Proceedings of the Virtual Transnational Linguistic Project on

Speech Production and Perception and its Cognitive Processing

2023

Cross-Linguistic Imitation: English - Turkish

Pelageina Nadia

[1]*Institute of Linguistics, University of Cologne, Germany*

npelagei@smail.uni-koeln.de

Abstract

This study delves into the intriguing phenomenon of cross-linguistic imitation between English and Turkish, investigating whether proficient imitation skills in one's native language (L1) can be transferred to a second language (L2). Specifically focusing on the imitation of Turkish sounds by English speakers, our research seeks to explore the extent to which native speakers of English can accurately mimic the phonetic features of Turkish. Through a series of controlled experiments and linguistic analyses, we aim to uncover insights into the potential transferability of imitation skills across languages. Additionally, we address two key questions: (1) Is the degree of imitation a person exhibits consistent across languages? and (2) Do "high imitators" have an advantage in phonetic and phonological second language acquisition? Our findings hold implications for theories of language acquisition and bilingualism, shedding light on the interplay between imitation in L1 and its impact on L2 phonology acquisition.

Keywords: speech perception, speech production, accent imitation, L2 phonology acquisition.

1. Introduction

1.1 L2 phonology acquisition

The acquisition of second language (L2) phonology and the nuanced art of accent imitation have long been recognized as formidable challenges within the realm of L2 learning, particularly among older learners. Extensive research underscores the complex interplay of multifaceted factors that influence an individual's ability to master these skills. Among the myriad determinants, age, motivation, the quality of input, and exposure to authentic spoken language emerge as pivotal influencers (Moyer, 2013; 2004; 1999). Furthermore, the intrinsic awareness of L2 phonology and the persistent influence of one's native language exert significant impact.

Delving deeper, the intricate tapestry of L2 phonological acquisition reveals the profound role played by innate talent, the capacity for mimicry, and musical aptitude.

Moyer (2014) sheds light on the multifarious ways in which these factors intersect to shape an individual's journey towards accent emulation. Indeed, the ability to accurately mimic sounds, intonations, and speech patterns has emerged as a hallmark attribute among successful L2 learners. This mimicking talent, often heralded as a distinct advantage, holds the potential to bridge the gap between linguistic proficiency and authentic communication. It serves as a conduit through which learners can immerse themselves in the intricacies of a new language, unlocking the nuances of pronunciation and intonation that imbue speech with cultural authenticity.

In essence, while the path to mastering L2 phonology and accent imitation may be fraught with challenges, the cultivation of mimicking talent emerges as a potent ally in this endeavor.

Through diligent practice, attentive observation, and a willingness to embrace the intricacies of linguistic expression, learners can harness the power of mimicry to navigate the complexities of accent emulation and forge meaningful connections in their linguistic journey.

1.2 Phonetic imitation

Phonetic imitation has been a focal point of linguistic inquiry, particularly within single-language contexts, where it has illuminated consistent behavioral patterns among speakers, as highlighted by Wade et al. (2021). However, the exploration of cross-linguistic imitation remains a relatively underexplored domain within the field of linguistics. Coles-Harris (2017) provides a comprehensive synthesis of existing literature on imitation, offering valuable insights into the motivational underpinnings of this phenomenon. Central to Coles-Harris's analysis are two overarching perspectives: one positing imitation as solely driven by innate mechanisms, while the other suggests its susceptibility to extrinsic social dynamics.

In alignment with Coles-Harris's discourse, our perspective leans towards viewing imitation predominantly as an automatic cognitive process, albeit one that is intricately intertwined with social influences (Jiang & Kennison, 2022; Zając & Rojczyk, 2014). While the automaticity of imitation implies a direct link between perception and production, the acknowledgment of social factors introduces a layer of complexity, highlighting the dynamic interaction between individual cognition and sociocultural contexts.

This nuanced interplay underscores the potential for imitation to transcend the boundaries of language, suggesting that its underlying mechanisms may operate on a fundamental level of human communication. By recognizing the role of social influences in shaping imitation, we gain insight into the diverse ways in which linguistic behavior is molded by cultural norms and societal expectations.

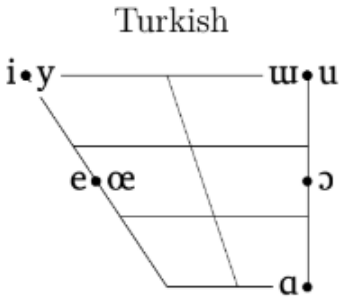
Moreover, the exploration of cross-linguistic imitation holds promises for advancing our understanding of the intricate interplay between cognitive processes, social influences, and linguistic behavior across diverse linguistic contexts. By delving into the mechanisms underlying phonetic imitation in multilingual settings, we stand to unravel the complexities of

human communication and gain deeper insights into the ways in which language is acquired, processed, and utilized in everyday interactions.

1.3 Turkish vowel system

The Turkish vowel system boasts eight distinct vowels, with six of them forming rounded-unrounded pairs (Sabev, 2019). Rounding serves as a vital cue in Turkish phonetics, facilitating the differentiation between vowel pairs with similar height and backness. Additionally, rounding plays a crucial role in Turkish vowel harmony, an essential aspect of the language's phonological structure, thus significantly contributing to effective communication in Turkish.

Graph 1 – Turkish vowel chart

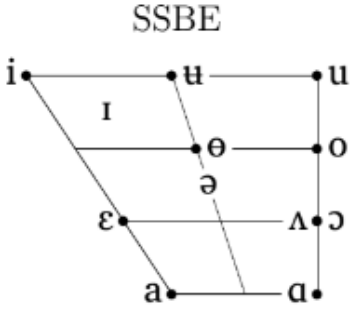


1.4 SSBE vowels system

Conversely, English, particularly Standard Southern British English (SSBE) in our case, features ten vowels that minimally utilize rounding as a distinguishing feature (de Jonge et al., 2022). Instead, English vowels primarily rely on duration and position in the vowel space for differentiation.

This stark contrast between the two languages offers a fascinating avenue for exploring cross-linguistic imitation.

Graph 2 – Standard Southern British English SSBE vowel chart.



When examining vowels shared between Turkish and English, particularly cardinal vowels, it's reasonable to expect that they will be produced with similar formant frequencies (F1 and F2), placing them approximately in the same region of the vowel space. This is especially true for vowels such as /i, ɑ, u/, and potentially including mid-vowels like /ɛ, ɔ/.

However, despite the similarities in formant frequencies, these shared vowels may still exhibit slight differences in quality. This can present a challenge for L1 English speakers, as they may instinctively categorize what they hear as a token of a familiar English vowel without adjusting their own vowel productions accordingly.

When encountering novel Turkish vowels, speakers may adopt one of two approaches. First, as these vowels are perceived as novel segments, they may be relatively easier to acquire due to their perceptual salience. Consequently, we can anticipate greater levels of imitation for these vowels.

On the other hand, there's a possibility that speakers might collapse novel Turkish vowels into the nearest equivalent English phoneme, which could impede their ability to acquire the phonetic contrasts present in Turkish. Pinpointing which English phoneme the Turkish vowels are perceived as can be challenging. For instance, research by de Jonge et al. (2022) suggests that American English speakers might categorize the Turkish vowel [œ] somewhere between the English vowels [ɛ] and [ʊ]. Similarly, it's conceivable that the Turkish vowel /y/ could be perceptually assimilated to the English vowel [u], while the vowel /ɯ/ might be spread across English vowels such as /ʊ/, /u/, or /ʌ/.

These complexities underscore the intricate process of acquiring and imitating unfamiliar phonetic elements from another language, highlighting the importance of perceptual salience and the potential for interference from existing phonetic categories in one's native language.

2. Methodology

2.1 Materials

Two sets of word lists, one in English and one in Turkish, have been compiled to prompt the pronunciation of vowels in each respective language. The English word list consists of ten words, each intended to elicit one of the ten monophthongs in Standard Southern British English (SSBE), all structured as Consonant-Vowel-Consonant (CVC). Additionally, six distractor words are included, resulting in a total of sixteen words displayed randomly on screen for participants to read.

Table 2

English word list with corresponding target vowels

Word	Vowel (IPA)
heed	/i/
hid	/ɪ/
head	/ɛ/
had	/æ/
hod	/ɔ/
hawed	/o/
hood	/ʊ/
who'd	/u/
hud	/ʌ/
heard	/ɚ/

The Turkish word list comprises eight words in CVC structure, with six of them forming minimal pairs in terms of rounding. Monosyllabic words were chosen to avoid complications related to Turkish vowel harmony and to align with the structure of the English word list. Eight distractor words are also included, maintaining consistency with the English list.

Table 3

Turkish word list with corresponding target vowels

Word (Orthography)	Word (IPA)	Vowel (IPA)
diş	/diʃ/	/i/
düş	/dyʃ/	/y/
deş	/deʃ/	/e/
döş	/dœʃ/	/œ/
kaş	/kaʃ/	/a/
dış	/dɯʃ/	/ɯ/
duş	/duʃ/	/u/
koş	/koʃ/	/o/

2.2 Participants

Monolingual English Speakers: Between 40 to 60 monolingual speakers of Standard Southern British English (SSBE), aged over 18, will participate. The only prerequisite for participation is no prior exposure to Turkish, whether formally or informally. Any exposure to or experience with other languages will be noted in a background questionnaire to assess potential effects on results.

Model Talkers:

- For the first phase, a single speaker of American/Canadian English provides stimuli to test imitation levels in English.

- For the second phase, four native Turkish speakers (two female, two male) are enlisted. Efforts are made to recruit

speakers from a single dialectal area to mitigate influences from dialectal variations. Multiple model talkers are employed in this phase to encourage imitation of general Turkish features rather than the speech of a specific individual. Each model talker reads the entire word list, with the words evenly divided among them for the shadowing portion to ensure equal exposure to all model talkers.

2.3. Procedure

The main experiment is designed to investigate cross-linguistic imitation across two phases. In the initial phase, participants engage in a shadowing task conducted in their native language, English. During this phase, participants listen to recordings of an English model talker uttering individual words and then proceed to repeat each word aloud. This process is repeated across two blocks, ensuring three repetitions for each word. The purpose of this phase is to gauge participants' proficiency in imitating English vowel sounds.

Following Phase 1, participants are categorized into either high imitators or low imitators based on their performance. High imitators are identified as those who exhibit a success rate exceeding 60%, indicating clear imitation in the majority of instances. Conversely, participants failing to meet this criterion are grouped as low imitators.

Phase 2 of the experiment shifts focus to the Turkish vowel system. Participants are tasked with shadowing a list of Turkish words, with each word representing one of the eight Turkish vowels. This phase replicates the structure of Phase 1, comprising two blocks of shadowing with four repetitions per word. Consequently, each participant completes a total of 64 shadowing trials, encompassing eight repetitions for each Turkish vowel.

Unlike Phase 1, participants are not required to initially read aloud the Turkish word list. Instead, they solely engage in the shadowing task, listening to the Turkish model talker and mimicking the pronunciation without prior verbalization. This adjustment aims to assess participants' ability to imitate Turkish vowel sounds directly, without the influence of initial reading or pronunciation attempts.

By implementing these two phases, the experiment aims to elucidate the extent to which phonetic imitation transcends language boundaries, particularly focusing on the adaptation of vowel sounds between English and Turkish. The meticulous design ensures comprehensive examination of participants' imitative abilities across both their native and a foreign language, providing valuable insights into the mechanisms underlying cross-linguistic phonetic imitation.

3. Expected findings

Based on prior research findings, we anticipate several key outcomes regarding the level of phonetic imitation exhibited by participants in their native language (L1) and its transfer to their second language (L2). First, building upon the work of Lewandowski & Jilka (2019) and Wade et al. (2021), we expect a strong correlation between individuals' imitative proficiency in their L1 and their ability to imitate phonetic features in their L2. This carries implications for the ease of phonetic acquisition and suggests that individuals who excel at imitating vowels in their L1 will likely demonstrate a similar level of proficiency when imitating vowels in their L2.

Furthermore, drawing from the findings of Spinu et al. (2018), we anticipate that participants categorized as high imitators in their L1 will exhibit a noticeable advantage over low imitators when pronouncing Turkish vowels. This suggests that individuals with a heightened ability to imitate phonetic features in their L1 are more adept at transferring this skill to the imitation of vowels in a different language, such as Turkish.

However, our expectations regarding the uniformity of this advantage across all vowels are tempered by insights from Kartushina et al. (2016). Their research suggests that certain vowels, particularly those that are phonetically similar or "proximal," may show less imitation when transitioning from L1 to L2. Specifically, we anticipate that Turkish vowels such as /i, e, u/, which have counterparts in Standard Southern British English (SSBE), may demonstrate reduced levels of imitation. This could potentially lead to these proximal Turkish vowels being replaced by their L1 counterparts during production by participants, thereby hindering the accurate reproduction of vowel sounds in the L2.

Overall, we expect to observe a strong correlation between imitative proficiency in L1 and L2, with high imitators demonstrating a clear advantage in pronouncing Turkish vowels. However, the extent of this advantage may vary depending on the phonetic similarity between vowels in L1 and L2, with proximal vowels potentially showing lower levels of imitation and greater susceptibility to replacement by L1 counterparts. These anticipated findings highlight the nuanced interplay between individual phonetic abilities, language-specific phonetic systems, and the process of cross-linguistic phonetic imitation.

4. Acknowledgements

I wish to express my appreciation to the course organizers for their efforts in facilitating a conducive learning environment.
Furthermore, I am deeply grateful to Professor Doris Muecke for her insightful lectures and guidance throughout the course. Her expertise and passion for teaching have greatly enriched our understanding of the subject matter and inspired us to delve deeper into our research.

5. References

Coles-Harris, E. H. (2017). Perspectives on the motivations for phonetic convergence [e12268LNCO-0713.R1]. Language and Linguistics Compass, 11(12), e12268. https://doi.org/https://doi.org/10.1111/lnc3.12268

de Jonge, K., Maxwell, O., & Zhao, H. (2022). Learning on the field: L2 turkish vowel production by l1 american english-speaking ngos in turkey. Languages, 7. https://doi.org/10.3390/languages7040252

Jiang, F., & Kennison, S. (2022). The impact of l2 english learners' belief about an interlocutor's english proficiency on l2 phonetic accommodation. Journal of Psycholinguistic Research, 51(1), 217–234. https://doi.org/10.1007/s10936-021-09835-7

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2016). Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. Journal of Phonetics, 57, 21–39. https://doi.org/https://doi.org/10.1016/j.wocn.2016.05.001

Lewandowski, N., & Jilka, M. (2019). Phonetic convergence, language talent, personality and attention. Frontiers in Communication, 4. https://doi.org/10.3389/fcomm.2019.00018

Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation and instruction. Studies in Second Language acquisition 21: 81-108.

Moyer, A. (2013). Foreign Accent. The Phenomenon of Non-Native Speech. Cambridge University Press.

Moyer, A. (2014). Exceptional Outcomes in L2 Phonology: The Critical Factors of Learner Engagement and Self-Regulation, Applied Linguistics, Volume 35, Issue 4, Pages 418–440, https://doi.org/10.1093/applin/amu012

Moyer, A. (2004). Age, accent and experience in second language acquisition. An integrated approach to critical period inquiry. The Cromwell Press Ltd.

Sabev, M. (2019). Spectral and durational unstressed vowel reduction: An acoustic study of monolingual and bilingual speakers of bulgarian and turkish [Doctoral dissertation, University of Oxford]. University of Oxford.

Spinu, L. E., Hwang, J., & Lohmann, R. (2018). Is there a bilingual advantage in phonetic and phonological acquisition? the initial learning of word-final coronal stop realization in a novel accent of english. International Journal of Bilingualism, 22(3),350–370. https://doi.org/10.1177/1367006916681080

Wade, L., Lai, W., & Tamminga, M. (2021). The reliability of individual differences in VOT imitation [PMID: 32772791]. Language and Speech, 64(3), 576–593. https://doi.org/10.1177/0023830920947769

Zając, M., & Rojczyk, A. (2014). Imitation of english vowel duration upon exposure to native and non-native speech. Poznan Studies in Contemporary Linguistics, 50(4), 495–514. https://doi.org/doi:10.1515/psicl-2014-0025

Effects of Social Provenance in Vowel Accommodation in Parisian Varieties

Inge Akyaa[1], Olivia Caffin[2], Suyuan Dong[2], Léna Gaubert[2]

[1]*University of Cologne, Germany*
[2]*University of Sorbonne Nouvelle, France*

iakyaa@smail.uni-koeln.de, olivia.caffin@sorbonne-nouvelle.fr,
suyuan.dong@sorbonne-nouvelle.fr, lena.gaubert@sorbonne-nouvelle.fr

Abstract

This study explores the influence of social factors on vowel accommodation in Parisian varieties of French. Female native French speakers from central Paris and Saint-Denis are engaged in a collaborative map task, adopting roles as instructors and followers, and producing target vowels /a/, /i/, and /u/. Findings suggest significant vowel convergence among speakers with similar backgrounds, while speakers from different backgrounds display a combination of convergence and divergence. These results suggest that speech accommodation is a multifaceted process, reflecting both an automatic process and strategic behaviour, deeply influenced by social factors.

