

NORTHWESTERN UNIVERSITY

Phonetic Accommodation after Auditory Exposure to Native and Nonnative Speech

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Linguistics

By

Midam Kim

EVANSTON, ILLINOIS

August 2012

© Copyright by Midam Kim 2012
All Rights Reserved

ABSTRACT

Phonetic Accommodation after Auditory Exposure to Native and Nonnative Speech

Midam Kim

We investigated native English talkers' phonetic accommodation to a native or nonnative model talker in a passive auditory exposure setting. We performed a phonetic accommodation experiment, following the procedure of Goldinger & Azuma (2004). Specifically, the participants read monosyllabic words, disyllabic words, and sentences before and after perceptual exposure to a model talker with a certain group level linguistic distance, namely, a native model talker with the same dialect, a native model talker with a different dialect, or a nonnative model talker. Additionally, participants' implicit attitudes towards foreigners were also measured by an implicit association task (IAT). We performed various acoustic measurements on monosyllabic and disyllabic words, and dynamic time warping (DTW) analyses and XAB perception tests on sentences. We found that dialect mismatch and L1 mismatch between participants and their model talkers did not inhibit participants' phonetic convergence in most acoustic measurements on words and XAB perception test results on sentences. Instead, within each group level linguistic distance, at the item level, the preexisting acoustic distances between model talkers and participants before auditory exposure positively affected their degrees of phonetic convergence, regardless of the direction of the change. That is, the farther the acoustic distance was before the auditory exposure, the larger the degree of phonetic convergence was. However, there were variations in the influence of participants' implicit attitudes towards foreigners on their phonetic accommodation to nonnative model talkers. Finally, the perceived phonetic convergence patterns by human listeners were predicted by the DTW analyses results. Overall, we found reliable

evidence of phonetic convergence to all native and nonnative model talkers from lower-level monosyllabic and disyllabic words to higher-level sentences after passive auditory exposure.

ACKNOWLEDGEMENTS

Finishing my days as a student officially, I have many people to thank. The 30 or more years of learning have been full of joy and pain. It has been a long time but, looking back, it was a short moment. Overall, I was totally blessed.

Like for many others, my first education started at home. When I was 4 years old, my mom taught me Hangul, the Korean orthography with her own handwriting of the letters and drawings that could easily represent each letter. Then my dad, a linguist, taught me the basic Korean grammar when I was 9 years old, which I really enjoyed. When I started learning English when I was 11 years old, I was totally fascinated by the new speech sounds and the letters that could represent the sounds, the International Phonetic Alphabets (IPA). My dad helped me practicing IPA transcriptions in English and Korean, and it was so fun! So, not only for their great parental support for me, but also I owe my mom and dad a lot for their leading my life as a linguist and phonetician.

Fast forward to my grad school days for MA in linguistics at Seoul National University. I first learned articulatory phonetics from Professor Hyunbok Lee and experimental phonetics from Il-jin Jeong. My advisor, Professor Hoyoung Lee, encouraged my thesis work on correlation between voice onset time and the fundamental frequency in the perception of Korean stops and affricates. Then a postdoctoral researcher at SNU and now a professor at Hongik University, Professor Hansang Park, helped my thesis through weekly meetings and thorough revisions. Professor Jeong-Im Han at Konkuk University introduced Northwestern University to me as a good place to continue my research in speech perception.

Through my PhD work at Northwestern University, I could grow as a psycholinguist. My primary advisor, Professor Ann Bradlow led my research interests to speech perception and production in global communications. In our group project to make a corpus on spontaneous conversations between native and nonnative talkers, the Wildcat Corpus, I became very interested in how native and nonnative talkers accommodate their speech to their partner's. Ann helped me develop the question into a well-designed research project. Through discussion with Professor Sid Horton in Psychology I could view this "me-search" in the frame of cognitive psychology. This great collaboration led to our paper published in *Journal of Laboratory Phonology*, "Phonetic convergence in spontaneous conversations as a function of interlocutor language distance" (Kim, Horton, and Bradlow, 2011). My interest in global communications also developed to my second qualifying paper, "Discourse markers during dialogues between native and nonnative speakers" with the guidance of Professor Brady Clark, Professor Sid Horton, and Professor Andrea Sims (then a postdoctoral researcher at NU and now a professor at The Ohio State University). For my dissertation, along with Ann and Sid, Professor Matt Goldrick joined the committee and greatly contributed in improving the statistical analyses in the dissertation. Professor Dan Ellis at Columbia University provided me of great consultation on the dynamic time warping techniques (DTW) through extensive and prompt email exchanges. Professor Valerie Hazan at University College London, Professor Jennifer Pardo at Montclair State University, and Professor Stephen Goldinger at Arizona State University gave me precious comments on my dissertation work in our meetings in conferences. I greatly appreciate all the helps from my academic advisors and seniors inside and outside of NU for their guidance, advice, encouragement, and patience for me.

