limited in the two conversations.

The two exploratory analyses provide additional insights into the main analysis. The results of Exploratory Analysis 1 showed that the host's speech in the solo talk differed from her speech in interaction, and she maintained a stronger alveolar-retroflex contrast in the solo speech. One possible reason could be that accommodation happened very quickly, such that some accommodation had already taken place in the early phase of the interaction. Alternatively, the host and the guest may have had already talked before the recording session and already reached a certain degree of convergence by that point. The fact that the solo episode was recorded on a different occasion does not seem to be a major concern, as the host's speech was overall comparable across the early phases of the two conversations, even though these were also recorded on different days. Nevertheless, the host's solo talk should still be interpreted with caution. In the solo episode, the host may have been more aware of her role as a podcaster addressing an audience, and may therefore have paid more attention to her pronunciation and the maintenance of a standard speaking style. In the conversations, by contrast, the guest was the primary interactional focus, which may have reduced such monitoring. Such difference in speaking style could also account for her stronger alveolar-retroflex distinction in the solo speech.

Similar to the main analysis, Exploratory Analysis 2 did not provide strong evidence for accommodation either. Although the four-way interaction in the extended phase of Conversation 2 reached significance, most of the changes were driven by the host, and her changes did not show a consistent pattern of convergence or divergence. This may simply reflect the fluctuation of her contrast in the interaction, rather than genuine accommodation, particularly given that she exhibited an overall greater variability (see Figure 2).

Taken together, the results of the main analysis and the two exploratory analyses do not allow the present study to determine whether the speakers had already accommodated by the early phase, or whether accommodation occurred on other features not examined here. What can be said with more confidence is that the degree of convergence, or accommodation, did not increase as the interaction progressed. This is unlikely to be due to insufficient interaction time, as the recordings had a duration of one to three hours, and the actual unedited interactions could be even longer, which further suggests that a longer duration does not necessarily lead to greater convergence between interlocutors.

> *(RQ2): If convergence occurs, is it symmetric or asymmetric between the two interlocutors?*
> *(H2): Convergence is expected to be asymmetric, which means the host would contribute more to the convergence. This prediction is based on two reasons. First, from a linguistic perspective, the guests' accent is more salient than that of the host, making the host more likely to adjust her speech than the guests. Second, from a social perspective, the host is*

expected to adapt more because her role is to engage with the guests and keep the conversation going smoothly.

For the second research question, the present study cannot provide a meaningful answer, given the limited evidence for accommodation. Nevertheless, the data did show that the two guests' productions were relatively stable, whereas the host showed more variation. Compared to their differences in accent background and interactional role, such variation may be more likely to reflect individual differences in speech production. Some speakers tend to articulate more clearly, whereas others exhibit greater variability in their speech. In natural speech, fast speaking rate may further contribute to this difference.

4.2 Possible Explanations for the Limited Evidence of Accommodation

The limited evidence of accommodation in the present data can be considered on two levels: the linguistic features themselves and the interlocutors who produced them.

First, why did convergence not emerge on the examined consonantal features? This finding is in fact consistent with previous Mandarin studies reporting no evidence for convergence over time for suprasegmental features in task-oriented conversations (Xia et al., 2014, 2023; Xia, 2024) and in talk shows (Sun, Ding, 2023, 2024). It is also compatible with Weise et al. (2019), who found that synchrony was more prevalent than convergence in English conversations, with both convergence and divergence emerging on specific suprasegmental features. A similar pattern also emerged in the present data. However, the underlying mechanisms of convergence on segmental features and suprasegmental features may be different. For suprasegmental features, such as speaking rate and pitch, interlocutors can adjust rapidly as feedback to a partner in conversation. For example, when one interlocutor becomes excited and raises their pitch, the other may also raise their pitch in response. But there seems to be little reason to expect interlocutors to pay close attention to the moment-to-moment realization of consonants and adjust in response to each other. Therefore, the synchrony observed in some specific conditions in the present data may be simply due to the fact that only two phases were examined.