Keywords: Vocalic Accommodation, Socio-Phonetics, Conversational Interaction, Map Task, Multilingualism, Virtual Reality

1. Introduction

Speech accommodation is a complex and dynamic process, conditioned by the cognitive organisation of a larger language system. This phenomenon, characterised by speakers' unconscious adaptation of their acoustic-phonetic attributes to those of their interlocutor, has been extensively documented in previous studies (Babel & Bulatov, 2012; Pardo, 2022). Several approaches are used to describe the multiple directions and extent of this phenomenon, such as alignment, convergence, divergence, imitation, and mimicry. Despite the terminological differences, these approaches underscore the nature of "inter/cognitive and/or external/social mechanisms" driving changes during the communicative procedure (Pardo, 2022). Besides, this phenomenon is manifested through various acoustic characteristics, such as VOT, f0, and vowel imitation. It is also suggested that this phenomenon is not only phonetic but also deeply influenced by other task-specific and talker-specific factors, such as gender, age, personality, linguistic background, cultural identity, and role during interaction. Furthermore, the extent and nature of accommodation are subject to both short/long term (*idem.*).

The present study aims to investigate the social dimensions of speech accommodation, particularly examining the influence of cultural background among speakers of different varieties within the same language, and also the impact of interactional roles during the task on the accommodation process. This prompts following research question:

How long does the process of vowel accommodation take with regards to the direction and extent of these vowel changes in terms of formant distribution, duration, diphthongization, and reduction?

1.1. Social and Cognitive Factors

Social factors that influence vocal accommodation include social status, power dynamics, social norms, visual cues, familiarity, and social identity. Furthermore, social knowledge about the model talkers can inhibit or facilitate the accommodation process (Babel, 2009a, 2009b; Babel & Bulatov, 2012; Kim et al., 2012, Giles, 2016). Recent research suggests that an integrated approach, considering both social and cognitive factors, may offer a more comprehensive understanding of phonetic processes, as opposed to relying solely on automatic mechanisms (Babel & Bulatov, 2012; Kim et al., 2012; Pardo, 2022).

1.2. Previous Studies

The Communication Accommodation Theory (CAT) focuses on why individuals adjust their speech either to converge or diverge from their interlocutors (Giles, 2016). It suggests social factors such as the desire for social approval and likability (idem.). For example, individuals may unconsciously mimic the speech patterns of someone they perceive as having higher social status or power (Ostrand & Chordoff, 2021). Conversely, they might deliberately avoid mimicking speech patterns of groups they do not identify with (Giles, 2016). In the context of North and South Korean immigration in London, Evans et al. (2023) state that North Korean refugees accommodate closely to South Korean immigrants, especially if they left North Korea throughout their childhood. However, because of diverging identities, both communities tend to diverge during conversational speech. While studying the dialect accommodation of Cherokee groups in Western North Carolina, Anderson (1997) showed that the less isolated group (and therefore the group that was the most in contact with English-speaking and white communities) showed stronger accommodation, specifically with the monophthongal vowel [a:]. Immigration in the Parisian suburbs has created a strong multicultural and multilingual context: the region of Île-de-France has the highest proportion of individuals having a "bilingual daily life", with one child out of four being in contact with another language than French at home (Gadet 2007: 128). This observation necessitates a focus on the dynamics inherent within the multilingual context under consideration.

1.3. Typological Background

Suburban Parisian French, while prevalent in various suburbs, is also spoken in some intra-muros parts of Paris. For this study, the focus will be on speakers from Saint-Denis, the largest city in the suburbs in the northeast of Paris, known for its multiculturalism and multilingualism. This variety of French has been the subject of numerous studies aiming to

describe its phonetic characteristics (Paternostro, 2012; Candea, 2016). A notable feature of Suburban Parisian French, derived in part from what is known as "popular French" of the working class, includes consonant variations such as simplification of consonant clusters, consonantal assimilation, and palatalization (Candea, 2016). The vowels, on the other hand, are distinctively influenced by a more multilingual context, particularly by Maghrebian languages. This influence is evident in the shortening and occasional elision of vowels, leading to vowel reduction, which is a phenomenon relatively rare in French (Fagyal, 2010: 96). Gadet et al. (2017) also identify vocalic harmony and closed realisation of vowels that are supposed to be open (e.g. <mère> /mɛʁ/ becomes /meʁ/). Despite the prevalent focus on consonants and prosody in existing research, there is a gap in studies addressing vowel variations in this variety. Considering this lacuna and the fact that the cardinal vowels /a/, /i/, and /u/ are not only present in both target varieties but also embody the multicultural attributes of the Maghrebian community, our study specifically focuses on the reduction of these vowels.

1.4. Hypothesis and Predictions

1 – Interlocutors with the same socio-geographical background will display an increased likelihood of observing vowel convergence.
2 – Interlocutors originating from different socio-geographical backgrounds may display different accommodation behaviours according to their subjective representations of their personal and interlocutor identity, of the context, and the perceived linguistic power dynamics within the interaction.
3 – Participants may show different accommodative behaviours depending on their role as *Follower* or as *Instructor*.
4 – Participants with greater familiarity will show a stronger tendency towards convergence in vowel production and vice versa. Social interaction during the ice-breaking phase may favour vowel accommodation.

2. Methods

As mentioned above, we hypothesise that social factors influence vowel accommodation. To investigate this, we conduct a collaborative map task with two distinct groups:

– Group A (with introduction phase):
Upon arrival, participants of this group will engage in an initial ice-breaking session involving self-introductions about names, ages, study programs, and residential areas. Furthermore, participants are prompted to share three popular expressions from their schools. This exchange aims to build social connections, set a conversational tone among them, and activate the use of their daily language in preparing for the subsequent tasks. After the ice-breaking phase, participants will proceed to the map task, where they will produce and record vowels.
– Group B (without introduction phase):
Unlike the first group, there will be no initial ice-breaking phase. Participants will remain anonymous to each other as they proceed to the map task. The missing social bond could significantly impact their degree of accommodation.

The experiment is designed to test four conditions: Paris-Paris (P-P), Suburb-Suburb (S-S), Paris-Suburb (P-S) and Suburb-Paris (S-P). Within these four conditions, the participants of the map task will take on two roles: *Instructor* and *Follower*.

2.1. Participants

To minimise the gender effects in speech, this study engages female native French speakers, aged from 16 to 18 (cf. Candea, 2016). Half of the participants are born and raised in Paris and exhibit minimal bilingual exposure, the Parisian French variety serves as the predominant language spoken within their households. Whereas the other half are of Maghrebian origins from Saint-Denis and actively participate in multilingual interactions in their everyday lives. Beyond French, they are exposed to at least one other language.

All participants are high school students with no specialised linguistic training and exhibit normal speech and hearing capabilities.

2.2. Procedure

A school located in Saint-Denis is partnered with a central Parisian high school, where the participants are separated into Group A and Group B. Participants are equipped with high-quality microphones and virtual reality (VR) headsets, which adds mobility to the experiment setup. This may potentially broaden the accessibility of participants for this study. The passing order is randomised yet structured to ensure equal representation of each condition, following the sequence: s-p-p-s-s-p-p-s. Each participant enrols two times by playing both *Instructor* and *Follower* in turn (1-2, 2-3, …, 8-1).

Before starting the main task, participants of each group engage in a preliminary testing exercise to ensure their comprehension of the procedure. Then during the pre-task session, the dyad in role needed to read individually a staging text written in the form of a diary, containing words with target vowels, appearing on the VR screen. This activity aims to immerse participants in the map-task scenario, and the recordings serve also as a baseline for the vowel comparison.
In the main map task, participants are presented individually with a map on their VR screen: the instructor's map detailed with place names, while the other with only road names and traffic information. The objective is to identify locations and purchase items needed collaboratively within one hour. Following the map task, each participant reads another text, similar in structure to the pre-task, to reflect on the completed activity and repeat the target words. This serves to assess any phonetic shifts post-interaction.
After the main task, participants are asked to complete a questionnaire regarding their feelings and perceptions of language variations during the whole experiment.

2.3. Data Collection Overview

Primary focus of this experiment is the elicitation of target words and the collection of data on vowel accommodation depending on the specific testing conditions as mentioned above. The target words are then analysed in PRAAT with regards to VOT, f0, and vowel imitation and reduction. From this we aim to gain insights on formant distribution, duration, diphthongisation.
Secondly, data on head movements and reaction times will be gathered to gain a better understanding of the effects of social interaction on cooperation.
Thirdly, the temporal aspects are investigated to determine at which point vocal processes become observable. In addition, for Group A, we seek to examine which duration of prior social interaction has a significant effect.

3. Possible Results

The figures and tables presented here are hypothetical results that only serve to visualise the execution of the analysis based on the experiment presented here.

The acoustic analysis of the pre-task productions of the cardinal vowels /a/, /i/, and /u/ may reveal a possible alignment with the predicted vowel distribution. Such results could suggest that vowels in the Maghrebian variety of French exhibit a more centralised vocalic space compared to those in Parisian French.

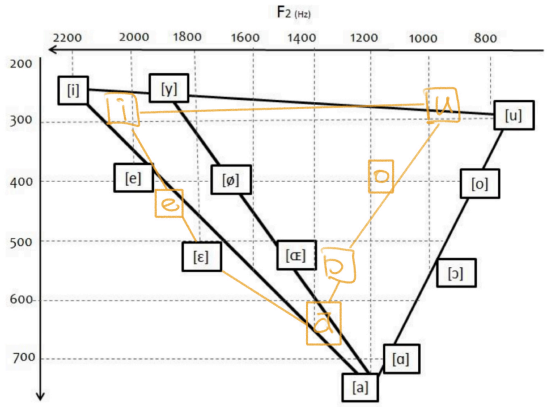


Fig. 1: *Modal French vowels vs. suburban vowels (in yellow)*

Vowels		F1	F2	F3
closed	i	308	2064	2976
	y	300	1750	2120
	u	315	764	2027
mid-closed	e	365	1961	2644
	ø	365	1417	2235
	o	383	793	2283
mid-open	ɛ	530	1718	2558
	œ	517	1391	2379
	ɔ	531	998	2399
open	a	684	1256	2503

Fig 2.: *Potential distribution in vowel formant frequencies in Hz of Suburban Parisian Variants*

When it comes to the map task, results may indicate a significant level of vowel convergence among participants sharing similar socio-geographical backgrounds (P-P and S-S dyads). In contrast, dyads of P-S and S-P participants could display a mixture of both convergence and divergence. This phenomenon may be possibly observed in both directions (P → S or S → P).

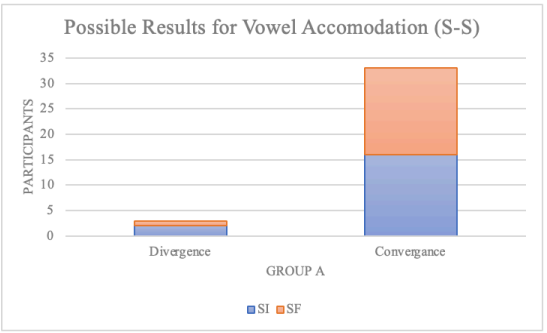


Fig. 3: *Comparison of the directions of vowel accommodation considering participants same social provenance (S-S)*

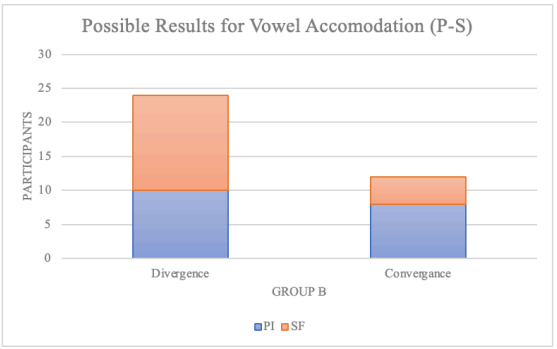


Fig. 3: *Comparison of the directions of vowel accommodation with different social provenance (P-S)*

Questionnaire responses could provide an enhanced understanding of participants' awareness of language and social identity, which potentially reveals a correlation with their degree of accommodation during the communicative interactions. However, the variation in roles (*Instructor* vs. *Follower*) within the map task may give insight into power dynamics and the influence of speaker familiarity in terms of vocalic accommodation. We assume that individual representation and personality may exert a substantial greater impact on speech patterns than the assigned roles within the interaction. Quantitative analysis of reaction times and head movements might yield insights into the impact of prior social interaction on the communicative event. Moreover, the observation of post-task accommodation may confirm the results of previous studies.

Further findings might emerge beyond the primary focus. These include potential shifts in prosody, syntax, variations in lexical choice, and the use of supralinguistic features such as clicks to express attitude. These aspects, while not central to the current study, may offer intriguing directions for future research on speech accommodation in multilingual contexts.

4. Discussion

Current debates show that world views are becoming more and more extreme. This raises the question of how to bring divided societies closer together again. Multicultural and multilingual realities harbour an opportunity for dialogue: social interaction has a significant impact on accommodation, hence why a lively, multilingual context can be useful for mutual understanding. In fact, the ability to communicate transculturally is crucial to a pluralistic society.
In Germany, the image of urban places with a high density of people with a migrant provenance often overlaps with the negative association of "sozialer Brennpunkt". In anglophone

countries, the positively connoted term *melting pot* seems to have recognised the potential of such linguistically and culturally diverse places. Multilingualism also has a positive effect on cognitive flexibility and fosters intercultural sensitivity (Gadet et al. 2017;Ikizier & Ramírez-Esparza, 2018). The question here is, to what extent multilingualism can be implemented institutionally? For example, bi-/multilingual school programs in which the languages that students speak at home can also be taken into greater consideration to foster those resources in a more and more globalised world.

Language wields power, which is also shown historically in France by the close interdependence between language, social status, and geographical affiliation. This power dynamic is also contrasted between "Parler Parisien," regarded as the linguistic standard and prestigious, and "Parler de Banlieue," often stigmatised (Candea, 2016). Integration is frequently expected of minority groups, but it inherently implies a power imbalance. It necessitates adaptation on their part. Whereas inclusion recognises this imbalance and emphasises the acknowledgement and value of diversity. Another aspect that should not be neglected is that multilingualism may have a beneficial effect on social mobility and could help individuals to adapt in a globalised world.

5. Conclusion

This study presented here may provide insights into the complex dynamics of speech accommodation within multicultural urban settings, emphasising the role of social factors in phonetic variations in popular French, focusing on the cardinal vowels /a/, /i/, and /u/ among speakers from Paris and Saint-Denis. The preliminary findings may indicate that cultural identity and multilingualism significantly influence speech accommodation. Such a study could make a useful contribution to the field of socio phonetics. This raises the question of whether there are similar patterns in other multicultural cities around the world, allowing for a cross-linguistic approach. Additionally, leveraging modern technologies such as Virtual Reality could standardise the experiment setup, making it scalable at a low threshold. This approach may have the potential to broaden the pool of participants, whereas the medium of VR headset may also have a positive effect on motivation.

A follow-up study with a shadowed AXB task could explore the perceptual boundaries of these vocalic changes. This would help us understand the speech accommodation not just from a production point of view but also its perceptual aspects. Furthermore, an adaptation of the method with an Electromagnetic articulography (EMA) map task could allow visualisation of the position and movements of the articulators in the vocal tract during vowel production. This approach could provide new insights into when and how speakers accommodate their speech in real time.

6. References

Anderson, Bridget L. (1997): Adaptive sociophonetic strategies and dialect accommodation: /ay/ monophthongization in Cherokee English. U. *Penn Working Papers in Linguistics.* Volume 4.1. Pennsylvania, USA.

Babel, Molly (2009): Selective Vowel Imitation in Spontaneous Phonetic Accommodation. *UC Berkeley PhonLab Annual Report* 5(5) doi:10.5070/P76c95w485.

Babel, Molly & Dasha Bulatov (2012): The Role of Fundamental Frequency in Phonetic Accommodation. *Language and Speech*. SAGE Publications Ltd. 55(2). 231–248. doi:10.1177/0023830911417695.

Candea, Maria (2016): L'accent dit „de banlieue", une mode ? Étude auprès de lycéens en ZEP inscrits dans une dynamique de réussite scolaire. Gilles Siouffi (ed.), *Modes langagières dans l'histoire*. Honoré Champion.

Candea, Maria (2018): Les parlers jeunes dans l'Île-de-France multiculturelle Paris, Ophrys, 2017, 174 p. Gadet, F. (ed.) *Langage et société*. Éditions de la Maison des sciences de l'homme. 163(1). 192–195. doi:10.3917/ls.163.0192. Paris.