I have many more people to thank at Northwestern University. Professor Ann Bradlow has been a strong academic advisor and supporter for me beyond expectation. With her help, I could only concentrate on my research, not worrying about anything else. Moreover, she was truly a great mentor for many aspects of life. Watching her life, I could envision my own life balanced between a good scientist and teacher and a nice and happy person. There is just no way I can appreciate enough Ann's good influence on me. The Director of Graduate Studies, Professor Janet Pierrehumbert, gave me a huge support during my coursework. Our department engineer, Chun Liang Chan, has been always the magical help for many technical issues. The director of our ESL program and the International Summer Institute (ISI), Julie Moore, was the first person who welcomed me to Evanston and NU and helped me find the apartment building I lived in for 7 years. Our department secretary, Irene Sakk, always resolved all my complex issues as an international student so smoothly. My lovely friends at NU, Lauren Ackerman, Rachel Baker, Melissa Baese-Berk, Katinka Baren, Michael Blasingame, Ann Burchfield, Susanne Brouwer, Arim Choi and Tyler Perrachione, Robert Daland, Caroline Engstler, Angela Fink, Elisa and James Sneed German, Hyunok Ham, Jordana Heller and Paul Heider, Lisa Hesterberg, Jen Alexander Kiefer, Soyoung Kim, Su Jung Kim and Tae Joon Moon, Ken Konopka, Meredith Larson, Kyounghee Lee, Jenna Luque, Brett Margolis, Julie Matsubara, Kelsey Mok, Heather Norbury, Songyi Park, Celina Pettis, Rajka Smiljanic, and Kristin Van Engen, have been wonderful collaborators and great joy in my life.

There are two special people I would like to remember, my grandpa, Yeonghwan Kim, and Eun Myoung Kim. My grandpa loved me so much and was always so happy about my school progresses. He was the first one in our family who came to Seoul National University on the day of announcement for successful college applicants. When I earned my MA in Linguistics

and when I received the admission letter from NU, I missed him so badly. Eun Myoung was a beautiful singer in our choir in the First Presbyterian Church of Evanston and took care of me just like my mom when I was struggling to survive my first and second years at NU, while she herself was suffering from cancer. Both my grandpa and Eun Myoung are now in heaven, and I know that they are always looking at me with our God. But I miss them so much on this day of my achievement. As I once told Eun Myoung, my love for them is an egoistic one, while theirs for me was a self-sacrificing one.

Finally, I would like to thank special ones in my life. My long-time friend and roommate in Evanston, Su Jung, has been a great supporter for my life. She is not actually a friend to me, she is my family. My mom and dad, and my brother, Mideun, have always been the largest and deepest rock for my life. Although I am so sorry that I am far from them and cannot share every moment of our lives together, I always feel that we are always strongly connected to one another. My husband, Min-Young, has been the greatest joy of my life for 15 years. I sincerely thank God for giving you to me as the love of my life. My in-law family has always been sincerely supporting my research and career. I know that I did not earn their love for me. I give love and thanks to God, who made me, saved me, and loves me forevermore.

Although I am finishing my official education, I will always be a student, studying science and learning life. I will try to appreciate all the priceless helps I have received by giving them back to others in my path.