On the other hand, the findings of the present study contrast with those of Wagner et al. (2021). They found significant convergence on the CoG of the fricative /s/ in Dutch, a pattern not found for Mandarin /s/ in the present study. This difference may be due to a setting difference. Wagner et al. (2021) used a repetition task, in which participants repeated a stimulus they had just heard. Since the phonetic information was still fresh in memory, participants were more likely to notice and imitate the target feature. In natural speech, however, speech is produced quickly and carries a great deal of information at multiple levels simultaneously, leaving less attention available for a single phonetic feature, even a relatively salient one. Additionally, Wagner et al. (2021) reported significant divergence on the fricative duration, which suggests that even within a single segment, convergence may not occur uniformly across all acoustic properties. Therefore, it is possible that the speakers in the present study were accommodating on other properties of these sibilants, such as frication duration, rather than on CoG, which primarily reflects place of articulation. More broadly,

accommodation may also have occurred in other segmental or suprasegmental features, rather than on sibilant production per se.

Additionally, accommodation could have been obscured by variability in the segments examined. While retroflex productions remained relatively stable throughout the interactions, alveolar productions showed greater variation. This may be due to the fact that alveolar sibilants are simply more variable in spontaneous speech, especially when speaking rate increases. Compared with retroflex sibilants, which require a more constrained tongue configuration, alveolar sibilants may be more easily affected by articulatory reduction and therefore more variable. Such variation may have obscured the already weak convergence. The influence of lip-rounding may have further blurred the picture. In the present data, sibilants followed by rounded vowels showed more variation, particularly for alveolar sibilants. As alveolar sibilants have a a shorter articulatory tract, lip-rounding may therefore have a much stronger effect on their CoG. Furthermore, some alveolar-rounded tokens were function words, which tend to be produced faster than content words, further contributing to the variation. Taken together, the greater variability of alveolar productions, together with the additional influence of lip-rounding, may have masked subtle accommodation effects, resulting in the inconsistent patterns observed in the present data.

Second, why did the two interlocutors not converge with each other? From a linguistic perspective, regardless of the debate on whether greater or smaller acoustic distance leads to greater convergence, most accounts agree that there should be at least some phonetic space for convergence to happen. Compared to Guest 1, Guest 2's initial acoustic distance from the host was relatively small, thus the difference may be too small to be salient to the host and leaves limited room for further convergence. However, this explanation could not fully account for the patterns observed in Conversation 1. Although Guest 1 showed a noticeable acoustic difference from the host, particularly in his retroflex productions, both speakers' retroflex productions remained stable in the interaction. Therefore, language distance seems to have a limited explanation power for the limited evidence for convergence in Conversation 1. Another possible explanation concerns limited language ability. Both guests were already in their seventies, and retroflex sibilants have been reported to be especially difficult for SHM speakers to acquire. If this contrast had remained difficult for them after a lifetime of speaking Mandarin, it seems unrealistic to expect a single conversation to bring about change. This interpretation is consistent with Babel's (2009) proposal that accommodation may be constrained by speakers' existing phonetic repertoires. Notably, except for Guest 2's decreased contrast in the rounded context in the extended phase, both guests showed slight increases in their alveolar-retroflex contrast. Although such increase was not statistically significant, it may still suggest a tendency toward reinforcing the contrast.

Social factors may also have contributed to the lack of accommodation. For the host, since the alveolar-retroflex merger is considered non-standard in Mandarin, this accent feature may not be attractive to her, and she therefore had little motivation to converge toward it. For the two guests, since they were both successful public figures, they may already be accustomed to public conversations, which could make their speech more stable and less likely to change within a single interaction.

In summary, several factors may help explain why only limited evidence of accommodation was observed in the present study. However, the absence of clear effects does not necessarily imply the absence of accommodation. It remains possible that accommodation occurred in ways that were not captured by the present study. Future studies could further examine this possibility, as detailed in the next section.

4.3 Limitations and Future Directions

Several limitations of the present study can be addressed by future research. First, the present study focused only on CoG as an acoustic measure of sibilant production. Future studies could examine multiple acoustic properties and combine them with perceptual judgments. As pointed out by Pardo (2013), acoustic measures combined with perceptual assessments may provide a more comprehensive understanding of the phenomenon. Second, the present study examined only two phases in Conversation 1 and three phases in Conversation 2. Future research could divide conversations into more fine-grained phases to better capture the dynamic development of accommodation over time. Third, only around 20 tokens were extracted per phase, per category, per speaker. Future studies could include more tokens to improve the reliability of the results.