Evans, Bronwen G., Rakgyun Shin & Justin J. H. Lo (2023): *Being North Korean in Koreatown : The role of identity & experience in accommodation in the London North Korean community.* Guarant International.

Fagyal, Zsuzsanna (2010): Rhythm Types and the Speech of Working-Class Youth in a Banlieue of Paris: The Role of Vowel Elision and Devoicing. De Gruyter Mouton. doi:10.1515/9781934078068.1.91.

Giles, Howard (2016): *Communication Accommodation Theory.* University of California, Santa Barbara, USA. doi:10.1002/9781118766804.wbiect056.

Ikizer, Elif G. & Nairán Ramírez-Esparza (2018): Bilinguals' social flexibility. *Bilingualism: Language and Cognition* 21(5). 957–969. De Gruyter Mouton. doi:10.1017/S1366728917000414.

Kim, Midam, William S. Horton & Ann R. Bradlow (2011): Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology* 2(1). De Gruyter Mouton. doi:10.1515/labphon.2011.004.

Ostrand, Rachel & Eleanor Chodroff (2021): It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. *Journal of Phonetics* 88. 101074. doi:10.1016/j.wocn.2021.101074.

Pardo, Jennifer S., Elisa Pellegrino, Volker Dellwo & Bernd Möbius (2022): Special issue: Vocal accommodation in speech communication. *Journal of Phonetics* 95. 101196. doi:10.1016/j.wocn.2022.101196.

Paternostro, Roberto (2012): Aspects phonétiques du "français parisien multiculturel" : innovation, créativité, métissage(s). *Cahiers de l'AFLS* 17(2). 32.

Socio-geographical Impact in Vowel Accommodation: Parisian and Suburban Varieties

Olivia Caffin

University of Sorbonne Nouvelle, France
olivia.caffin@sorbonne-nouvelle.fr

Abstract

This study explores the influence of social factors linked to multiculturality on vowel accommodation in Parisian varieties of French. Eight female native French speakers from Paris and Saint-Denis engaged in quasi-spontaneous task-oriented interactions elicited through a map-task, in which they were given roles of instructors and followers. The productions of vowels before, during and after the task were analyzed. Results suggest significant vowel convergence among speakers with similar backgrounds, while speakers from different backgrounds displayed a combination of convergence and divergence. These results suggest that speech accommodation is a multifaceted and complex process, reflecting a hybrid procedure combining automatic processing and strategic behavior influenced by social factors.

Keywords: Vowel Accommodation, Sociophonetics, French Varieties, Conversational Interaction, Map-Task, Multicultural Context

1. Introduction

Speech accommodation, a complex and dynamic process, is intertwined with the cognitive organization of a broader language system. This phenomenon, characterized by the subconscious adjustment of phonetic attributes by speakers to align with those of their interlocutors, has been extensively documented in previous studies (Babel & Bulatov, 2012; Pardo, 2022). Several approaches are used to describe the multiple directions and extent of this phenomenon, such as alignment, convergence, divergence, imitation, and mimicry. Despite the terminological differences, these approaches underscore the nature of "inter/cognitive and/or external/social mechanisms" driving changes during the communicative procedure (Pardo, 2022). Besides, this phenomenon is manifested through various acoustic characteristics, such as VOT, f0, and vowel imitation. Every segmental and supra-segmental cues are subject to accommodation, and it can go even further through lexical and syntactical accommodation. It is also suggested that this phenomenon is not only phonetic but also deeply influenced by other task-specific and talker-specific factors, such as gender, age, personality, linguistic background, cultural identity, and role during interaction. Furthermore, speakers do not imitate every acoustic phonetic attribute, while convergence can happen with some acoustic dimension, divergence can occur with another one.

The present study aims to delve into the social dimensions of speech accommodation, particularly examining the influence of cultural background among speakers of different varieties within the same language. Additionally, the investigation seeks to explore the influence of interactional roles assumed during communicative tasks on the accommodation process.

1.1. Social factors

The dynamics of vocal accommodation within social contexts are shaped by several influential factors, contributing to a complex understanding of this communicative phenomenon. A comprehensive exploration reveals social status, power dynamics, social norms, visual cues, familiarity, and social identity all play pivotal roles in the course of accommodation. Furthermore, social knowledge about the model talkers can inhibit or facilitate the accommodation process (Babel, 2009a, 2009b; Babel & Bulatov, 2012; Kim et al., 2012). Individuals may unconsciously mimic the speech patterns of those they perceive as holding higher social status or power, as highlighted in the work of Ostrand & Chordoff (2021). For example, within the context of North and South Korean immigration in London, Evans (2023) states that North Korean refugees often align their speech with that of South Korean immigrants, particularly if their migration occurred during childhood. However, the divergence in identities of these communities becomes palpable during conversational exchanges in which their speech tends to diverge from one another. Examining linguistic behaviors in different situations, Candea (2016) shed light on the variability of highschoolers' speech. A noteworthy observation emerges as one teenager from the Parisian suburbs recorded during a debate with his teachers transitioned from standard French to a more Suburban-oriented speech when personal topics arose. This variability in speech and accommodation hints at its nuanced nature, dispelling the notion that it is a purely automatic process. Instead, it suggests that personal identities have an impact on the degree and nature of accommodation.

1.2. French and multiculturalism

Since the 19th century, France has evolved into a significant epicenter of immigration within Europe (Gadet et al., 2017). Paris, being the capital city, stands out as a pivotal point of cultural confluence and migration, new-comers gathered in the more popular parts of the area, that being the Parisian suburbs. This immigration has engendered a strong multicultural and multilingual context: it is estimated that a quarter to a third of the French living in France can trace their ancestry to immigrant origins. Furthermore, 17% of the population of the

region of Île-de-France comes from immigration (Gadet et al., 2017) this region also has the highest proportion of individuals having a "bilingual daily life", with one child out of four being in contact with another language than French at home (Hansen 2023: 94). The manifestation of bilingualism, however, exhibits a spectrum of variation, often manifesting in a more passive form predominantly used within familial contexts (Paternostro, 2012). The heritage language, as a means of cultural identity, assumes a nuanced and sometimes paradoxical character: oscillations between being a source of both shame and pride, depending upon various factors, including the interlocutor and the relative linguistic prestige. The 'suburban' speech, alternately denominated as 'young's speech', 'street language' or 'ghetto speech', is subjected to stigmatized designations. Hambye and Paternostro, with the latter introducing the term 'Parisian multicultural accent', reject these pejorative classifications, trying to accentuate the linguistic richness instead. Those designations cited earlier are commonly linked to specific areas and considered as a social marker, and are stereotyped as not being 'real' French. This perception of a linguistic gap creates a socio-geographical gap as well (Hambye, 2008). Those variations are rejected from the standardized French norm, this rejection being known and acknowledged by the user's of these specific variations. Hambye's (2008:34) interaction with a teenager, encapsulates this sentiment, the interlocutor characterizing the French he speaks as not "very very French"[1], attributing his way of speaking to the streets. The observations related to the Suburban French and its use necessitates a focus on the dynamics inherent within the multilingual context under consideration.

1.3. Typological background

Suburban Parisian French, while prevalent in various suburbs, is also spoken in some intra-muros parts of Paris. This investigation directs its focus toward speakers from Saint-Denis, the largest city in the northeast suburbs of PAris; renowned for its multiculturalism.This variety of French has been the subject of numerous studies aiming to describe its phonetic characteristics (Paternostro, 2012; Candea, 2016; Hansen, 2023). A distinctive feature of Suburban Parisian French, derived in part from what is known as "popular French" of the working class, includes consonant variations such as simplification of consonantal clusters (eg. <il y a> /ilja/ becomes /ja/, <peut-être> /pøtɛtʁ/ becomes /pøtɛt/), consonantal assimilation (eg. <je sais pas> /ʒøsɛpa/ becomes /ʃepa/, and palatalization (eg. <tu dis> /tydi/ becomes /tydʒi/). This last variation is argued to be more salient in North African immigrant descent's speech, suggesting a variation not only connected to popular French but also to the influence of language contacts, albeit lacking consensus in existing research (Hansen, 2023). The vowels, on the other hand, are distinctively influenced by a more multicultural context, particularly by Maghrebian languages (Gadet et al., 2007). This influence materializes in the shortening and occasional elision of vowels (other than the /ə/, often eluded in French's casual speech), leading to vowel reduction - a phenomenon relatively uncommon in French (Fagyal, 2010: 96). This could be influenced by the syllable structures present in Arabic, constructed around consonants (Hansen, 2023). The figure 1 compares the vowel trapezoid of standard French and Maghrebi Arabic[2]. Glottal stops can occur at the beginning of a vowel starting word, which is something rare in standard French. This phonetic feature is related to language contact by Gadet et al. (2017) and Hansen (2023). Gadet et al. (2017) also identify vocalic harmony (eg. <aujourd'hui> /oʒuʁdɥi/ becomes /ɔʒɔʁdɥi/) and closed realization of vowels that are supposed to be open (e.g. <mère> /mɛʁ/ becomes /meʁ/). The /a/ vowel being articulated less forward is either described as being related to popular French or Multicultural Suburban French, depending on the studies (Hansen, 2015). Despite an extensive focus on consonants and prosody in research, a notable lacuna exists in studies addressing vowel variations within this linguistic variety. Considering this gap and the fact that the cardinal vowels /a/, /i/, and /u/ are not only present in both target varieties but also embody the multicultural attributes of the Maghrebian community, our study focuses on the reduction of these vowels, and the variability of the centroid (in terms of heightness) vowels /y/, /e/- /ɛ/, /œ/-/ø/, /o/-/ɔ/, not phonologically present in Maghrebi Arabic.

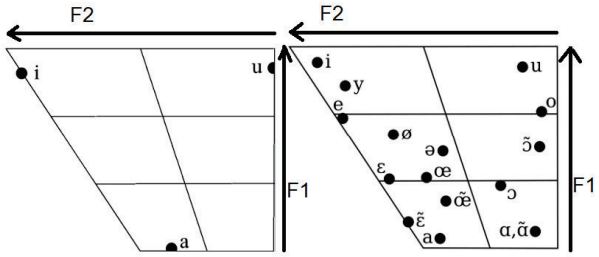


Figure 1. The vowel trapezoid of Maghrebi Arabic (on the left) and French (on the right). Because of the few phonological vowels present in Arabic, they are quite variable.

1.4. Hypothesis and predictions

1 – **Vowel convergence within similar socio-geographical backgrounds**: Interlocutors sharing a common socio-geographical background are hypothesized to exhibit an increased likelihood of vowel convergence.
2 – **Diverse accommodation behaviors across socio-geographical backgrounds**: Interlocutors with divergent socio-geographical origins may display varying accommodation behaviors influenced by personality traits, subjective representations of personal and interlocutor identities, contextual dynamics, and perceived linguistic power dynamics within the interaction.
3 – **Role-based accommodative behaviors**: Participants might showcase distinct accommodative behaviors contingent upon their roles as Followers or Instructors, aligning with insights from Pardo et al. (2010), where assigned roles of participants impacted observed convergence.
4 – **Listeners' identification**: Listeners may find more difficulties in designating a speech as being produced by a Parisian or a Suburban after an interaction between those two types of participants.

2. First experiment: map-task

The experiment is designed to test four conditions: Paris Instructor-Paris Follower (PP), Suburb Instructor-suburb Follower (SS), Paris Instructor-Suburb Follower (PS) and Suburb Instructor-Paris Follower (SP). Participants assumed two roles in the map task: Instructor and Follower.

[1] "[...] lui et ses amis ne parlaient pas un français "très très français [...]"

[2] Maghrebi Arabic, also called 'darija'is the vernacular spoken in Maghreb, forming an Arabic varieties continuum. Since the mutual intelligibility is high and the phonology is similar between these varieties, only one vowel trapezoid was used as an illustration.

2.1. Participants

The speakers recruited for the experiment were unacquainted with each other. To minimize the gender effects in speech, this study engaged eight female native French speakers, aged 16 to 18. This age range has been chosen because teenagers are usually more innovative in their speech, bringing changes in the spoken language (Gadet et al., 2017). Four participants were born and raised in Paris, while the other four are of Maghrebian origins from Saint-Denis. All participants were high school students with no specialized linguistic training and exhibited normal speech and hearing capabilities. They were paid 15 euros per hour for their participation.

2.2. Procedure

The experiment was conducted in a sound-insulated room in the Phonetic and Phonology Laboratory in Paris. Participants were seated facing one another, equipped with AKG 520 microphones and a computer monitor on the desk to view instructions. The passing order was randomized yet structured to ensure equal representation of each condition, following the sequence: s-p-p-s-s-p-p-s. Each participant enrolled two times by playing both *Instructor* and *Follower* in turn, but not with the same interlocutor (1-2, 2-3, …, 8-1). A total of 8 hours of data was collected.

Upon arrival, participants participated in an initial ice-breaking session, which encompassed comprehensive self-introductions involving details such as names, ages, study programs, and residential areas. This exchange served the dual purpose of fostering interpersonal connections among participants, establishing a convivial conversational atmosphere, and activating the use of their daily language in anticipation of the subsequent tasks.

Before starting the main task, participants engaged in a preliminary testing exercise to ensure their understanding of the procedural aspects of the study. Then during the pre-task session, each participant was recorded while reading out-loud a staging text presented in a diary format, containing words with target vowels, appearing on the computer screen. The diary consisted of sentences regarding the map-task and the places they had to go to in the scenario: eg. “aujourd’hui, je dois aller à l’épicerie pour acheter du lait, elle se trouve dans la rue de l’égalité, proche du parc aux roses [...][3]”. This activity aimed to immerse participants in the map-task scenario, and the recordings serve also as a baseline for the vowel comparison.

In the main map-task, participants were presented individually with a map on their computer: the instructor’s version detailed place names like the grocery store, the church, etc., while the other exclusively featured road names and traffic information.

The objective was to collaboratively identify locations in order to procure specified items listed in the pre-task, within a prescribed one-hour timeframe. This lapse of time was chosen in order to let the participants be more comfortable with each other the more time passed by, and the limited time gave a sense of challenge in order to entice them.

Following the map task, each participant individually read another text mirroring the structure of the pre-task material, eg: “aujourd’hui, je suis bien allée à l’épicerie où j’ai pu trouver du lait. J’ai dû remonter la rue de l’égalité et contourner le parc [...]. This post-task engagement served the purpose of reflecting on the completed activity and repeating the target words, facilitating an assessment of any phonetic shifts post-interaction.

Participants completed a questionnaire summarizing the information conveyed during the initial self-presentation in the ice-breaking session. The questions then delved into participants’ subjective sentiments and perception regarding language variations throughout the entirety of the experiment. They answered closed questions (eg. “did you feel some kind of barrier with certain participants during the experiment?”, “were you more comfortable being the Instructor than the Follower?”, “did you notice differences in your way of speaking throughout the task?”) with answers going from “absolutely yes” to “a bit”, “not really” and “not at all”. Broader open questions were asked (eg “can you explain what you felt when communicating with the participants from the other city ?”) in order to give more choice and space for their answers. This part of the questionnaire aims to check participants’ representations concerning Parisian French, Suburban French, and their respective speakers. This also sought to uncover potential disjunctions between participants’ conceptualizations and their actual linguistic production, providing valuable insights into the interplay between perception and production in the context of language variation.

2.3. Acoustic measures

For the three sets of recordings, the vowel formants (F1, F2 and F3) of /i, y, u, e, ɛ, œ, ø, o, ɔ, a/, reflecting vowel height, advancement and rounding (Thomas, 2013) were measured at the midpoint of each vowel. Ten utterances of each target vowel were selected and the formant measures were combined in order to create a mean F1, F2, F3 of each vowel for each participant in each task. The vocalic duration was also measured, and the elisions were counted. These measures were visually inspected through spectrograms and extracted in Praat. The formant measures were normalized using the Labov technique, as used by Pardo et al., in 2016, through the vowels package for R. This technique eliminates variation caused by physiological differences, while preserving sociolinguistic and dialectal differences, permitting cross-talker comparisons (Pardo et al., 2016; Thomas, 2013).

A regression analysis was done in order to assess the relationship between the predictor variables (eg., socio-geographical context, role in the task, participant) and the dependent variable (formant measures), examining the impacts of socio-geographical factors on acoustic convergence in vowels.

3. Second experiment: Identification task

3.1. Participants

The listeners recruited will be ten French students in phonetics recruited in the Sorbonne-Nouvelle University, ages between 21 to 35 years old. They all exhibited normal speech and hearing capabilities

3.2. Procedure

In a forced-choice identification task, listeners categorized short sentences from both pre and post-tasks as either being Parisian or Suburban accents. The task involved the listeners

[3] This could be translated as “today, I have to go to the grocery store to buy milk, it is located on rue de l’égalité, near the rose park“

facing a computer, where they clicked on their chosen answer. A certainty scale, offering options “completely sure”, “quite sure” or “not really sure”, accompanied each response given.