TABLE OF CONTENTS

ABSTRACT.....	3
ACKNOWLEDGEMENTS.....	5
LIST OF TABLES.....	11
LIST OF FIGURES	12
1. INTRODUCTION	14
1.1. Background	14
1.1.1. Theoretical accounts of phonetic accommodation.....	15
1.1.2. Intervening factors in phonetic accommodation.....	17
1.1.3. Generalizability of phonetic accommodation	24
1.1.4. Measurements on phonetic accommodation.....	25
1.2. The current study and the outline of the paper.....	29
2. GENERAL METHODOLOGY.....	35
2.1. Definition of phonetic accommodation: revisited and controls	35
2.2. Experiments.....	39
2.2.1. Phonetic Accommodation Experiment	40
2.2.2. Implicit Association Task (IAT).....	47
2.2.3. Final questionnaire.....	54
2.2.4. Statistical analyses	55
3. MONOSYLLABIC WORDS	60
3.1. Methods.....	60
3.1.1. Materials	60
3.1.2. Participants.....	61
3.1.3. Procedure	61
3.1.4. Analyses.....	62
3.2. Results	62
3.2.1. Linguistic talker distance and generalizability for phonetic accommodation	63
3.2.2. Implicit attitudes and generalizability of phonetic accommodation	81
4. DISYLLABIC WORDS	94
4.1. Methods.....	94
4.1.1. Materials	94
4.1.2. Participants.....	96
4.1.3. Procedure	96
4.1.4. Analyses.....	96

	10
4.2. Results	97
4.2.1. Linguistic talker distance and generalizability of phonetic accommodation.....	99
4.2.2. Implicit attitudes and generalizability of phonetic accommodation	107
5. SENTENCES	114
5.1. Methods.....	114
5.1.1. Materials	114
5.1.2. Participants.....	116
5.1.3. Procedure	116
5.1.4. Analyses.....	116
5.2. Results	121
5.2.1. Perceived phonetic accommodation	121
5.2.2. Dynamic time warping analyses	123
5.2.3. Prediction of perceived phonetic convergence with mechanically judged phonetic convergence	126
6. GENERAL DISCUSSION	131
6.1. Summary and discussion.....	131
6.2. Implications, limitations, and future studies	137
REFERENCES	145

LIST OF TABLES

Table 1. Facilitating factors for phonetic accommodation	18
Table 2. Numbers of talkers for all conditions and groups for the phonetic accommodation experiment.....	43
Table 3. Native and foreign names used in the IAT	49
Table 4. Positive and negative adjectives used in the IAT	50
Table 5. Two sets of English monosyllabic words	60
Table 6. Model talker average values for each acoustic measurement type	63
Table 7. Additional fixed effect factors and multicollinearity condition number for each linear mixed effects regression model	64
Table 8. Additional fixed effect factors and multicollinearity condition number for each regression model	82
Table 9. Two sets of disyllabic words	95
Table 10. Model talker average values for each acoustic measurement type	98
Table 11. Additional fixed effect factors and multicollinearity condition number for each regression model	99
Table 12. Additional fixed effect factors and multicollinearity condition number for each regression model	107
Table 13. Two sets of sentences	115
Table 14. Overall phonetic accommodation patterns	132

LIST OF FIGURES

Figure 1. Schematic description of acoustic measurements from the phonetic accommodation experiment.....	36
Figure 2. Schematic description on the experiment procedure for each group.	44
Figure 3. VOT accommodation patterns for exposed and unexposed monosyllabic words in <i>same-L1-same-dialect</i> , <i>same-L1-different-dialect</i> , and <i>different-L1</i> conditions.....	66
Figure 4. Vowel duration accommodation patterns on exposed and unexposed monosyllabic words in the <i>same-L1-same-dialect</i> , <i>same-L1-different-dialect</i> , and <i>different-L1</i> conditions.....	70
Figure 5. F0-max accommodation patterns on exposed and unexposed monosyllabic words in the <i>same-L1-same-dialect</i> , <i>same-L1-different-dialect</i> , and <i>different-L1</i> conditions.....	72
Figure 6. F1 accommodation patterns on exposed and unexposed monosyllabic words in the <i>same-L1-same-dialect</i> , <i>same-L1-different-dialect</i> , and <i>different-L1</i> conditions.....	75
Figure 7. F2 accommodation patterns on exposed and unexposed monosyllabic words in the <i>same-L1-same-dialect</i> , <i>same-L1-different-dialect</i> , and <i>different-L1</i> conditions.....	78
Figure 8. VOT accommodation patterns in the <i>different-L1</i> condition	83
Figure 9. Vowel duration accommodation patterns in the <i>different-L1</i> condition.....	85
Figure 10. F0 accommodation patterns in the <i>different-L1</i> condition	87
Figure 11. F1 accommodation patterns in the <i>different-L1</i> condition	89
Figure 12. F2 accommodation patterns in the <i>different-L1</i> condition	91
Figure 13. V2/V1 duration ratio accommodation for exposed and unexposed disyllabic words in the <i>same-L1-same-dialect</i> , <i>same-L1-different-dialect</i> , and <i>different-L1</i> conditions.....	101