More broadly, the present study highlights the value of investigating accommodation in natural speech. Although spontaneous interaction introduces more variability and makes accommodation more difficult to identify, it provides an important test of whether patterns observed in laboratory settings generalize to real-world communication, and therefore deserves much more attention in future research. In addition, accommodation in Mandarin remains seriously understudied, particularly on segmental features. Compared with the extensive literature on English and other European languages, research on Mandarin is quite little. This gap is particularly striking given that Mandarin contains a rich range of regional varieties and accents, making it a very interesting context to study the phenomenon. More research is therefore needed on Mandarin, particularly on how accommodation works between Mandarin speakers with different dialect or accent backgrounds.

5. Conclusion

The present study investigated consonantal convergence in Mandarin natural interaction. It analyzed two podcast conversations between the same Standard Mandarin-speaking host and two different guests with Shanghai-accented Mandarin. The alveolar-retroflex merger, a Shanghai accent feature, was targeted and measured by the center of gravity (CoG) of the frication portion of the sibilants. Specifically, the study examined whether the interlocutors' sibilant productions became more similar over the course of interaction by comparing the early and late phases of each conversation, and also explored whether any such accommodation was symmetric or asymmetric between speakers. Two exploratory analyses were also conducted: the first examined the host's solo speech to establish her sibilant productions without interaction, and the second included an additional later phase of

Conversation 2 to investigate whether any accommodation continued to develop over a longer interaction.

Overall, the results provided little evidence for accommodation. Neither conversation showed significant convergence over time, and the observed changes were generally small. The exploratory analyses provided some additional insights: the host's solo speech showed a stronger alveolar-retroflex distinction than her speech in conversation; the extended phase of Conversation 2 showed inconsistent accommodation patterns and thus did not provide clear evidence for convergence. Several factors may have contributed to the patterns observed. In particular, variation in the host's speech and the relatively high variability of alveolar productions may have obscured subtle accommodation effects. Notably, both guests showed slight increases in their alveolar-retroflex contrast, suggesting a possible tendency toward convergence.

The present study is among the first to investigate segmental convergence over time in natural Mandarin conversations. Although little evidence was found for accommodation, the study highlights the importance of examining accommodation in naturalistic speech, despite its challenges outside laboratory settings. More importantly, the study calls for greater attention to accommodation research in Mandarin, which would offer a new perspective through which to understand this complex phenomenon better.

6. Acknowledgments

I would like to express my deepest gratitude to my supervisor, Dr. Silke Hamann. She has been so patient and encouraging throughout this process, always willing to think through ideas with me and answer my questions. Her guidance helped keep me on track whenever my thinking wandered too far, allowing me to see a way forward when I felt stuck. I am truly grateful for the time she devoted to this project. I would also like to thank Dr. Paul Boersma and my friend Xintong Xu for their valuable statistical suggestions.

This is the first time I have carried out such a formal research project, taking a question that arose from my everyday observations and developing it into a completed thesis through scientific inquiry. I feel truly grateful for this experience, and I look forward to carrying this curiosity into whatever comes next.

References

Abel, J. (2012). Phonetic imitation, accommodation, and audience design in Canadian radio interviews. In A. Black & M. Louie (Eds.), *UBC qualifying papers 2 (2011–2012) (University of British Columbia Working Papers in Linguistics*, Vol. 34, pp. 1–14). University of British Columbia.

Ashby, P. (2013). *Understanding phonetics*. Routledge.

Barón-Birchenall, L. (2023). Phonetic accommodation during conversational interactions: An overview. *Revista Guillermo de Ockham, 21*(2), 493–517. https://doi.org/10.21500/22563202.6150

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

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*(4), 437–456. https://doi.org/10.1017/S0047404510000400

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40*(1), 177–189.

Boersma, P., & Weenink, D. (2025). *Praat: Doing phonetics by computer* (Version 6.4.67) [Computer software]. University of Amsterdam. https://www.praat.org/

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*(2), B13–B25.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 22*(6), 1482–1493. https://doi.org/10.1037//0278-7393.22.6.1482

Cao, W. (2019). Phonetic convergence of Hong Kong English: Sound salience and the exemplar-based account. In *Proceedings of the 19th international congress of phonetic sciences, Melbourne, Australia* (pp. 2334–2338).