The set of the eight pre-task samples served as a baseline, displaying the speech each participant had before any interaction. The eight post-task samples, evenly split between interactions involving participants with a similar socio-geographical background (PP, SS) and those with different backgrounds (SP, PS), aimed to explore convergence and divergence, comparing the designations of speech attributed to a participant depending on the interaction context. The sentences used contained as many vowels as possible, with centroid vowels being the main target. This design also allowed assessing whether accommodation is perceptible to individuals external to the interaction.
Following this, a comprehensive analysis was conducted using a regression model. This model is used to understand the interaction between acoustically observed convergence, listeners’ designations of accents, and the level of certainty expressed by the listeners in these choices.

4. Possible results and discussion

The acoustic analysis of the pre-task productions of the cardinal vowels /a/, /i/, and /u/ is consistent with the hypothesized vowel distribution. This result implies that vowels in the Maghrebian-influenced variety of French exhibit a more centralized vocalic space compared to those in Parisian French. This is particularly salient with the centroid vowels: the pairs /e/- /ɛ/ and /œ/-/ø/ exhibit minimal distinction, F1 being lower in the Saint-Denis participant’s speech compared to their Parisian counterparts. While the pair /o/-/ɔ/ showcases a significant difference between the two phonemes, it is not as noticeable as in the Parisian speech.
In the context of the map task, results suggest a significant level of vowel convergence among participants sharing similar socio-geographical backgrounds (P-P and S-S dyads), as expected. Notably, speech intelligibility diminished as the participants progressed through the task, with vowel targets being hypo-articulated, also displaying several vowel elisions. This phenomenon might be from either a fatigue effect or because the participants felt more comfortable after extended interaction, reducing the necessity for high degree of articulation. In contrast, dyads of P-S and S-P participants displayed a mixture of both convergence and divergence. This phenomenon was observed in both directions (P-Instructor → S-Follower or S-Instructor → P-Follower). It is interesting to note that in those cases, the participants sustained a more articulated speech throughout the map task compared to dyads with shared backgrounds. Dyads displaying the most divergence were marked by an exaggeration of their articulation, in comparison to their articulation during the pre-task. Questionnaire responses from these participants indicate a higher awareness of language identity, implying a correlation between linguistic divergence and a projection of one’s identity during the communicative interactions. Indeed, from these answers, we can speculate that the participants wanted to display their identities by furthering their pronunciation from one another. The observed variation in roles (*Instructor* vs. *Follower*) within the map task may not provide significant differences in terms of vocalic accommodation, suggesting that individual representation and personality may exert a more substantial influence on speech patterns than the assigned roles within the interaction. This is on par with Lewandowski and al.,’s research from 2019, suggesting that personality had an impact on the degree of convergence. Additionally, participants displaying more openness to the other, as shown by initiating conversations or exhibiting positive gestures such as smiling, displayed more convergence occurrences. This observation is not subjected to quantitative measurement. Although bilingualism does not exhibit a significant impact on convergence in the current context, it is noteworthy that this aspect was not the primary focus of the study. Besides, the observation of post-task accommodation mirrors the previous studies and implies the enduring nature of accommodation even after the immediate task.
Regarding the identification task, the designations of pre-task and post-task in SS and PP contexts mirror themselves, listeners are able to identify the accents heard. However, a small number of listeners wrongly designated the post-task Suburban accent as being a Parisian one, although the certainty scale was never the highest in these situations. This may be caused by the inherent nature of the task, as reading out loud tends to display more hyper-articulation traits than casual speech. The post-tasks involving SP and PS interactions showcase mixed results. The samples used after interaction holding signs of convergence are either not designated as they truly were, or correctly assigned but with a low certainty score. This is an indicator of convergence staying even after the main task, and being unconsciously perceptible by external people. The regression model shows interaction between uncertainty and samples showing convergence, displaying the perception of both accents to the listener and difficulty to designate. On the other hand, the post-task samples occurring after an interaction showing divergence are majoritarily correctly designated by the listeners. This illustrates the point made earlier concerning awareness of linguistic identity and its projection in speech.
Further findings may include potential shifts in prosody, syntax and variations in lexical choice. The examination of supralinguistic features, such as clicks, should also be explored, given their use in French. Dental clicks, as exemple, serve to express discontentment, while lateral click and labiodental inspired fricatives (known as ‘tchip’) are used in suburban varieties to convey specific attitudes. From a perceptual standpoint, Candea’s study (2016) demonstrated participants’ ability to distinguish sentences as either exhibiting a standard or suburban accent during a perception task. However, those participants struggled to pinpoint the specific acoustic cues underpinning their judgments. A better understanding of both segmental and supra-segmental acoustic cues produced in this linguistic variety could offer a better understanding of accent perception mechanisms and may offer interesting future research on speech accommodation in multilingual contexts.

5. References

Babel, Molly (2009a): Phonetic and Social Selectivity in Speech Accommodation.

Babel, Molly (2009b): Selective Vowel Imitation in Spontaneous Phonetic Accommodation. UC Berkeley PhonLab Annual Report 5(5) doi:10.5070/P76c95w485.

Babel, Molly & Dasha Bulatov (2012): The Role of Fundamental Frequency in Phonetic Accommodation. Language and Speech. SAGE Publications Ltd. 55(2). 231–248.

Candea, M. (2016). L’accent dit ‘de banlieue', une mode? Étude auprès de lycéens en ZEP inscrits dans une dynamique de réussite scolaire.

Evans, Bronwen G., Rakgyun Shin & Justin Lo (2023): Being North Korean in Koreatown: the role of identity & experience in accommodation in the London North Korean community.

Fagyal, Z. (2010). Rhythm types and the speech of working-class youth in a banlieue of Paris: the role of vowel elision and devoicing. Preston & Niedzielski (éds.), 91-132.

Gadet, F., Cappeau, P., Conein, B., Guerin, E., Moreno, A., Paternostro, R., ... & Wachs, S. (2017). Les parlers jeunes dans l'Ile-de-France multiculturelle. Éditions Ophrys.

Hambye, P. (2008). Des banlieues au ghetto: La métaphore territoriale comme principe de division du monde social. *Cahiers de sociolinguistique*, (1), 31-48.

Hansen, A. B. (2015). Sensibilité et insensibilité devant la variation phonétique: une étude perceptive sur le français de la région parisienne. *Langage & société*, (1), 45-65.

Hansen, A. B. (2023). Palatalized/affricated plosives in Paris French. A sociophonetic production-perception study of a dynamic working-class and/or language contact phenomenon among middle-class speakers. *Globe: A Journal of Language, Culture and Communication*, *15*, 93-116.

Kim, Midam, William S. Horton & Ann R. Bradlow (2011): Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. Laboratory Phonology 2(1) doi:10.1515/labphon.2011.004.

Lewandowski, N., & Jilka, M. (2019). Phonetic convergence, language talent, personality and attention. *Frontiers in Communication*, *4*, 18.

Ostrand, Rachel & Eleanor Chodroff (2021): It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. Journal of Phonetics 88. 101074. doi:10.1016/j.wocn.2021.101074.

Paternostro, R. (2012). Aspects phonétiques du "français parisien multiculturel": innovation, créativité, métissage (s). Cahiers de l'AFLS, 17(2), 32-54.

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics*, *72*(8), 2254-2264.

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, *79*, 637-659.

Pardo, Jennifer S., Elisa Pellegrino, Volker Dellwo & Bernd Möbius (2022): Special issue: Vocal accommodation in speech communication. Journal of Phonetics 95. 101196. doi:10.1016/j.wocn.2022.101196.

Thomas, E. R. (2013). Sociophonetics. The handbook of language variation and change, 108-127.

Effects of Cultural Background in Vowel Accommodation: A Case Study of Parisian French Varieties

Suyuan DONG[1], *Olivia Caffin[1], Inge Akyaa[2], Léna Gaubert[1]*

[1]*University Sorbonne Nouvelle (France)*
[2]*University of Cologne (Germany)*
suyuan.dong@sorbonne-nouvelle.fr, olivia.caffin@sorbonne-nouvelle.fr,
iakyaa@gmail.uni-koeln.de, lena.gaubert@sorbonne-nouvelle.fr

Abstract

This study explores the influence of social factors on vowel accommodation in Parisian varieties of French. Eight female native French speakers from Paris and Saint-Denis engaged in a map task, adopting roles as Instructors *and* Followers*, and producing target vowels /a/, /i/, and /u/. Findings suggest significant vowel convergence among speakers with similar backgrounds, while speakers from different backgrounds displayed a combination of convergence and divergence. These results suggest that speech accommodation is a multifaceted process, reflecting both an automatic process and strategic behavior, deeply influenced by social factors.*

Keywords: Vowel Accommodation, Socio-Phonetics, Conversational Interaction, Map Task, Multicultural Context

1. Introduction

From a sociophonetic perspective, language is considered as inherently dynamic which enables speakers to adapt to social situations in which they are exposed (Thomas, 2011:2). A key manifestation of this adaptability is speech accommodation.

Characterized as a multifaceted and dynamic process, speech accommodation is suggested to be conditioned by the cognitive organization of a broader language system. This phenomenon involves speakers' unconscious adaptation of their acoustic-phonetic attributes to those of their interlocutor in respect to the context. This adaptation has been extensively documented in previous research (Babel & Bulatov, 2012; Pardo, 2022; etc.). These studies exploit several approaches to describe its multiple directions and extent, such as alignment, convergence, divergence, imitation, and mimicry. Despite the terminological differences, these approaches emphasize the nature of "inter/cognitive and/or external/social mechanisms" that drive changes during the communicative situation (Pardo, 2022). Notably, speech accommodation is observable through diverse acoustic features, including Voice Onset Time (VOT), fundamental frequency (F0), voice quality and vowels. The extent and nature are subject to both short and long term, influenced by both the level of interactivity in communicative tasks and other factors relative to interlocutors (idem.). Thus, it is also suggested that this phenomenon is not only phonetic, but also deeply influenced by other task-specific and talker-specific factors, such as gender, age, personality, linguistic background, cultural identity, and role during interaction (e.g. information giver vs. receiver).

This study aims to delve into the social dimensions of speech accommodation, with a particular focus on exploring the influence of cultural background among speakers of different varieties within the same language. Additionally, it examines the impact of interactional roles within the communicative task on the accommodation process.

1.1. Cultural background, power dynamics and linguistic representation

A multitude of social factors are known to influence speech accommodation, including social identity, social norms, familiarity, and power dynamics.

First, previous studies indicate that self-awareness and knowledge of social identity regarding the interlocutor can either inhibit or facilitate the accommodation process (Babel, 2009a, 2009b; Babel & Bulatov, 2012; Kim et al., 2012). For instance, individuals might unconsciously mimic the speech patterns of those perceive as socially superior (Ostrand & Chordoff, 2021). Evans et al. (2023) observe that North Korean refugees in London closely accommodate to South Korean immigrants' speech. However, both communities tend to diverge in their conversational speech, likely to insist on a distinct identity. A similar case is noted among Hebrew speakers, who display a stronger tendency for divergence compared to other speakers of Indo-European languages (Weise et al., 2020). Moreover, the role assumed during interaction, particularly in contexts with power asymmetry (e.g., police interview task), has been shown to influence the direction of accommodation, often leading to increased divergence (Earnshaw, 2021).

These observations might link to the concept of "linguistic representation", wildly explored in sociolinguistics. A representation can be both a collective/shared and an individual/heterogeneous construct. Abric (1994) defines it as always being "a representation of something for someone", which is appropriated by the individual (or group), shaped by their previous experiences and the surrounding social context, reconstructed in their cognitive system, and integrated into their value system (Abric, 1994: 16-17). This concept suggests that individuals (or groups) perceive reality based on their own reference system, then assigning specific meaning to their perceptions. Consequently, these "mental images" are both long-term, shared and co-constructed by group members, and short-term shaped by the immediate context (idem.: 18). From this perspective, representations are considered not only to emerge from linguistic practices, but also partially guide the linguistic behaviors and circulate through language.

In this study, the focus is on Parisian French. This language and its varieties are particularly relevant due to the high immigration rate in this Parisian region, creating a strong multicultural and multilingual context. The region of Île-de-France, for instance, has the highest proportion of population bathing a "bilingual daily life" in France, with one in four children exposed to a language other than French at home (Gadet 2007: 128). Despite this rich linguistic diversity, there is a lack of literature addressing the linguistic dynamic within this multicultural

context which, according to us, necessitating further exploration.

1.2. Typological background: Suburban Parisian French vs. Parisian French

Suburban Parisian French, prevalent in various suburbs, is also spoken in some intra-muros areas of Paris. This study focuses on speakers from Saint-Denis, the largest administrative division in the northeastern suburbs of Paris, notable for its multiculturalism.

Numerous studies have analyzed the phonetic characteristics of this French variety (Paternostro, 2012; Candea, 2016). Partially derived from the "popular French" of the working class, this variety contains notable linguistic features addressing consonant variations such as simplification of consonantal clusters, consonantal assimilation, and palatalization. On the other hand, the vowels are more distinctly influenced by a multicultural context, particularly by Maghrebian languages. This influence manifests in vowel shortening and occasional elision, leading to vowel reduction — a phenomenon relatively rare in standard French (Fagyal, 2010: 96). Gadet et al. (2017) also noted vocalic harmony and the closer realization of open vowels.

Existing studies majorly focus on consonants and prosody, leaving a research gap regarding vowel variations in this variety. Given this gap and considering the multilinguistic attributes of Maghrebian community reflected in cardinal vowels /a/, /i/, and /u/, our study specifically focuses on the centralization of these vowels.

1.3. Hypothesis and predictions

The primary objective of this study is to determine how socio-geographical-cultural background within the same language influences vowel accommodation during interactive communication. According to Giles' Communication Accommodation Theory (CAT), the two main directions of accommodation—convergence and divergence — represent expressions of social closeness or distance among interlocutors (Soliz & Giles, 2014). While convergence is generally considered as the unmarked pattern, divergence may occur when speakers aim to increase social distance or counteract an exaggerated speech pattern of their interlocutor, potentially trying to induce convergence.

Based on these observations, the hypothesis for this study are as follows:

1) Interlocutors sharing the same socio-geographical background will likely exhibit a high propensity for vowel convergence, indicating increased social closeness.

2) Conversely, interlocutors from different socio-geographical-cultural backgrounds may display varied accommodation behaviors, influenced by their subjective perceptions of personal and interlocutor identity, the context, and the perceived linguistic power dynamics within the interaction.

3) Participants' accommodative behaviors may vary depending on their role as either *Follower* or *Instructor*.

2. Methods

To address the study's objective, the experiment test four conditions: Paris-Paris (PP), Suburb-suburb (SS), Paris-Suburb (PS) and Suburb-Paris (SP). Participants performed in two roles during the map task: *Instructor* and *Follower*.

2.1. Participants and experiment setup

To control the number of variables, gender and age effects in speech is eliminate in this study. Thus, the experiment engaged eight female native French speakers, aged from 16 to 18. Four of them were born and raised in Paris, while the other four were of Maghrebian origin from Saint-Denis. All participants were high school students with no specialized linguistic training and exhibited normal speech and hearing capabilities.

The experiment was conducted in a sound-insulated room. Participants were seated facing one another, equipped with high-quality microphones and a computer monitor to view instructions. The passing order was randomized yet structured to ensure equal observations of each condition, following the sequence: S-P-P-S-S-P-P-S. Each participant engaged twice in the map-task, alternating roles as *Instructor* and *Follower*, order from n1-n2, n2-n3, …, n8-n1.

2.2. Procedure: Ice-breaking, pre-task, map-task, post-task, questionnaire

Upon arrival, an initial ice-breaking session was conducted where participants introduced themselves, providing their names, ages, study programs, and residential areas. Furthermore, they were also encouraged to share three popular expressions from their schools. This exchange aimed to bound connections and set a conversational tone among them, also to activate the use of their daily language in preparing for the subsequent tasks. Then, this section was followed by a preliminary testing exercise to ensure their comprehension of the map-task procedure.

During the pre-task session, participants who were about to enroll in the map-task needed to individually read a staging text. The text is written in a diary format with informal register, containing words with target vowels. The following sentence exemplifies a part of the text: *Fallait qu'on allie au Market Aroma pour acheter des chips, des patates, des coulis, […].* While one speaker read, the others listened to music through their headphones to prevent potential influences during this phase. This activity was aiming to immerse participants in the map-task scenario, and the recordings served as a baseline for vowel comparison (before, during, and after task).

In the main map task, each participant was given a unique map on their computer: the *Instructor*'s map detailed place and road names, while the *Follower*'s map provided only road names and traffic information. The objective was to collaboratively identify locations and purchase necessary items within a one-hour limit. Following the map task, each participant read another text, similar in structure to the pre-task text, to reflect on the completed activity and reproduce the target words. This served to assess any post-interaction phonetic shifts. After the post-task, the final stage involved participants completing a questionnaire regarding their reflections and representations on language variations experience during the whole experiment and their daily life.