Figure 14. V2/V1 F0 ratio accommodation for exposed and unexposed disyllabic words in the <i>same-L1-same-dialect</i> , <i>same-L1-different-dialect</i> , and <i>different-L1</i> conditions.....	103
Figure 15. V2/V1 amplitude ratio accommodation for exposed and unexposed monosyllabic words in the <i>same-L1-same-dialect</i> , <i>same-L1-different-dialect</i> , and <i>different-L1</i> conditions....	105
Figure 16. V2/V1 duration ratio accommodation patterns in the <i>different-L1</i> condition	109
Figure 17. V2/V1 F0 ratio accommodation patterns in the <i>different-L1</i> condition	110
Figure 18. V2/V1 amplitude ratio accommodation patterns in the <i>different-L1</i> condition	112
Figure 19. Results of XAB perception test	122
Figure 20. Full sentence DTW accommodation in the <i>same-L1-same-dialect</i> , <i>same-L1-different-dialect</i> , <i>different-L1</i> conditions.....	123
Figure 21. Hum DTW accommodation in the <i>same-L1-same-dialect</i> , <i>same-L1-different-dialect</i> , <i>different-L1</i> conditions.....	125
Figure 22. Scatterplots of z -normalized full sentence DTW and XAB perception test results (top) and scatterplots of z -normalized hum DTW and XAB perception test results (bottom) in the <i>same-L1-same-dialect</i> , <i>same-L1-different-dialect</i> , and <i>different-L1</i> conditions.....	127
Figure 23. Schematic plots of perceived phonetic convergence predicted by hum DTW convergence and full sentence DTW convergence in the <i>same-L1-same-dialect</i> , <i>same-L1-different-dialect</i> , and <i>different-L1</i> conditions	129

1. INTRODUCTION

1.1. Background

When we hear other people's speech, we are often exposed to phonetic details of the speech that may be different from our own speech and from speech patterns that we have encountered previously. This is because the acoustic-phonetic realizations in the speech are always variable both within and across talkers. Does the speech variation that we experience in our everyday lives influence our own speech production in some way? That is, do we speak somewhat differently after hearing speech that differs from our own speech or from our previous experience? If so, do we change our speech all the time or do we do so differently to different talkers or in different situations? Moreover, how do we judge whether talkers modified their speech in response to hearing other talkers? The current study investigates how native talkers accommodate their speech styles after hearing a native or nonnative talker and how we can measure their potential speech accommodation.

Many previous studies have found evidence that speakers change their speech production in response to variability in the speech input, and have referred to this phenomenon by various names, such as accommodation (e.g. Babel, 2009; Kim, Horton, & Bradlow, 2011; Namy, Nygaard, & Sauerteig, 2002; Shepard, Giles, & Le Poire, 2001), convergence (e.g. Natale, 1975; Pardo, 2006), phonetic imitation (e.g. Babel, 2012; Nielsen, 2011), alignment (e.g. Kraljic, Brennan, & Samuel, 2008), or mimesis (Delvaux & Soquet, 2007). Among these terms, accommodation, drawn from Communication Accommodation Theory (CAT) (Giles, Coupland, & Coupland, 1991; Shepard et al., 2001), is the most inclusive term covering all three possible categories of accommodation: convergence, maintenance, and divergence. In fact, phonetic maintenance and divergence have been observed (e.g. Babel, 2010; Bourhis & Giles, 1977;

Gallois, Giles, Jones, Cargile, & Ota, 1995; Kim et al., 2011) as well as phonetic convergence (e.g. Babel, 2012; Delvaux & Soquet, 2007; Goldinger & Azuma, 2004; Honorof, Weihing, & Fowler, 2011; Nielsen, 2011; Pardo, 2006; Shockley, Sabadini, & Fowler, 2004). Thus, in the current study, we use accommodation as our main term to refer to talker changes in speech production after particular speech perception situations.