Clopper, C. G., & Dossey, E. (2020). Phonetic convergence to Southern American English: Acoustics and perception. *Journal of the Acoustical Society of America, 147*(1), 671–683. https://doi.org/10.1121/10.0000555

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica, 64*(2–3), 145–173. https://doi.org/10.1159/000107914

Dippong, J. (2020). Status and vocal accommodation in small groups. *Sociological Science, 7*, 291–313. https://doi.org/10.15195/v7.a12

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

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, N. Coupland, & J. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge University Press. https://doi.org/10.1017/CBO9780511663673.001

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251

Gregory Jr, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions.

Journal of Personality and Social Psychology, 70(6), 1231–1240. https://doi.org/10.1037/0022-3514.70.6.1231

Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association, 32*(2), 141–174. https://doi.org/10.1017/S0025100302001020

Hamann, S., & Avelino, H. (2007). An acoustic study of plain and palatalized sibilants in Ocotepec Mixe. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 949–952).

Kania, A. N., & Roselani, N. G. A. (2025). English alveolar /t/ and approximant /ɹ/ convergence phenomenon in Indonesians' interactions: Reflection of social dynamics. *Lexicon, 12*(1), 13–24. https://doi.org/10.22146/lexicon.v12i1.104977

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. https://doi.org/10.1515/labphon.2011.004

Levitan, R., Gravano, A., Willson, L., Beňuš, Š., Hirschberg, J., & Nenkova, A. (2012). Acoustic-prosodic entrainment and social behavior. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 11–19).

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proc. Interspeech 2011* (pp. 3081–3084).

Lewandowski, E. M., & Nygaard, L. C. (2018). Vocal alignment to native and non-native speakers of English. *Journal of the Acoustical Society of America, 144*(2), 620–633. https://doi.org/10.1121/1.5038567

Li, A., & Wang, X. (2003). A contrastive investigation of standard Mandarin and accented Mandarin. In *Proc. Eurospeech 2003* (pp. 2345–2348).

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science, 36*(8), 1404–1426. https://doi.org/10.1111/j.1551-6709.2012.01269.x

Muir, K., Joinson, A., Cotterill, R., & Dewdney, N. (2017). Linguistic style accommodation shapes impression formation and rapport in computer-mediated communication. *Journal of Language and Social Psychology, 36*(5), 525–548. https://doi.org/10.1177/0261927X17701327

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology, 21*(4), 422–432. https://doi.org/10.1177/026192702237958

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21*(4), 337–360. https://doi.org/10.1177/026192702237953

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

Paquette-Smith, M., Schertz, J., & Johnson, E. K. (2022). Comparing phonetic convergence in children and adults. *Language and Speech, 65*(1), 240–260. https://doi.org/10.1177/00238309211013864

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Pardo, J. S. (2010). Expressing oneself in conversational interaction. In E. Morsella (Ed.), *Expressing oneself/expressing one's self: Communication, cognition, language, and identity* (pp. 183–196). Psychology Press.

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology, 4*, Article 559. https://doi.org/10.3389/fpsyg.2013.00559

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics, 72*(8), 2254–2264. https://doi.org/10.3758/APP.72.8.2254

Pardo, J., Urmanche, A., Wilman, S., & Wiener, J. (2016). Phonetic convergence and talker sex: It's complicated. *Journal of the Acoustical Society of America, 139*(4), 2105–2106. https://doi.org/10.1121/1.4950256

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. https://doi.org/10.1016/j.wocn.2018.04.001

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056

R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Reverdy, J., Koutsombogera, M., & Vogel, C. (2019). Linguistic repetition in three-party conversations. In A. Esposito, M. Faundez-Zanuy, F. C. Morabito, & E. Pasero (Eds.), *Neural approaches to dynamics of signal exchanges* (pp. 359–370). Springer Singapore. https://doi.org/10.1007/978-981-13-8950-4_32

Ross, J., Lilley, K. D., Clopper, C. G., Pardo, J. S., & Levi, S. V. (2021). Effects of dialect-specific features and familiarity on cross-dialect phonetic convergence. *Journal of Phonetics, 86*, Article 101041. https://doi.org/10.1016/j.wocn.2021.101041

Schweitzer, A., & Lewandowski, N. (2014). Social factors in convergence of F1 and F2 in spontaneous speech. In *Proceedings of the 10th International Seminar on Speech Production* (ISSP 2014) (pp. 391–394).