3. Possible Results and Discussion

The acoustic analysis of the pre-task productions of vowels /a/, /i/, and /u/ aligns with the predicted vowel distribution. As illustrated in **Figure 1**, the Maghrebian Parisian variety demonstrates a more centralized vocalic space compared to those in Parisian French.

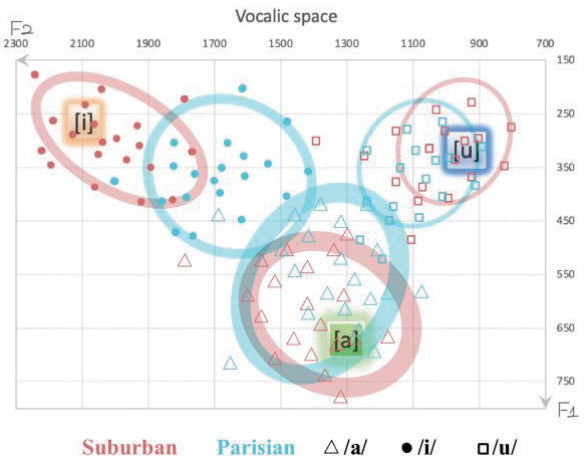


Figure 1: V*owel space illustrating a more centralized tendency in Suburban Parisian (in red) compared to Parisian French (in blue). The points, triangles and squares in the figure represent respectively the sounds /i/, /a/ and /u/, as produced by participants during the pre-task reading. The vowel /a/ seems to be more variable compared to the other two vowels. Cardinal vowels are marked with IPA symbols.*

In the map task, results may suggest significant vowel convergence among participants sharing similar socio-geographical backgrounds (P-P and S-S dyads). Conversely, dyads of P-S and S-P participants displayed a mixture of both convergence and divergence. This phenomenon was observed in both directions (P → S or S → P) and was marked by high intra-individual variation. Responses from the questionnaire indicate a heightened awareness of language identity among these participants, correlating with their divergence behavior during interactions. Additionally, a residual effect of accommodation was observed after the map-task, as illustrated in **Figure 2**.

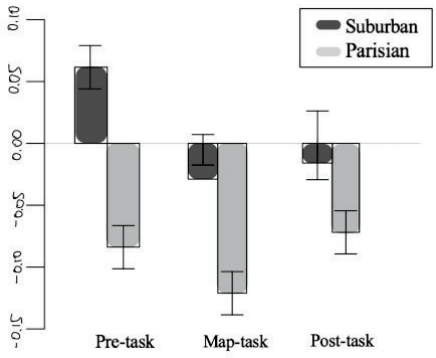


Figure 2: *This figure, inspired by Babel and Bulatov (2012: 239), illustrates the general tendency of accommodation across two linguistic communities during three phases of the experiment. The y-axis represents the vowels' formants average difference in distance, with negative values indicating convergence, and positive values indicating vowel divergence.*

However, the variation in roles (Instructor vs. Follower) within the map task may not provide significant differences in terms of vocalic accommodation. This suggests that individual representation and personality traits may exert a more substantial influence on speech patterns than the assigned roles within the interaction.

Further findings may emerge beyond the primary focus, including potential shifts in prosody, syntax, variations in lexical choice, and the use of supra-phonetic features such as clicks to express attitudes. While these aspects are not central to the current study, they may offer intriguing directions for future investigations into speech accommodation in multilingual contexts.

4. References

Abric, J. C. (1994). Pratiques sociales et représentations. Paris. Presses Universitaires de France.

Babel, M. (2009a). Phonetic and Social Selectivity in Speech Accommodation.

Babel, M. (2009b). Selective Vowel Imitation in Spontaneous Phonetic Accommodation. UC Berkeley PhonLab Annual Report 5(5).

Babel, M. & Bulatov, D. (2012). The Role of Fundamental Frequency in Phonetic Accommodation. Language and Speech. SAGE Publications Ltd. 55(2). 231–248.

Bronwen G. E., Shin R. & Lo, J. (2023). Being North Korean in Koreatown: the role of identity & experience in accommodation in the London North Korean community. ICPhS 2023, Prague.

Candea, M. (2016). L'accent dit « de banlieue », une mode ? Étude auprès de lycéens en ZEP inscrits dans une dynamique de réussite scolaire. Gilles Siouffi. Modes langagières dans l'histoire, Honoré Champion, 2016. hal-01397470

Earnshaw, K. (2021). Examining the implications of speech accommodation for forensic speaker comparison case work: A case study of the West Yorkshire FACE vowel. Journal of Phonetics, 87.

Fagyal, Z. (2010). Rhythm types and the speech of working-class youth in a banlieue of Paris: the role of vowel elision and devoicing. Preston & Niedzielski (Eds.), 91-132.

Gadet, F., et al. (2017). Les parlers jeunes dans l'Ile-de-France multiculturelle. Éditions Ophrys.

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).

Ostrand, R & Chodroff, E. (2021). It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. Journal of Phonetics 88. 101074.

Paternostro, R. (2012). Aspects phonétiques du « français parisien multiculturel »: innovation, créativité, métissage (s). Cahiers de l'AFLS, 17(2), 32-54.

Pardo, J. S., Pellegrino, E., Dellwo, V. & Möbius, B. (2022). Special issue: Vocal accommodation in speech communication. Journal of Phonetics 95. 101196.

Soliz, J., & Giles, H. (2014). Relational and identity processes in communication: A contextual and meta-analytical review of Communication Accommodation Theory. In E. Cohen (Ed.), Communication yearbook 38 (pp. 106-143). Thousand Oaks, CA: Sage.

Thomas, E. R. (2011). Sociophonetics: an introduction. New York, Palgrave Macmillan.

Weise, A., Silber-Varod, V., Lerner, A., Hischberg, J., & Levitan, R. (2020). Entrainment in spoken Hebrew dialogues. Journal of Phonetics, 83.

Exploring rhoticity perception: cross-linguistic insights from differential F3 values in Mandarin Chinese and American English

Marie Huygevelde[1],

[1]*Université Paris Cité, PARIS*

marie.huygevelde@etu-u.paris.fr

Abstract

Highlighted in Heselwood & Plug (2022), the role of F3 became subject of experimental interest to account for acoustic cues responsible for rhoticity perception. Indeed, acoustic prominence in F2 region assumed as main rhoticity percept calls for more investigation by diversifying languages and vocalic contexts. Following the proposal of an *experimental study on cross-linguistic perception of rhotic vowels preceding post-vocalic /r/ by native Mandarin and American English speakers (Huygevelde et al., 2023), this final project intends to explore further cross-linguistically* the *rhoticity percept by choosing a position showing discrepancies in terms of F2 and F3 values for the retroflex /ɹ/, i.e. prevocalic position.*
Exploiting language-specific acoustic values of the rhotic /ɹ/, the present work aims to show the importance of targeting between-languages variation and its related perception to envision more comprehensively the rhoticity perception.

Keywords: rhoticity, speech perception, cross-linguistic, low F3, Mandarin Chinese, American English

1. Introduction

If inventories of world's languages (e.g. Maddieson, 1894) provide insights on distribution of rhotics, also call "r-sound" or "r-coloured" (Clark & Yallop, 1990), debate remains in the literature on the articulatory and acoustic features that could unify the rhotics under the same natural class (Wiese, 2011). Characterized by a high variability, rhotics call for further exploratory work to clarify complementarity between acoustic and perceptual dimensions.
Aiming at contributing via an experimental study to these questions, this project intends to take advantage of cross-linguistic variation to highlight perception of rhoticity.
For this purpose, we will focus on two languages presenting both rhotics in their phonological systems, Mandarin Chinese (MC) and American English (AE) (Ladefoged & Maddieson, 1996). These languages were also proposed as study subject in a previous group proposal (Huygevelde et al., 2023). The rhotic /r/ and its acoustic correlate of a low F3 in postvocalic position were intended to be tested in production and perception, and similar identification patterns were expected. Differently, the present study relies on cross-linguistic discrepancies, instead of similarities, by testing perception of the retroflex /ɹ/ in prevocalic position showing different F2 and F3 values between the two languages (Chen & Mok, 2021). To the extent of considering Mandarin Chinese as less rhotic in this position than AE, valuable insights could emerge from exploring such variation.
On the other hand, as mentioned in Heselwood & Plug (2022) study on English perception, it is generally admitted that a low F3 constitutes the main rhoticity percept but experimental evidence remains insufficient. By attenuating alternatively the intensity of F2 and F3 in the signal, their findings revealed that a unique formant in F2 region caused by the proximity of F2-F3 was responsible for rhoticity perception. Separation with higher spectrum being crucial, attenuating F2 created the reverse effect than attenuated F3 by making the signal perceived less rhotic by American English listeners.

Aligning with their conclusion that more languages and more vocalic contexts should be tested to enrich scientific knowledge on this domain, this project investigates further rhoticity perception on a vocalic position displaying language-specific acoustic differences in formants values. Indeed, from frameworks such as Perceptual Assimilation Model (Best et al., 2001) or in experimental literature on perception of rhotics (e.g. Ingram & Park, 1998; Yamada & Tohkura, 1992), we know that phonological system of native language impacts tremendously perception but the latter can also differ from predictions. For instance, while perception of the non-phonemic contrast /l/ vs /r/ by Korean native listeners was expected to be difficult, they performed surprisingly high, hypothetically due to allophonic contrast [l] vs [ɾ] (Kochetov & Smith, 2009).

This brings us to consider variation in perception highly impacted by native language phonetics and phonology. For this reason, this study intends to explore experimentally differences in acoustic characteristics of the rhotic /ɹ/ between Mandarin Chinese (MC) and American English (AE). Does language-specific acoustic dimension impact rhoticity perception or are language-general assumptions on F2-F3 sufficient to account for the phenomenon? In other terms, does the hypothesized F-rho salience induce similarly rhoticity perception of languages having rhotics with higher F2-F3 values, such as the rhotic /ɹ/ in prevocalic position in Mandarin Chinese?

To address these questions, perception of rhoticity will be tested with native Mandarin Chinese and American English on sequences that display differential formants values between these languages, i.e. in prevocalic position. We will also account for the role of F3 in perception by including a progressive attenuation of F3 in the signal.
Unlike previous proposal predicting similar identification patterns between the two groups on continua involving rhotics in postvocalic position, this final project postulates differential perception due to differences in acoustic characteristics of the rhotic /ɹ/ in prevocalic position. Indeed, articulatory and acoustic evidence (Chen & Mok, 2021) showed an effect of language as well as syllabic position for values of F2 and F3 of /ɹ/. While it appears similar for Mandarin Chinese and American English in postvocalic position, prevocalic position revealed significant differences with much higher F2 and F3 values for Mandarin Chinese, compared to American English. As mentioned before, Mandarin /ɹ/ seems less rhotic than its English counterpart in this position.

To explore the impact of language-specific values of formant patterns, two experiments in production and perception will be conducted. Based on the assumption that coarticulation patterns in perception and production are intrinsically linked (Beddor et al., 2013), both dimensions are accounted for. In fact, production of native speakers of Mandarin Chinese and American English will allow us to define a baseline in acoustic values for the rhotic and to give better interpretation of tendencies in case of highly individual variation. Nonetheless, perception constitutes the core of our experimental investigation via an identification task assessing whether language-specific values range of the rhotic in acoustic signal induces participants differential perception.

2. Experiment 1

The experiment 1 will analyze acoustic features of /ɹ/ through production of native speakers of Mandarin Chinese (group 1) and American English (group 2) and draw a baseline for formants values to ensure congruent results with the literature (Chen & Mok, 2021; Liu, 2016)

2.1. Methods

2.1.1. Stimuli

Stimuli imply monosyllabic pseudo-words involving /ɹ/ at word-initial following by various vocalic contexts. To balance the number of vowels between languages and ensure similar weight for acoustic analysis, 4 vowels for Mandarin Chinese and 4 vowels for American English are chosen. Vocalic tokens in Chinese are /ɻ a ɤ u/ to draw parallels with Chen & Mok (2021) findings, and [ɑ] to compare with potential results from the group proposal mentioned above (Huygevelde et al., 2023). For AE, vowels gathering similar distribution in aperture degree and front-backness feature, /a ɑ u ʌ/ are selected from their phonological inventory. For the non-attested sequences whose orthography is not transparent, existing words were given to help participants identifying the pronunciation of the vowel in question, e.g. [ɹʌ] as the pronunciation of the first syllable in "Renault" [ɹʌ'noʊ].

2.1.2. Participants

Participants will be recruited to take part of both experiments at one week of interval. Thus, recruitment will proceed with respect to criteria of inclusion for both perception and production tasks.
In addition, to maintain sufficient statistical power and following the procedures found in the experiment literature in perception, 20 native speakers of Mandarin Chinese (group 1) and 20 native speakers of American English (group 2) within an age range of 18-35 years will be recruited. No restriction on gender will be applied. Group 1 will be asked to speak a rhotic variety of Northern China (including the capital Beijing), knowing for its rhoticity (Lee, 2005; Lee & Zee, 2003; Chen & Mok, 2021).
A questionnaire will be distributed to the participants to assess mainly their proficiency in the other language (Mandarin Chinese for group 2 and American English for group 1). To avoid interference with linguistic knowledge, the main criteria of exclusion would be a prolonged sojourn in the other concerned country, bilingualism and proficiency (speaking and listening) in the other language. Due to the spreading of English in education, less drastic proficiency criterion for participants in group 1 will be applied with low proficiency in English accepted.

2.1.3. Procedure

Participants production will be recorded under laboratory conditions, in a sound-proof booth. To control the recording and maintain similar settings with the procedure from the previous group proposal, participants will be asked to repeat these pseudo-words embedded in sentences displayed on a screen: "/ʈʂɤ kɤ ___ ba/ "This is ___" for group 1 and "what a ___ again" for group 2. Each stimuli should be repeated 8 times, for a total of 32 repetitions by each group.

2.2. Results

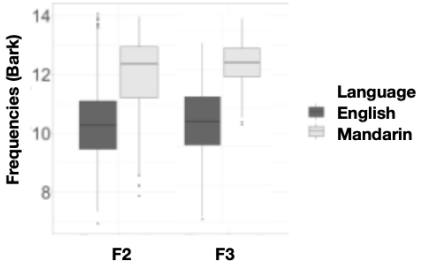


Figure 1: F2 and F3 values /ɹ/ for both groups in prevocalic position. *Adapted from Chen & Mok (2021)*

Based on the literature and Bark scale, formants values mean in Hertz for F2 and F3 in prevocalic position should be higher in Mandarin /ɹ/, about 3000Hz, than in American English that would be within a range of 2200-2500Hz (Chen & Mok, 2021; Liu, 2016). Regarding F2 values, Mandarin Chinese tokens would involve frequencies around 2500Hz and around 1800Hz for American English.
After converting the values in Bark, a statistical analysis based on a linear mixed-effect model will be applied on F2 and F3 values when the /ɹ/ sound reaches its lowest point, following Chen & Mok (2021) analysis. Taking stimuli and participants as random effects, we expect to see an effect of language with statistical significance related to the prevocalic position with higher values of F2 and F3 of Mandarin /ɹ/ than that of English. (Fig.1).

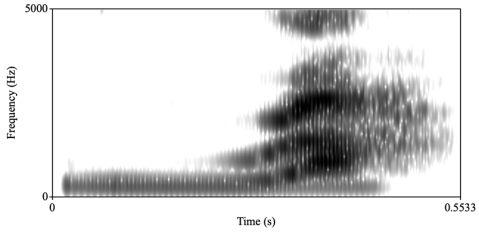


Figure 3: *Spectrogram [ɹa] sound non attenuated (personal recording, similar to potential AE token)*

3. Experiment 2

Experiment 2 will assess identification of stimuli varying in F3 along /ɹ/ (Mandarin Chinese) - /ɹ/ (American English) continua.

The identification task involves rhoticity judgments of the stimuli perceived or not as 'r-sound'. Due to various F2-F3 values in Mandarin Chinese linguistic system depending on syllabic position, differential results are expected with Mandarin listeners more sensitive to rhotics perception than English participants.

3.1. Methods

3.1.1. Stimuli

Three 11-step continua of synthesized stimuli to control maximally to the acoustic characteristics involved will be created based on production of tokens by two female speakers, respectively native speaker of American English and Mandarin Chinese. Monosyllabic pseudo-words extracted from production data (Exp. 1), averaged on duration, are selected to avoid as much as possible lexical knowledge inference. Sequences [ɪa] and [ɪɑ] take place as experimental items because of the low back vocalic segments [a] and [ɑ] proved to be r-coloured in American English (Clark & Yallop, 1990), and are included among vowels presenting a low F3 due to rhoticity (Huang, 2010) in Mandarin Chinese. Furthermore, same vocalic qualities [a] and [ɑ] chosen in the group proposal allows comparisons with its potential results as well as ensures stronger effects of F3 manipulation as shown in rhoticity judgments presented in Heselwood & Plug (2022).