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics, 66*(3), 422–429. https://doi.org/10.3758/BF03194890

Smith, B. J., Mielke, J., Magloughlin, L., & Wilbanks, E. (2019). Sound change and coarticulatory variability involving English /ɹ/. *Glossa: A Journal of General Linguistics, 4*(1), Article 63. https://doi.org/10.5334/gjgl.650

Sun, Y., & Ding, H. (2023). Speech entrainment in Chinese story-style talk shows: The interaction between gender and role. In *Proc. Interspeech 2023* (pp. 3537–3541).

Sun, Y., & Ding, H. (2024). Unpacking the gender-role interaction of prosodic entrainment in Chinese long-and-short turn-taking: evidence from perceptual and acoustic similarities. *Humanities and Social Sciences Communications, 11*(1), Article 1618. https://doi.org/10.1057/s41599-024-04137-4

Thomson, R., Murachver, T., & Green, J. (2001). Where is the gender in gendered language?. *Psychological Science, 12*(2), 171–175. https://doi.org/10.1111/1467-9280.00329

Tobin, S. J. (2022). Effects of native language and habituation in phonetic accommodation. *Journal of Phonetics, 93*, Article 101148. https://doi.org/10.1016/j.wocn.2022.101148

Ulbrich, C. (2024). Phonetic accommodation on the segmental and the suprasegmental level of speech in native–non-native collaborative tasks. *Language and Speech, 67*(2), 346–372. https://doi.org/10.1177/00238309211050094

Wagner, M. A., Broersma, M., McQueen, J. M., Dhaene, S., & Lemhöfer, K. (2021). Phonetic convergence to non-native speech: Acoustic and perceptual evidence. *Journal of Phonetics, 88*, Article 101076. https://doi.org/10.1016/j.wocn.2021.101076

Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology, 6*, Article 546. https://doi.org/10.3389/fpsyg.2015.00546

Ward, A., & Litman, D. (2007). Dialog convergence and learning. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Artificial intelligence in education: Building technology rich learning contexts that work* (pp. 262–269). IOS Press.

Weise, A., Levitan, S. I., Hirschberg, J., & Levitan, R. (2019). Individual differences in acoustic-prosodic entrainment in spoken dialogue. *Speech Communication, 115*, 78–87. https://doi.org/10.1016/j.specom.2019.10.007

Wieling, M., Tiede, M., Rebernik, T., de Jong, L., Braggaar, A., Bartelds, M., Medvedeva, M., Heisterkamp, P., Freire Offrede, T., Sekeres, H., Pot, A., Ploeg, M. van der, Volkers, K., & Mills, G. (2021). A novel paradigm to investigate phonetic convergence in interaction. In M. Tiede, D. H. Whalen, & V. Gracco (Eds.), *Proceedings of the 12th International Seminar on Speech Production* (pp. 1–4). Haskins Press.

Xia, Z., Hirschberg, J., & Levitan, R. (2023). Investigating prosodic entrainment from global conversations to local turns and tones in Mandarin conversations. *Speech Communication, 153*, Article 102961. https://doi.org/10.1016/j.specom.2023.102961

Xu, Y., & Reitter, D. (2016). Entropy converges between dialogue participants: Explanations from an information-theoretic perspective. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 537–546). Association for Computational Linguistics.

Yao, Y. (1988). A brief account of Shanghai-accented Mandarin. *Language Teaching and Linguistic Studies, 4*, 120–128. [In Chinese]

Ye, J. (2012). A dynamic analysis of accent-related linguistic features in Shanghai-accented Mandarin. *Applied Linguistics, 2012*(S1), 130–138. [In Chinese]

Yu, J. (2004). *An acoustic comparative study of the vowel systems of Shanghai-accented Mandarin and Standard Mandarin* [Master's thesis, Zhejiang University]. [in Chinese]

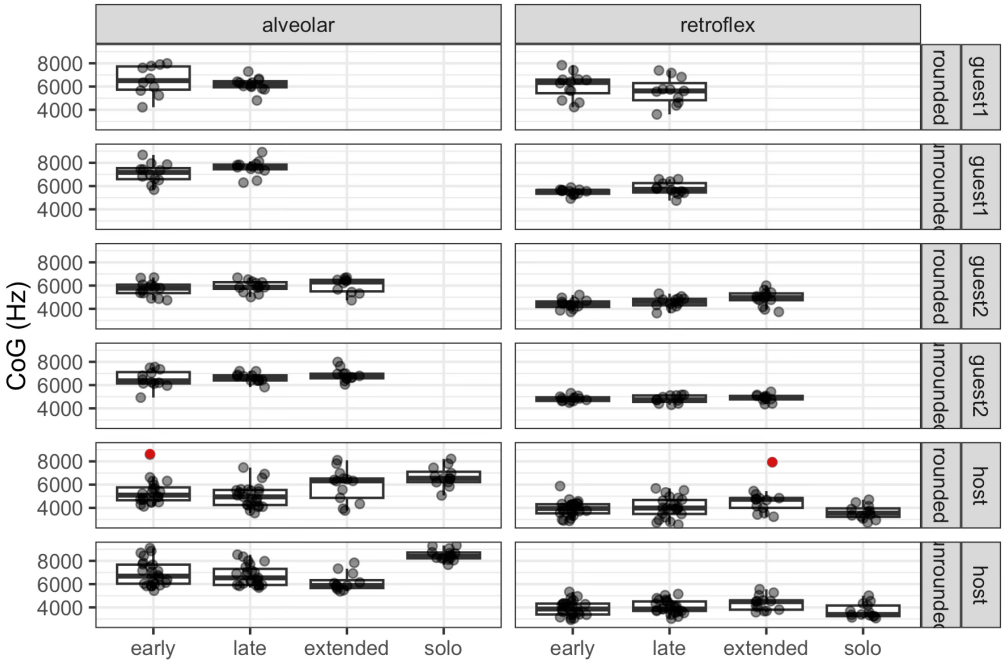
Yu, J., Li, A., & Wang, X. (2008). An acoustic comparison of retroflex vowels in Shanghai -accented Mandarin and Standard Mandarin. *Contemporary Linguistics, 10*(3), 211–219. [In Chinese]

Appendix 1 — Script:

```
# 1. Read input files

sound = Read from file: "conv01_early.wav"
textgrid = Read from file: "conv01_early.TextGrid"

# 2. Find tiers by name

selectObject: textgrid
speakerTier = 1
tokenTier = 2
contextTier = 3

writeInfoLine: "speaker, token, context, duration(s), CoG(Hz)"

# 3. Get number of intervals

n = Get number of intervals: tokenTier

# 4. Loop through all non-empty intervals

for i from 1 to n

    selectObject: textgrid
    token$ = Get label of interval: tokenTier, i

    if token$ <> ""
        speaker$ = Get label of interval: speakerTier, i
  context$ = Get label of interval: contextTier, i

        start = Get start time of interval: tokenTier, i
  end = Get end time of interval: tokenTier, i
        duration = end - start

        # 7. Extract interval

  selectObject: sound
  Extract part: start, end, "rectangular", 1, "yes"

  part = selected("Sound")

        # 8. Convert to Spectrum

  To Spectrum: "yes"
  spectrum = selected("Spectrum")
```

```
                # 9. Measure CoG

        cog = Get centre of gravity: 2

                # 10. Print results
                appendInfoLine: speaker$, ", ", token$, ", ", context$, ", ", fixed$(duration, 3),
", ", fixed$(cog, 0)

                selectObject: part
                plusObject: spectrum
                Remove
            endif
endfor

# 11. Remove all extra objects
selectObject: sound
Remove

selectObject: textgrid
Remove
```

Appendix 2 — Data distribution with removed outliers:



Appendix 3 — Complete results of statistical analyses:

Main Analysis, Conversation 1:

Effect	Estimate	95% CI	*t* value	*p* value
(Intercept)	5542.59	[5425.27, 5659.91]	93.26	<.001
phase-Early+Late	-27.92	[-262.56, 206.72]	-0.23	.8146
speaker-Host+Guest	1504.13	[1269.49, 1738.77]	12.65	<.001
place-Retro+Alveo	1631.63	[1396.99, 1866.27]	13.73	<.001
context-Round+Unrou	543.05	[308.41, 777.69]	4.57	<.001
phase-Early+Late:speaker-Host+Guest	29.90	[-439.38, 499.18]	0.13	.9001
phase-Early+Late:place-Retro+Alveo	52.23	[-417.05, 521.51]	0.22	.8264
speaker-Host+Guest:place-Retro+Alveo	-1045.07	[-1514.35, -575.79]	-4.40	<.001
phase-Early+Late:context-Round+Unrou	379.04	[-90.24, 848.32]	1.59	.1127
speaker-Host+Guest:context-Round+Unrou	-281.13	[-750.41, 188.15]	-1.18	.2386
place-Retro+Alveo:context-Round+Unrou	1372.96	[903.68, 1842.24]	5.78	<.001
phase-Early+Late:speaker-Host+Guest:place-Retro+Alveo	150.82	[-787.74, 1089.38]	0.32	.7515
phase-Early+Late:speaker-Host+Guest:context-Round+Unrou	809.62	[-128.94, 1748.18]	1.70	.0904
phase-Early+Late:place-Retro+Alveo:context-Round+Unrou	-390.20	[-1328.76, 548.36]	-0.82	.4130
speaker-Host+Guest:place-Retro+Alveo:context-Round+Unrou	-388.45	[-1327.01, 550.11]	-0.82	.4151
phase-Early+Late:speaker-Host+Guest:place-Retro+Alveo:context-Round+Unrou	940.20	[-936.92, 2817.32]	0.99	.3242

Main Analysis, Conversation 2:

Effect	Estimate	95% CI	*t* value	*p* value
(Intercept)	5301.67	[5201.37, 5401.98]	104.34	<.001
phase-Early+Late	35.06	[-165.56, 235.67]	0.34	.7306
speaker-Host+Guest	232.74	[32.13, 433.36]	2.29	.0232
place-Retro+Alveo	1719.88	[1519.26, 1920.49]	16.92	<.001
context-Round+Unrou	792.91	[592.29, 993.52]	7.80	<.001
phase-Early+Late:speaker-Host+Guest	171.13	[-230.10, 572.36]	0.84	.4010
phase-Early+Late:place-Retro+Alveo	-210.99	[-612.22, 190.24]	-1.04	.3007
speaker-Host+Guest:place-Retro+Alveo	-356.25	[-757.48, 44.98]	-1.75	.0815
phase-Early+Late:context-Round+Unrou	-51.19	[-452.42, 350.04]	-0.25	.8015
speaker-Host+Guest:context-Round+Unrou	-492.32	[-893.55, -91.09]	-2.42	.0165
place-Retro+Alveo:context-Round+Unrou	1168.04	[766.81, 1569.27]	5.75	<.001
phase-Early+Late:speaker-Host+Guest:place-Retro+Alveo	623.14	[-179.32, 1425.60]	1.53	.1272
phase-Early+Late:speaker-Host+Guest:context-Round+Unrou	-218.11	[-1020.57, 584.35]	-0.54	.5923
phase-Early+Late:place-Retro+Alveo:context-Round+Unrou	467.41	[-335.05, 1269.87]	1.15	.2518
speaker-Host+Guest:place-Retro+Alveo:context-Round+Unrou	-1432.76	[-2235.22, -630.31]	-3.52	<.001
phase-Early+Late:speaker-Host+Guest:place-Retro+Alveo:context-Round+Unrou	-875.12	[-2480.04, 729.79]	-1.08	.2833

Exploratory Analysis 1, host solo talk:

Effect	Estimate	95% CI	*t* value	*p* value
(Intercept)	5215.33	[5090.51, 5340.15]	82.66	<.001
speechConversation_vs_Solo	-612.21	[-870.82, -353.61]	-4.68	<.001
speechConv1_vs_Conv2	398.60	[85.87, 711.33]	2.52	.0129
place-Retro+Alveo	2751.17	[2501.52, 3000.81]	21.80	<.001
context-Round+Unrou	886.19	[636.55, 1135.84]	7.02	<.001
speechConversation_vs_Solo:place-Retro+Alveo	-1765.96	[-2283.16, -1248.75]	-6.76	<.001
speechConv1_vs_Conv2:place-Retro+Alveo	-6.47	[-631.93, 618.99]	-0.02	.9837
speechConversation_vs_Solo:context-Round+Unrou	-98.62	[-615.83, 418.58]	-0.38	.7066
speechConv1_vs_Conv2:context-Round+Unrou	313.64	[-311.82, 939.10]	0.99	.3230
place-Retro+Alveo:context-Round+Unrou	1725.52	[1226.23, 2224.81]	6.84	<.001
speechConversation_vs_Solo:place-Retro+Alveo:context-Round+Unrou	-32.65	[-1067.05, 1001.76]	-0.06	.9503
speechConv1_vs_Conv2:place-Retro+Alveo:context-Round+Unrou	-565.39	[-1816.31, 685.53]	-0.89	.3729

Exploratory Analysis 2, extended phase of Conversation 2:

Effect	Estimate	95% CI	*t* value	*p* value
(Intercept)	5349.08	[5264.13, 5434.02]	124.00	< .001
phaseEarly_vs_Later	97.40	[-84.63, 279.43]	1.05	.2930
phaseLate_vs_Extended	124.68	[-81.25, 330.62]	1.19	.2343
speaker-Host+Guest	301.10	[131.21, 470.99]	3.49	< .001
place-Retro+Alveo	1680.54	[1510.65, 1850.43]	19.48	< .001
context-Round+Unrou	614.19	[444.30, 784.08]	7.12	< .001
phaseEarly_vs_Later:speaker-Host+Guest	230.89	[-133.17, 594.95]	1.25	.2128
phaseLate_vs_Extended:speaker-Host+Guest	119.53	[-292.34, 531.39]	0.57	.5682
phaseEarly_vs_Later:place-Retro+Alveo	-217.25	[-581.31, 146.81]	-1.18	.2410
phaseLate_vs_Extended:place-Retro+Alveo	-12.51	[-424.38, 399.36]	-0.06	.9524
speaker-Host+Guest:place-Retro+Alveo	-282.96	[-622.74, 56.82]	-1.64	.1022
phaseEarly_vs_Later:context-Round+Unrou	-306.47	[-670.53, 57.59]	-1.66	.0986
phaseLate_vs_Extended:context-Round+Unrou	-510.56	[-922.43, -98.70]	-2.44	.0153
speaker-Host+Guest:context-Round+Unrou	-224.73	[-564.51, 115.05]	-1.30	.1939
place-Retro+Alveo:context-Round+Unrou	961.49	[621.71, 1301.27]	5.57	< .001
phaseEarly_vs_Later:speaker-Host+Guest:place-Retro+Alveo	577.29	[-150.84, 1305.41]	1.56	.1197
phaseLate_vs_Extended:speaker-Host+Guest:place-Retro+Alveo	-91.70	[-915.43, 732.03]	-0.22	.8266
phaseEarly_vs_Later:speaker-Host+Guest:context-Round+Unrou	237.81	[-490.31, 965.93]	0.64	.5207

phaseLate_vs_Extended:speaker-Host+Guest:context-Round+Unrou	911.83	[88.10, 1735.56]	2.18	.0302
phaseEarly_vs_Later:place-Retro+Alveo:context-Round+Unrou	40.72	[-687.40, 768.85]	0.11	.9124
phaseLate_vs_Extended:place-Retro+Alveo:context-Round+Unrou	-853.37	[-1677.10, -29.63]	-2.04	.0424
speaker-Host+Guest:place-Retro+Alveo:context-Round+Unrou	-750.23	[-1429.79, -70.66]	-2.17	.0306
phaseEarly_vs_Later:speaker-Host+Guest:place-Retro+Alveo:context-Round+Unrou	367.47	[-1088.78, 1823.71]	0.50	.6197
phaseLate_vs_Extended:speaker-Host+Guest:place-Retro+Alveo:context-Round+Unrou	2485.18	[837.72, 4132.64]	2.97	.0033