In addition, to plainly focus on the category of rhotics, tokens produced in Mandarin Chinese and American English of the same rhotic are used as endpoints of the continua. Overall, the continua will display three main conditions, apart from the vocalic quality, i.e. (1) non attenuation of F3 (/ɹ/ (Chinese - /ɹ/ AE; or reverse), (2) progressive attenuation of F3 towards the Mandarin Chinese token (considered less rhotic), and (3) towards the AE token (considered more rhotic).
In fact, aiming at contributing to the research on the role of F3 in rhoticity perception, the use of another consonant for one of the endpoints of the continuum such as /l/ is avoided. It would imply other variables to control for, such as the locus of F3, the onset frequency of F2 and vocalic abruptness in raising F1 (Ingram & Park, 1998; Yamada & Tohkura, 1992).

Continua synthesized steps will be generated with the Klatt cascade formant synthesizer (Klatt, 1980) varying values F2 and F3 altogether.
Moreover, based on the literature, formants values for F2 and F3 in prevocalic position should be higher in Mandarin /ɹ/, about 3000Hz, than in American English that would be within a range of 2200-2500Hz (Chen & Mok, 2021; Liu, 2016). Regarding F2 values, Mandarin Chinese token would involve frequencies around 2500Hz and around 1800Hz for American English.
As highlighted in the Experiment 1 results section, F3 values for group 1 should be around 3000Hz, and around 2300 Hz for group 2. We hypothesized maximal values around 3400 Hz for MC tokens and 2300 for AE tokens, that will be used as endpoints. Eleven steps of regular increase or decrease in frequencies values will be generated, about 100Hz at each step between the Mandarin /ɹ/ considered as less rhotic in prevocalic position (Chen & Mok, 2021) and the American English /ɹ/. Regarding the attenuated continua, taking into account data from production experiment, intensity of F3 will vary at each step of 1.5dB using *SpeechStation 2* as filtering method to cover reasonably the 11 steps of the continua, in comparison with the 3dB for 8 steps in Heselwood & Plug (2022). This attenuation will decrease for the two continua: towards AE token (lower /ɹ/ values) and towards the MC token (higher /ɹ/ values).

3.1.2. Participants and Procedure

Same participants from the first experiment will be asked to perform the perception experiment a week later. Overall, we should obtain results from 20 speakers of Mandarin Chinese (group 1) and 20 speakers of American English (group 2) within an age range of 18-35 years. Both groups will encounter all three conditions of stimuli that will be randomized as well as items. Fillers of monosyllabic sequences involving the bilabial /b/ and the nasal /m/ associated with vowels /a, i, u/, present in the phonology of both languages, will be added to the task.

Participants will sit in front of a computer in a sound-proof booth, equipped with headphones such as AKG K 271 MK II. The identification task involves labelling stimuli representing each step of the 11-continua by choosing whether the sound is "r" or "not r". Different words will be heard in a familiarization phase to exemplify what we call "r" sound (e.g., [ɹɛd] for *red*). During the task, participants in front of a computer will be asked to press the corresponding key r for "r" or l for "not r", keys adapted for this task since distanced on the keyboard. Including the endpoints of the continua, participants will be tested on 13 stimuli for each condition (non attenuated, attenuated towards low values, attenuated towards high values). Repeated 3 times, the task will consist of 117 experimental trials (13 stimuli x 3 conditions x 3 continua) with supplementary 2/3 of filling items.

4. Predicted Results and Discussion

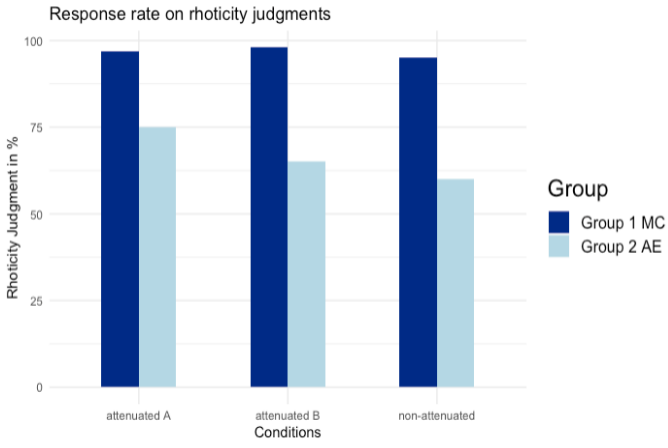


Figure 2: *Predicted overall response rate for both groups on rhoticity judgments across 3 conditions (from left to right: attenuated A (towards AE token), attenuated B(towards MC token) and non-attenuated).*

Grounded on the findings by Heselwood & Plug, we expect more 'rhotic' judgments the more F3 is attenuated. However, we postulate that language-dependent discrepancies in F2 and F3 values would impact rhoticity perception. Indeed, language-specific preferences might be triggered facing more prototypical values associated with the vocalic position tested, i.e. prevocalic position. If we could expect similar identification

patterns for postvocalic position, here, the prevocalic position might reveal a distinctive distribution in responses because of significant differences of F2 and F3 values between Mandarin Chinese and American English. Furthermore, more variation encountered by speakers of Mandarin Chinese in values depending on the syllabic contexts might characterize a more spread rhotic area in perceptual space.
Overall, we expect that attenuation of F3 produces a salience in rhoticity perception for both groups. Thus, among the three conditions, less rhoticity judgments are expected under non-attenuated condition. (Fig.3)

As hypothesized, for the group 1 (MC), more flexibility in rhoticity judgments is foreseen due to discrepancies in F2 and F3 values depending on the syllabic position and more diversity in acoustic signal characterizing rhotics, as mentioned above. Facing this variability, Mandarin Chinese listeners are expected to consider stimuli of the whole non attenuated continuum as rhotic even if individual preferences might emerge for a certain range of values. Similar tendencies should be found for the two other continua, with overall a high response rate towards 'rhotics'. Thus, their judgments should be more 'rhotic' than American English listeners, around 95-98%.

In comparison, for the group 2 (AE), more rhoticity should be assessed in the non attenuated condition towards tokens falling under a more common prototypical range of formant values (condition attenuated A) for rhotics in American English, i.e. less than 2700 Hz for F3 and less than 2100Hz for F2. Among the other conditions, a higher percentage in rhoticity judgments is expected for group 2 under the condition with great F3 attenuation on more AE-like tokens, at around 75%. For the attenuation on more MC-like tokens, even if the values are supposed non-prototypical, we postulate nonetheless an enhancement of rhoticity percept in comparison with non-attenuated condition so attenuated F3 should trigger more "r" responses.

Overall, we expect a strengthening in rhoticity percept due to attenuation of F3, aligning with accounts by Heselwood & Plug. However we postulate an effect of language-specific phonetic characteristics that might impact the perceptual space of rhoticity and thus the judgments associated to its identification. This assumption brings us to hypothesize that Chinese Mandarin over American English listeners might be accustomed to more variation in the acoustic signal when it comes to rhotics, and this fact should be reflected in our results.

5. Conclusion

Under general terms, this project intended to contribute to the research related to the class of rhotics and the definition of their common features. By diversifying the sequences via the prevocalic position and the languages tested, this study aimed to clarify the role of F3 and linguistic experience in perception of rhoticity. As Heselwood & Plug highlighted, few experimental work has been conducted to verify the general assumption on which a low F3 would be responsible for rhoticity perception. If a perceptual peak in F2 region is indeed inducing perception of rhoticity and if language-specific phonetic/phonological characteristics play a role, we should observe realization of the predicted results of this project. Indeed, linguistic habits are thought to shape perceptual space and thus perception of rhoticity might rely not only on an acoustic peak in F-rho but as well on acoustic values of F3 more frequently manifested in the native language.
Further work using eye tracking methods would be beneficial to study the temporal dimension of coarticulatory patterns perception in the vocalic portion and thus put under the light particular acoustic cues that drive in real-time rhoticity judgments.

6. References

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, *133*(4), 2350-2366.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775-794.

Chen, S., & Mok, P.P. (2021). *Articulatory and Acoustic Features of Mandarin /ɻ/: A Preliminary Study. 2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP), 1-5.*

Clark, J., & Yallop, C. *(1990). An introduction to phonetics and phonology*. Oxford: Blackwell.

Huang, T. (2010). Er-suffixation in Chinese monophthongs: phonological analysis and phonetic data. In *Proceedings of the 22nd North American Conference on Chinese Linguistics (NACCL-22) & the 18th International Conference on Chinese Linguistics (IACL-18)* 2010. Vol 1, 331-344.

Huygevelde, M., Xiang, Y, Uddin Ahmed, A., Yi, D. (2023). *Cross-Linguistic Perceptual Variations in Rhotic Vowels between Native Mandarin and American English Speakers* (non published), proposal in Speech Perception, transnational seminar Université Paris Cité & University of Cologne.

Ingram, J. C., & Park, S. G. (1998). Language, context, and speaker effects in the identification and discrimination of English/r/and/l/by Japanese and Korean listeners. *The Journal of the Acoustical Society of America*, *103*(2), 1161-1174.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *the Journal of the Acoustical Society of America*, *67*(3), 971-995.

Kochetov, A., & Smith, J. (2009). Cross-language perception of Russian plain/palatalized laterals and rhotics. *The Journal of the Acoustical Society of America*, *126*(4), 2313.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell Publishers.

Lee W.S. (2005), "A phonetic study of the "er-hua" rimes in Beijing Mandarin," In *Proceedings of Interspeech (2005).*

Lee, W.S., & Zee, E. (2003)."Standard Chinese (Beijing)," *Journal of the International Phonetic Association*, 33(1).

Liu, A. (2016). Etude acoustique et représentation phonologique sur/ə-/suffixe rhotique en mandarin (Acoustic study and phonological representation of the rhotic suffix/ə-/in mandarin). In *Actes de la conférence conjointe JEP-TALN-RECITAL 2016. Volume 1: JEP*, 292-300.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Yamada, R.A. & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics* **52**, 376–392.

Wiese, R. (2011). The representation of rhotics. *The Blackwell companion to phonology*, 1-19.

Cross-Linguistic Imitation: English to Turkish

Erik Morris[1], Dilara Balaban[1], Doğa Ancin[2], Nadya Pelagheina[2]

[1] *Université Paris Cité*
[2] *Universität zu Köln*
erik.morris@etu.u-paris.fr

Abstract

This study aims to examine the relationship between phonetic imitation in a speaker's native language and phonetic imitation in a foreign language. It aims to expand on previous research on imitation in vowels to determine whether degree of imitation in L1 carries over to an L2. The study examines speakers of SSBE and exposes them to Turkish stimuli to assess this relationship. The experiment procedure consists of two tasks: (1) reading the English word list displayed on a computer screen (baseline condition), (2) producing the English words after exposure to a native Canadian English model talker, and (3) producing novel Turkish words spoken by native Turkish model talkers (shadowing conditions). The expected results are that high imitators in their L1 will display the same degree of imitation in Turkish. We posit this could provide an advantage in L2 acquisition, with high imitators producing novel vowels in a more native like manner, earlier.

Keywords: imitation, L2 acquisition, speech production, vowel rounding, Turkish

1. Introduction

As any of us who have attempted to learn a new language can likely attest, learning the vocal qualities and accent of a language can be rather difficult. Some of us may find it surprisingly easy and it may just be that phonetic imitation plays a role. Phonetic imitation (accommodation, convergence) can loosely be defined as "an increase in the perceptual or acoustic similarity between one speaker and another". This phenomenon has been studied extensively within speakers of a single language, and speakers have been shown to imitate consistently (Wade, Lai, and Tamminga 2021), but only a few have investigated cross-linguistic imitation (Jiang and Kennison 2022; Kartushina et al. 2016; Lewandowski and Jilka 2019). Whether statistically proven "high imitators" in their first language also display the same level of imitation in a foreign language, to which they have never been exposed, is an intriguing question. If speakers do, could this confer an advantage in their SLA of the target language's accent. This study investigates whether the degree to which individuals imitate is consistent across languages, and whether "high imitators" have an advantage in the phonetic and phonological acquisition of a second language.

2. Background Literature

There are several models that seek to describe the mechanisms by which humans learn languages. One such model is the Revised Speech Learning Model (Flege and Bohn 2021) which points to a strong connection between perception and production, with neither taking the lead but rather evolving together. Inability to identify a distinction between a new sound and one already part of the L1 can lead to production errors. Particularly, native sounds that are similar (or proximal as we will call them) to a new sound will interfere, and could even block, the formation of a new category for the L2 sound. On the other hand, new sounds that are distinct from an L1 sound are often said to be easier to notice, and to some extent acquire, due to their perceptual salience. We should note that awareness of the distinction does not necessarily equate to ease of production for said distinction. However, it is reasonable to expect a greater degree of imitation for sounds with no L1 correlate due to this increased salience.

2.1. Factors in Imitation

Imitation is a somewhat nebulous concept that defies precise explanation of what is and is not a factor in the phenomenon. Age has been found to be a factor in imitation (Schertz and Johnson 2022) and gender (Babel 2009) could also be relevant to how much a person converges phonetically. Talent can play a role, as Lewandowski and Jilka (2019) discovered, those with phonetic talent demonstrated more convergence towards a dialogic partner throughout a conversation than those who were considered to have no phonetic talent. Participants were native German speakers with various levels of experience in English, who completed a Diapix task with two native English speakers of different dialects. To control for familiarity, a Hindi speaker was also used in the perception portion of the experiment. Participants would have had similar, if not equal amounts, of experience with German, while Hindi would have been a language all were completely unfamiliar with. To assess phonetic talent three tasks were used that examined speech production, perception, and imitation (which combines the former). Results from Lewandowski and Jilka (2019) suggest that phonetic talent in perception and production is required for imitation to occur.

Bilingualism and the effect it has on language processing has been tested in relation to imitation as well. Spinu, Hwang, and Lohmann (2018) found that bilinguals more consistently and closely imitated glottalized stop rates than monolinguals, even after brief exposure

and training to a novel accent. After training, bilinguals displayed greater retention rates of glottalization, where the monolinguals returned to their baseline. Imitation, as a factor, was seen to facilitate short-term phonetic and phonological learning, which could suggest that being a high imitator confers an advantage in these areas.

Wade, Lai, and Tamminga (2021) found that individuals seem to imitate at similar levels over time in the imitation of voice-onset time (VOT). They found that participants do not target precise VOT values for individual tokens, and that participants showed signs of returning to their baseline between visits. Despite this, regardless of other factors such as model talker, segment, and session, the high imitators tended to be consistent in their level of imitation. If a speaker is consistent in their imitation within a language it is plausible that they would be in another language, even a new one.

Sociolinguistics and psychology are at work also, as found in two studies where perception of an interlocutor appeared to determine the degree of imitation. An L2 learner may converge or diverge based on whether they perceive their interlocutor to be native in their L2 or not (Jiang and Kennison 2022; Zając and Rojczyk 2014). In their study, Jiang and Kennison (2022) found that Chinese L2 English learners phonetically converged to native English speakers rather than non-native, which matches a preference in the questionnaire expressing favour for native English pronunciation over non-native. This psychological motivation matches results from Zając and Rojczyk (2014) that direction of imitation may be influenced by a participant's attitude toward L2 pronunciation. Overall, these studies suggest that there is a strong desire to "sound native" when one is learning another language, which could be a driving force in the imitation of an L2 accent.

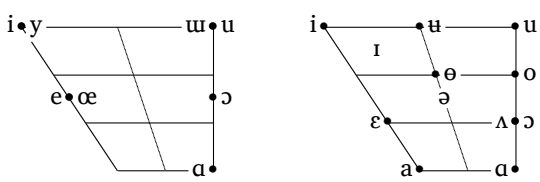
Coles-Harris (2017) discusses some of the literature on imitation and assumptions on its nature, which found the basis of this study as well. They discuss the two primary views on the motivation for imitation, whether it is a wholly automatic process or one that is motivated by social factors. We, like Coles-Harris, view imitation as a largely automatic process that can be mediated by external, social factors (Jiang and Kennison 2022; Zając and Rojczyk 2014). This automatic process is based on a direct link between perception and production and it is this link that suggests imitation would cross language boundaries.

2.2. Vowels

Vowels are often chosen for imitation studies (Babel 2009; Zając and Rojczyk 2014) for a number of reasons, including the ease with which measuring a imitation can occur through vowel duration and formants. Kartushina et al. (2016) used vowels in their experiment (though not directly an imitation study), which took French speakers and trained them to produce foreign vowels (Russian and Danish). The Danish vowel overlapped with a French vowel (proximal) while the Russian vowel had no overlap with any French vowel (distant). This research demonstrated that training had a greater effect on the production of vowels further from native vowels, while it had little impact on the proximal Danish vowel. In the early stages of learning it is the distant sound that may be produced less accurately due to requiring new patterns of articulation, in line with proposals by Flege and Bohn (2021). In this study, this means that we may see lower imitation in the proximal vowels as listeners may not consider them distinct from their L1, and that we could see greater imitation of distant vowels due to increased salience.

A recent study by de Jonge, Maxwell, and Zhao (2022) examined the acquisition of Turkish rounded vowels by L1 speakers of American English who had been residing in Turkey as part of their work for NGOs. All were required to pass a B2 proficiency test to qualify for their current roles so their experience with the Turkish language was already well established. Rather than exploring acquisition of novel vowels with naive participants, their goal was to examine the effect of language experience and length of residence on their productions of these rounded vowels. They found that some of their highest performers were those with the shortest lengths of residence, which, in the context of our study, could indicate that these participants were more readily predisposed to imitate. Their production results show the wide variation in producing the rounded vowels, all of which English lack (see Table 1). In order to acquire these rounded-unrounded pairs, L1 English speakers must learn to appropriately add or remove rounding.

Table 1: *Turkish vowel system, left, and Standard Southern British English, right.*



This is especially important because these three rounded-unrounded pairs are phonologically contrastive in rounding and position (Sabev 2019). Rounding is actively used to distinguish between vowel pairs of the same height and backness, and it is important in Turkish vowel harmony, making this feature key to proficient communication. English, on the other hand, makes greater use of position in the vowel space and duration to distinguish vowels. Rounding is present in English, though largely made redundant by the aforementioned cues, only in the differentiation between high and non-high front versus back vowels (de Jonge, Maxwell, and Zhao 2022). It is for these reasons that Turkish and English are the chosen languages to examine cross-linguistic imitation in our study.

3. Experiment

We are choosing to expand on existing literature on imitation in vowels, since this is a well-established area of research with clear guidelines on how to measure said imitation (i.e., duration and formants). Along the lines of Lewandowski and Jilka (2019), we will be conducting a two-phase experiment where participants will be grouped based on results from the first phase. The first phase will be an English shadowing task. The second phase will be

a Turkish shadowing task. Those who show signs of high imitation in phase one will be placed in one group, with the rest placed in the low imitation group. This is purely analytical in nature; the participants will not know, and the experiment takes place in a single session. Participants will also be analyzed on an individual basis, with each participant compared to themselves across phases.

As previously mentioned, six of the eight Turkish vowels come in rounded-unrounded pairs that are phonologically contrastive (Table 2). English, in our case Standard Southern British English (SSBE), has twelve vowels that make little use of rounding contrastively. Rather, English vowels are primarily distinguished by duration and position. This combination provides us with two vowel systems that share some vowels but use them in different ways, as well as vowels that exist in one but not the other.

Table 2: *Example of Turkish vowel distinctions.*

Rounding	Length
kotʃ, katʃ	kan, ka:n
kuʃ, kaʃ	atʃ, a:tʃ
kɯl, kul	bu, bu:
diʃ, dyʃ	er, e:r

We can safely assume that vowels shared between the two languages, especially cardinal vowels, will be produced with a similar F1 and F2 placing them in approximately the same region of the vowel space. These would be /i, ɑ, u/ primarily, though possibly with the addition of the mid-vowels /ɛ, ɔ/. These vowels will very likely differ slightly in quality, and could pose a problem to L1 English speakers as they may automatically categorize what they hear as a token of that L1 vowel and not alter their productions. The novel Turkish vowels may be treated in one of two ways. First, as novel segments they will be easier to acquire due to their perceptual salience and we can expect to see greater imitation of these vowels. Second, they could be collapsed into the nearest English phoneme and this will interfere with acquisition of the contrasts. Which phoneme they are considered a token of is a little difficult to pin down. According to de Jonge, Maxwell, and Zhao (2022) American English speakers have been shown to categorize [œ] as somewhere between /ɛ/ and /ʊ/. We can posit that [y] will be perceptually assimilated to SSBE /u/, and [ɯ] could be spread across /ʊ/, /u/, or /ʌ/ (de Jonge, Maxwell, and Zhao 2022).

This study aims to explore the cross-linguistic persistence of imitation and whether a participant who imitates highly in their L1 also does so in an L2. We can then question whether this persistence would confer an advantage in language acquisition. This is reflected in our two questions:

(1) Is the degree of imitation a person exhibits consistent across languages?

(2) Do "high imitators" have an advantage in phonetic and phonological second language acquisition?

4. Methods

4.1. Equipment

Recordings will be taken using a Tascam DR-40X field recorder with an Audio Technica AT2020 XLR microphone at 44.1 kHz, 24-bit PCM WAV format. For playback of the stimuli, participants will be wearing a pair of Philips Fidelio X2HR open-back, over-ear headphones, which should allow them to hear their own voice when speaking.

4.2. Materials

Two lists of words, one English and one Turkish, constructed to elicit the vowels of each language are used. The English word list contains ten words designed to elicit the ten monophthongs in SSBE with one word per vowel in CVC structure (Table 3). Included are six distractor words for a total of sixteen words which are displayed on screen in random order for the participants to read.

Table 3: *English word list with corresponding target vowels.*

Word	Vowel (IPA)
heed	/i/
hid	/ɪ/
head	/ɛ/
had	/æ/
hod	/ɔ/
hawed	/o/
hood	/ʊ/
who'd	/u/
hud	/ʌ/
heard	/ɚ/

Table 4: *Turkish word list with corresponding target vowels.*

Word (Orthography)	Word (IPA)	Vowel (IPA)
diş	/diʃ/	/i/
düş	/dyʃ/	/y/
deş	/deʃ/	/e/
döş	/dœʃ/	/œ/
kaş	/kaʃ/	/a/
dış	/dɯʃ/	/ɯ/
duş	/duʃ/	/u/
koş	/koʃ/	/o/

The Turkish word list contains eight words in CVC structure with six of the words forming minimal pairs in rounding (Table 4). We have attempted to compensate for what de Jonge, Maxwell, and Zhao (2022) saw as a shortcoming in their experiment by selecting words without a lateral approximant after the vowel. Likewise, we have used monosyllabic words to avoid Turkish vowel harmony and also to be in keeping with the English word list. Eight distractor words are included to bring the list to sixteen words in keeping with the English portion.

4.3. Participants

All participants and model talkers in the study are required to be over the age of eighteen years old, with no diagnosed hearing or speech difficulties.

All participants will complete a demographic questionnaire asking participants for their age, gender, their place of birth/where they grew up, languages spoken, their exposure to other foreign accents, and whether they have had any diagnosed hearing/speech disorders.

4.3.1. Monolingual English Speakers

A total of 40-60 monolingual Standard Southern British English (SSBE) speakers will partake in the study. The only additional requirement for participation is zero prior experience with Turkish, which here means that they have not had any formal or informal instruction. Exposure to, and experience with, other languages will be collected on the background questionnaire in the event that it could affect results.

4.3.2. Model Talkers

A single speaker of Canadian English provides the stimuli for the first phase of the experiment testing imitation levels in English.

Four native Turkish speakers, two female and two male, provide the stimuli for the second phase. To control any influence stemming from dialectal variations we will attempt to recruit speakers from a single dialectal area. Multiple model talkers are used in this portion of the experiment to encourage imitation of the general features of Turkish – not of one speaker specifically. The word list will be read in full by all speakers and divided equally between the model talkers in the shadowing portion, to ensure participants are exposed to all model talkers equally.

4.4. Experiment Procedure

Collection of stimuli for both phases of the experiment is conducted via a self-paced reading of the word list presented on slides. The list is read three times with the best tokens extracted and intensity normalized to 70 dB using Praat (Boersma and Weenink 2023) prior to use in the main experiment.

The main experiment is built in OpenSesame (Mathôt, Schreij, and Theeuwes 2012) with participants wearing headphones for the shadowing portions. It consists of two phases, one for assessing imitation in English and one for testing how much this imitation carries over into Turkish. Both phases will include a practice block to allow for familiarization with with experiment format.

Phase 1: English Shadowing Task. This first phase of the experiment tests participants for imitation in English. A baseline will be taken prior to any experimental phase. Participants read the English word list, three times through. The second read-through acts as the baseline. The third reading (and then the first) are used in case of any recording issues (i.e., loud noise, stuttering etc.). After recording the baseline, participants will then complete two blocks (3 repetitions per block) of a shadowing task, where they will hear the English model talker, one word at a time, and then repeat the word aloud. This will result in 66 tokens total or 6 tokens/vowel for each participant.

Results from this phase will then be used to divide the participants into one of two analytical groups: high imitators and low imitators. To be considered a "high imitator" a participant must show a greater than 60% success rate – meaning that in more than 60% of tokens they display clear imitation – with the remainder of participants going into the other group. This is a purely statistical division for the purposes of comparing the groups relevant to our hypothesis, participants will not be physically divided into different experimental groups.

Phase 2: Turkish Shadowing Task. For the second phase, which will occur immediately after the first, participants will complete two blocks (3 repetitions per block) shadowing the Turkish word list. Like Phase 1, there will be one word per Turkish vowel (8 words total). In total there will be 48 tokens of Turkish shadowing, or 8 tokens per vowel for each participant. Unlike Phase 1, participants will not be reading the Turkish word list aloud first and will only complete a shadowing task.

After the main experiment portion, participants will re-read the baseline English word list three times. Their productions pre- and post-experiment will be compared to assess the longevity of imitation.

4.5. Measurements

All measurements for the vowels will be taken using Praat (Boersma and Weenink 2023) with durations marked out via TextGrid. A custom script will iterate through each sound file extracting the duration of each vowel as well as the formant from five timepoints across the duration of the vowel: 0, 25, 50, 75, and 100 percent.

Duration of the vowels will be taken, along with the first three formants (F1, F2, F3) for all English participants as well as the Turkish model talkers. As Turkish phonology does not use duration as a contrastive feature to the extent it is in English (e.g. [hit] vs [hɪt]), duration will be used to assess whether English speakers are shortening their normally longer vowels to match Turkish. For the vowel pairs in Turkish distinguished primarily by rounded vs unrounded (e.g. [i] vs [y]) the third formant will be used to aid in assessing the degree of imitation in this factor.

The extent of imitation based on these values will be assessed in two ways based on those used by Schertz and Johnson (2022). Proximity is the absolute value of the difference between the participant's production and the model talker's production, with smaller values representing more imitation. Change In is the difference between a participants' baseline and shadowed productions which can be used to catch broader signs of imitation. Proximity values per token, per speaker, will be calculated and the average across all speakers per target vowel can be used to show general patterns.

4.6. Analysis

To visualize the distribution of tokens and the overall trends in imitation, tokens will be plotted using R (R Core Team 2023) inside a vowel space with the mean ellipses for each vowel (Figure 1). The centroid of each vowel as produced by the Model Talkers will be the comparison point to which the centroid of each participants'

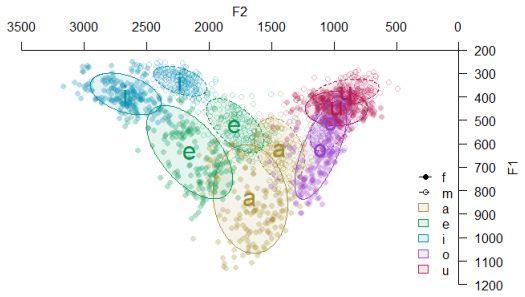


Figure 1: *Vowel Clouds with Mahalanobis Ellipses.*

productions will be compared. The Euclidean distance between these centroids will give an indication of how a speaker performs overall, with lower values for this distance indicating imitation in the F1 x F2 plane.

To compare the other qualities, duration and F3, we rely on Proximity and Change In as mentioned above. Lower values for Proximity would indicate a closer production between participant and model talker. Change In would be an indication of how much a speaker altered their production *relative to their baseline*. In practice this can be both towards (convergence) a model talker or away from them (divergence). In either case this is an indication that a speaker is attempting to alter their production and can be taken as a sign of imitation.

A comparison of the baseline English word list reading and the post-test reading will capture any traces of alterations to vowel quality resulting from exposure to Turkish. Altered English vowel productions are another indication of imitation and an additional point of comparison when determining how much a speaker has altered their own productions. It may also hint at some of the long-term effects that exposure to additional languages can have on L1 sounds as discussed by Flege and Bohn (2021).

5. Expected Results

Due to results from previous research on imitation (Lewandowski and Jilka 2019; Wade, Lai, and Tamminga 2021) we can expect that degree of imitation will carry over from L1 to an L2. Likewise, just as in research by Spinu, Hwang, and Lohmann (2018) we can expect high imitators to have a perceivable advantage over the lower imitators in their productions of Turkish vowels. This will not apply to all vowels equally, as found by Kartushina et al. (2016) we expect that proximal vowels will show less imitation and may be entirely supplanted by the L1 vowel. This is likely to be the case regardless of whether a speaker is a high or low imitator though we can posit less instances of this with the high imitators.

It is still possible that there will be no appreciable difference between high and low imitators. Our study is relatively small and examines only vowels in very limited contexts. The assessment of rounding on vowels is not entirely precise either. Rounding of vowels is not tied to a single indicator, while F3 has been shown to indicate rounding it can also be tied with F2. This is not a longitudinal study either, and it may take a longer period of exposure, or more consistent exposure, for any results to show. Our study consists of a single test session that exclusively examines the production side of the equation. We do not test participants on their discrimination of the contrasts used in the imitation stimuli. Because of this we have only one side of the story and no long-term data regarding the benefits of imitation in language learning. Adding a perception portion and additional testing sessions, spread apart, would allow us to test more of the connection described by Flege and Bohn (2021).

We expect that high imitators will display the greatest imitation of Turkish vowels, however it is also possible that other factors will either be responsible for it or hinder it. Recall that age, gender, and subjective opinions of the interlocutor have had effects on imitation (Babel 2009; Jiang and Kennison 2022; Schertz and Johnson 2022; Zając and Rojczyk 2014). The methodology of the study should be able to tease apart the influences of these factors, but it is still a cautionary note.

While we will be looking at participants as individuals we will also be looking at them as a whole, continuous group of monolingual English speakers. In the overall, generalized viewpoint of the study we do expect to see that imitation carries over from L1 to L2. The size of the effect may be severely reduced or inflated depending on the specific set of participants we manage to recruit. Once again the study is fairly small and it is possible that chance, as it were, could have it that we recruit a set of people skewed in favour of one or the other.

6. Acknowledgements

The author would like to thank Professors Doris Mücke and Hiyon Yoo for teaching, as well as Elise Herbig and Gyong Min Oh for facilitating (and putting up with my many questions), the EduVEnture Cologne (IVAC) program funded by the DAAD for making the exchange possible, and Universität zu Köln, IfL Phonetics for hosting.

7. References

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

Boersma, P. and D. Weenink (2023). *Praat: doing phonetics by computer*. Retrieved 30 November 2023. Version 6.4.01. Computer Program. URL: `http://www.praat.org`.

Coles-Harris, E. H. (2017). "Perspectives on the motivations for phonetic convergence". In: *Language and Linguistics Compass* 11.12. e12268 LNCO-0713.R1, e12268. URL: `https://doi.org/10.1111/lnc3.12268`.

de Jonge, K., O. Maxwell, and H. Zhao (Sept. 2022). "Learning on the Field: L2 Turkish Vowel Production by L1 American English-Speaking NGOs in Turkey". In: *Languages* 7. URL: `https://doi.org/10.3390/languages7040252`.

Flege, J. E. and O.-S. Bohn (2021). "The Revised Speech Learning Model (SLM-r)". In: *Second Language Speech Learning: Theoretical and Empirical Progress*. Ed. by R. Wayland. Cambridge University Press, pp. 3–83. URL: `https://doi.org/10.1017/9781108886901.002`.

Jiang, F. and S. Kennison (2022). "The Impact of L2 English Learners' Belief about an Interlocutor's English Proficiency on L2 Phonetic Accommodation". In: *Journal of Psycholinguistic Research* 51.1, pp. 217–234. URL: `https://doi.org/10.1007/s10936-021-09835-7`.

Kartushina, N., A. Hervais-Adelman, U. H. Frauenfelder, and N. Golestani (2016). "Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training". In: *Journal of Phonetics* 57, pp. 21–39. URL: https://doi.org/10.1016/j.wocn.2016.05.001.

Lewandowski, N. and M. Jilka (2019). "Phonetic Convergence, Language Talent, Personality and Attention". In: *Frontiers in Communication* 4. URL: https://doi.org/10.3389/fcomm.2019.00018.

Mathôt, S., D. Schreij, and J. Theeuwes (2012). "OpenSesame: An open-source, graphical experiment builder for the social sciences". In: *Behavior Research Methods* 44.2, pp. 314–324. URL: https://doi.org/10.3758/s13428-011-0168-7.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Version 4.3.2. Computer Software. Vienna, Austria: R Foundation for Statistical Computing. URL: https://www.R-project.org/%20(EPEL%20Repo).

Sabev, M. (2019). "Spectral and durational unstressed vowel reduction: an acoustic study of monolingual and bilingual speakers of Bulgarian and Turkish". PhD thesis. University of Oxford.

Schertz, J. and E. K. Johnson (2022). "Voice Onset Time Imitation in Teens Versus Adults". In: *Journal of Speech, Language, and Hearing Research* 65.5, pp. 1839–1850. URL: https://doi.org/10.1044/2022_JSLHR-21-00460.

Spinu, L. E., J. Hwang, and R. Lohmann (2018). "Is there a bilingual advantage in phonetic and phonological acquisition? The initial learning of word-final coronal stop realization in a novel accent of English". In: *International Journal of Bilingualism* 22.3, pp. 350–370. URL: https://doi.org/10.1177/1367006916681080.

Wade, L., W. Lai, and M. Tamminga (2021). "The Reliability of Individual Differences in VOT Imitation". In: *Language and Speech* 64.3. PMID: 32772791, pp. 576–593. URL: https://doi.org/10.1177/0023830920947769.

Zając, M. and A. Rojczyk (2014). "Imitation of English vowel duration upon exposure to native and non-native speech". In: *Poznan Studies in Contemporary Linguistics* 50.4, pp. 495–514. URL: https://doi.org/10.1515/psicl-2014-0025.

Final Project

Yunzhuo XIANG, 22201420, Paris, Université de la Sorbonne Nouvelle

Topic: A psychoacoustic study in language background, durational and the quantity of spectral change factors in the perception of monophthong-diphthong continuum [a- ai], [ɑ - ɑu] of Beijing Mandarin.

Introduction

The perception of diphthong is a topic with ongoing debates. Some difficulties come from the complexity of factors in diphthong perception, such as onset frequency value, offset frequency value, transition slope, duration of steady-state, etc. The main hypothesizes around the most important acoustic cues of diphthong perception were summarized in *Gottfried et al. 1993*, *Morrison & Nearey 2007* and *Morrison 2013*. They are hypotheses of "Onset + Offset", "Onset + Slope" and "Onset + Direction", all claiming that diphthong onset frequency value is important but they differ in what is the nature of the dynamic part. Early study of *Gay 1968,1970* reported that different English diphthongs have different F2 transition slopes (rate of change: ΔF2/t) by which they are distinguished. *Gay 1970* then proposed the "Onset + Slope" hypothesis. Gay's study was criticized by *Bladon 1985*, insisting that the frequency of diphthong offset formants is the most important cues and not the F2 transition slope. *Pols 1977* studied the Dutch vowels and proposed the "Onset + Direction" hypothesis which claimed that neither the offset value nor the slope is important but the direction of F2 transition (positive, negative, steady-state). The three hypotheses are linked to each other in real analysis. Given the onset and the offset F1-F2 values and the transition duration, one may easily calculate the slope (rate of change) and the direction. The current study doesn't want to test three hypotheses, but focusing on the monophthong-diphthong distinguishing.

Diphthong differs from the monophthong by its dynamic formant spectral change. This conclusion encounters a problem that it has been shown that all vowels have a certain amount of spectral change (*Nearey & Assmann, 1986*) and a spectral transition do not always been perceived, causing assimilation between monophthong and diphthong (*Gay 1970, Bladon 1985*). Imaging a situation, how can we transcript a vowel with a small spectral change, when it's an unknowing language that we know nothing about its phonological system? By now, it is not clearly studied to what extend the presenting spectral change is significant (can be perceived by a native speaker) and is this perceptual competence is influenced by listener's language background. The current study aims to addressing this question by conducting a psychoacoustic study of monophthong-diphthong continuum perception.

The studies of diphthong perception revealed several factors in monophthong-diphthong discrimination, for example, *Bladon 1985* use [ia], [iɛ], [ie] sounds produced by a phonetically trained speaker, he then cut the sounds by keeping the onset and matching the onset with different duration of transitions (s1, s2, s3, s4 and

s5 in **Figure 1**), so he gradually increasing the duration and the amount of spectral change at a same time. The listeners are also phonetically trained people and they were asked to label the sound that they heard. The result showed that, the duration and the amount of spectral change are both needed for a certain diphthong to be perceived. Result also showed a cross effects from duration and spectral change. For example, given a 50ms of vowel duration, the [ia] stimulus was already labeled as [iɪ], demonstrating that phonetically trained people perceived the spectral change of 50ms [ia] stimulus. On the contrary, 50 ms [iɛ] and [ie] were labled as monophthongs. This is because [ia] has a greater rate of change than [iɛ] and [ie] and thus, by 50ms, [ia]'s amount of change is greater than [iɛ] and [ie]. Similar phenomenon can also be found from *Gay 1970, He 1985, Nábělek et al. 1994.*

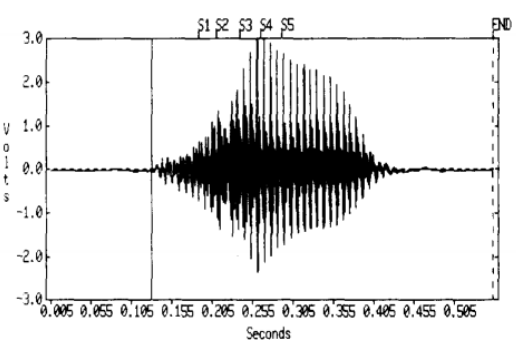


Fig. 2. Waveform of [ia], used in Experiment 1. Male speaker, monotone pitch, as analyzed on N.E.D. Sample-to-Disk computer after digitization at 50 kHz. Curtailed stimuli were created extending from the vertical line to each of the cutoff points S1 to S5.

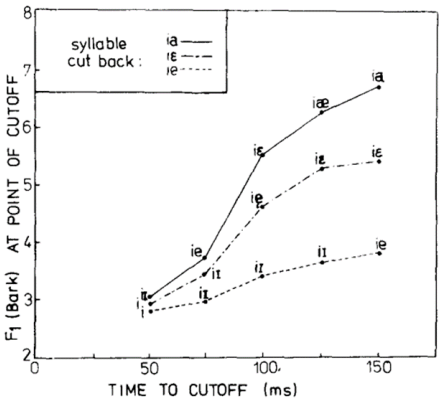


Fig. 3. Preferred identification (shown as a label) assigned to the curtailed stimuli of Experiment 1. Each data point represents a stimulus, plotted as its F1 frequency (Bark) versus its time to cutoff.

Figure 1. The used stimuli (left) and the result (right) of *Bladon 1985*.

Previous studies showed a need to test the effects from the duration and the spectral change respectively, which is not well-controlled in the existing literatures. At the same time, it's also interesting to test whether the perception of spectral change is influenced by listener's language background. The current study aims to run the experiment by matching different duration with different amount of spectral change and to see the difference between two groups of participants.

Method

1. **Stimulus:**

   1.1 Standard diphthongs and monophthongs

   The stimuli are synthesized Beijing Mandarin female speaker' diphthongs [ai], [ɑu], monophthongs [a], [ɑ] and a series of continuum from the monophthong to the diphthong. The data of vowels are from a production experiment conducted during November 2023 – January 2024. 10 Female speakers of Beijing Mandarin residing in Paris have been recruited to pronounce different diphthongs and monophthongs in different contextual

and speech rate environment. The following figure 2 shows a GAMMs analysis of [ai] and [ɑu] of normal speech rate without consonant attack:

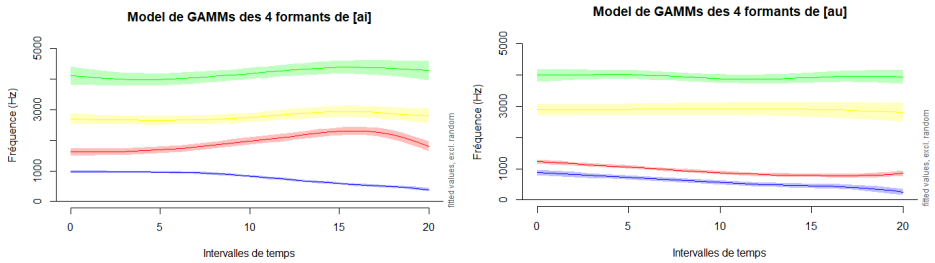


Figure 2. GAMMs of formants of diphthongs [ai] and [ɑu] in Beijing Mandarin.

The band widths of F1 – F4 were measured and the synthesis options of the monophthongs [a], [ɑ] and diphthongs [ai], [ɑu] are shown in the following table 1. The band widths and F3, F4 won't be changed among all stimuli. All stimuli have a duration of 200 ms. To match the production data, diphthong [ai] has a 30% (60 ms) initial steady-state thus its F1, F2 change from the 30% to 100%. For diphthong [ɑu], it doesn't have any steady-state, so the [ɑu] would be a continuous transition from onset to offset.

Onset/monophthong	F1	B1	F2	B2	F3	B3	F4	B4
a	1000	200	1600	300	2900	400	4000	500
ɑ	900	200	1300	150	2900	400	4000	300
Offset/diphthong	F1	B1	F2	B2	F3	B3	F4	B4
(a)i	500	200	2400	300	2900	400	4000	500
(ɑ)u	500	200	900	150	2900	400	4000	300

Table 1. Formants and Band widths of synthesized diphthongs and monophthongs

1.2 Continuum

The continuum range systematically in duration and spectral change quantity. The durations tested are 50ms, 60ms, 70ms, 80ms, 90ms, 100ms. For diphthong [ai], it always has a 30% steady-state portion and a 70% spectral change portion. So, the duration structure of [ai] is:

Total duration	50 ms	60 ms	70 ms	80 ms	90 ms	100 ms
Steady-state	15 ms	18 ms	21 ms	24 ms	27ms	30 ms
Spectral change	35 ms	42 ms	49 ms	56 ms	63 ms	70 ms

Table 2. Duration structure of [ai] in different duration conditions.

The amount of spectral change is measured using the Euclidean distance in the F1 – F2 frequency space, the formula is:

$$\sqrt{\left\{\left(F1_{\{n+1\}} - F1_{\{n\}}\right)^2 + \left(F2_{\{n+1\}} - F2_{\{n\}}\right)^2\right\}} = \textit{Total amount}$$

The $F1_{n+1}$ and $F2_{n+1}$ is the terminal spectral change point (offset) and the $F1_n$ and $F2_n$ are the starting spectral change point. Because we already know the values of diphthong onset, given a certain amount of spectral change, it's not difficult to calculate the offset F1 and F2 value. Another concerning is the rate of change, for diphthong [ai], it's F2 rate of change is set as 1.5 times bigger than F1 to match the production data. For [ɑu], the F1/F2 are parallel downwards which means that they have a similar rate of change. The continuum of amount of spectral change is shown in the table below, the first line is the total amount of spectral change, the corresponding F1 and F2 are showed in frequency value (Hz), rounded as integer.

Total Amount	50	100	150	200	250	300	350	400
[ai]$F1_{offset}$	972	945	917	889	861	833	805	778
[ai]$F2_{offset}$	1642	1683	1725	1766	1808	1850	1891	1933
[ɑu]$F1_{offset}$	865	829	794	759	723	688	653	617
[ɑu]$F2_{offset}$	1265	1229	1194	1159	1223	1088	1053	1017

Table 3. F1, F2 of the monophthong-diphthong continuum.

So there would be 2 diphthong × 6 durational conditions × 8 spectral change conditions + [a],[ai],[ɑ],[ɑu] of 200 ms = 100 stimuli in total. All stimuli have a falling tone of which the F0 decrease from 200 Hz to 160 Hz.

2. Participants

The participants would be young native Beijing Mandarin speakers and young native Xunyang Mandarin speakers without any phonetical training. The [ai] and [ɑu] are monophthongized in the Xunyang Mandarin (a dialect of Mandarin of Central Plaine) speakers' production, phonetically transcribed as [ɛ] and [ɔ]. The minimal number of listeners would be 20 for each group. We would like to see a difference between Beijing Mandarin listeners and Xunyang Mandarin listeners.

3. Procedure

A forced-choice task will be run using the online experimental tool PsyToolkit[1]. In each trial, the listener will hear one stimulus and he/she is forced to choose between a monophthong morpheme 啊 ([a]/ [ɑ], Mandarin doesn't opposite front [a] and back [ɑ] so the two monophthongs will be

[1] https://www.psytoolkit.org/

represented by a single morpheme) or the diphthong morphemes 爱[ai], 奥[ɑu].

Discussion of result predicted

The following figure 3 shows the result expected of 50 ms condition V.S. 100 ms condition. The blue line is the result of Xunyang Mandarin group and the red line is the Beijing Mandarin group:

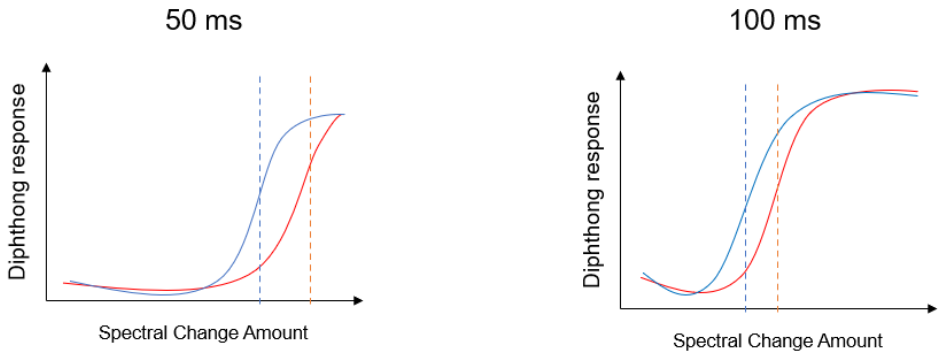


Figure 3. Result expected of 50 ms condition and 100 ms condition.

First, in general, as the amount of spectral change increase, the response ration of diphthong is higher and higher. And because it's a forced-choice take, so we would get a classical s-shape result, where the dotted lines indicate the shifting area.

Second, given that stimuli range from 50 ms to 100 ms, we expected that, having the same spectral change amount, shorter duration stimuli would be more difficult for all listeners to perceive the spectral change. Thus, the shifting area is later in 50 ms stimuli than 100 ms stimuli.

Third, we expected that the monophthongization of [ai] and [ɑu] in Xunyang Mandarin will make listeners of this dialect more sensitive to the spectral change of the vowels. Beijing Mandarin speaker often produces diphthong with a greater spectral change amount than all stimuli in the experiment, so a small amount of spectral change might be ignored due to magnitude effect and they may usually perceive the spectral change when it's greater than Xunyang group.

References

Gottfried, Michael, James D. Miller, and Donald J. Meyer. "Three Approaches to the Classification of American English Diphthongs." Journal of Phonetics 21, no. 3 (July 1, 1993): 205–29. https://doi.org/10.1016/S0095-4470(19)31337-3.

Geoffrey Stewart Morrison and Terrance M. Nearey, "Testing Theories of Vowel Inherent Spectral Changea)," The Journal of the Acoustical Society of America 122, no. 1 (June 4, 2007): EL15–EL22, accessed August 18, 2023, https://doi.org/10.1121/1.2739111.

Geoffrey Stewart Morrison, "Theories of Vowel Inherent Spectral Change," in Vowel Inherent Spectral Change, ed. Geoffrey Stewart Morrison and Peter F. Assmann (Berlin, Heidelberg: Springer Berlin Heidelberg, 2013), 31–47, https://doi.org/10.1007/978-3-642-

14209-3_3.

Anthony Bladon, “Diphthongs: A Case Study of Dynamic Auditory Processing,” Speech Communication 4, no. 1–3 (1985): 145–154.

Thomas Gay, “A Perceptual Study of American English Diphthongs,” Language and Speech 13, no. 2 (1970): 65–88, accessed January 15, 2024, http://journals.sagepub.com/doi/10.1177/002383097001300201.

Terrance M. Nearey and Peter F. Assmann, “Modeling the Role of Inherent Spectral Change in Vowel Identification,” The Journal of the Acoustical Society of America 80, no. 5 (November 1, 1986): 1297–1308, accessed August 25, 2023, https://doi.org/10.1121/1.394433.

Ningji, He, “Durational factors in the perception of Beijing Mandarin diphthongs,” The Record of Experiments of Beijing Languages, Press of Beijing University, 1985. (贺宁基, “北京话二合元音感知中的时间因素, 北京语言实验录” ,北京: 北京大学出版社, 1985).

Anna K. Nábělek, Zbigniew Czyzewski, and Hilary Crowley, “Cues for Perception of the Diphthong /ɑ i / in Either Noise or Reverberation. Part I. Duration of the Transition,” The Journal of the Acoustical Society of America 95, no. 5 (May 1, 1994): 2681–2693, accessed January 16, 2024, https://pubs.aip.org/jasa/article/95/5/2681/622583/Cues-for-perception-of-the-diphthong-i-in-either.