1.1.1. Theoretical accounts of phonetic accommodation

One theoretical framework for explaining the mechanism of phonetic accommodation is the interactive alignment account of dialogue (Garrod & Pickering, 2007a, 2007b, 2009; Pickering & Garrod, 2004). In this account, the language perception and production systems are viewed as sharing the same representation within individual talkers (the parity between comprehension and production or the perception-behavior link). Also, the purpose of a successful conversation, a joint activity between interlocutors (Clark, 1996), is to align the interlocutors' situational models, and this alignment is assumed to occur automatically. When conversation occurs between interlocutors, the alignment of their situational models begins. As part of this process, the alignment of their language perception and production systems occurs through the parity between comprehension and production. Importantly, alignment at one level (e.g. language perception and production) percolates up to other levels (e.g. situational model) during conversation. Thus, in this view, linguistic representations of talkers are automatically aligned as are their mental representations over the course of a conversation.

Despite the intuitive appeal of this account, there is a substantial body of evidence against the claim of automatic alignment. It does not occur in all interactions, and it does not seem to occur to the fullest degree in every case. For example, talkers do not consistently show alignment

to certain types of talkers, such as talkers from different dialectal backgrounds (Bourhis & Giles, 1977; Kim et al., 2011) or nonnative talkers (Kim et al., 2011), and talkers show selective patterns of convergence to certain vowels (Babel, 2012), certain direction of phonetic change (Nielsen, 2011), or certain conversational roles (Pardo, 2006). From a theoretical point of view, Costa, Pickering, and Sorace (2008) propose that interactions that involve nonnative talkers might induce more non-automatic or conscious processing than interactions between native talkers, and this can impair the alignment process between the interlocutors. Thus, an important omission from the dialogue model in this interactive alignment account is any mediating constraint based on linguistic/psychological/social factors that are important aspects of the relationship between the interlocutors.

As an alternative account, the Communication Accommodation Theory (CAT) (Giles et al., 1991; Shepard et al., 2001) views speech accommodation as reflecting talkers' intentions to adjust their social distance from their interlocutors. Talkers converge towards their interlocutors because they want to be socially closer to them, while they diverge from their interlocutors or maintain their own speech styles to be distant from the interlocutors or to distinguish themselves from the other group. Therefore, in CAT, phonetic accommodation processes mainly depend on talkers' social motivations. Evidence that supports this view includes a travel agent's phonological convergence to her customers with various sociolinguistic backgrounds when she desired approval from the customers and better communicative efficiency (Coupland, 1984), homestay hosts' convergence to nonnative guests in terms of reusing the guests' jokes and idiom translations in their cooperative conversations (Burt, 1998), and nonnative English talkers' convergence or divergence in /r/, /l/, and word-final /z/ to an interlocutor with the intent of ethnic threat depending on the degree of their ethnic identification (Zuengler, 1982).

Defending themselves from the criticism mentioned above, Pickering and Garrod (2004) explain that their notion of automaticity is conditional automaticity developed for social cognition (e.g. Dijksterhuis & Bargh, 2001). That is, phenomena called “automatic” actually require specific set of conditions, and the characteristics that are assumed to comprise automaticity, namely, unawareness, effortlessness, unintentionality, autonomousness, and uncontrollability, do not always co-occur. Rather, combinations of such characteristics vary across different cognitive processes. Specifically, the automatic alignment Pickering and Garrod (2004) argue for is assumed to occur at the post-conscious level (Bargh, 1989), so that talkers are aware of their alignment behavior and the possibility that other social or psychological factors intervene in the process is open. It is on this very point that the two theories, namely, CAT and the interactive alignment account, meet to explain phonetic accommodation. Both accounts converge on the idea that talkers can be aware of phonetic accommodation and that it can depend on the talkers’ intentions and various linguistic/psychological/social conditions, although the two theories still put their main foci on opposite directions (automatic alignment versus social distance control).

1.1.2. Intervening factors in phonetic accommodation

Previous studies have found various intervening factors in phonetic accommodation ranging from linguistic factors to social and procedural factors. Table 1 lists such factors and identifies some related papers depending on the criterion of whether the study found the factor facilitative to phonetic accommodation. Each type of intervening factors will be discussed in further detail below.

Table 1. Facilitating factors for phonetic accommodation

Type	Facilitating factor	Experiment setting		
		Passive exposure	Shadowing	Conversation
Linguistic	Close interlocutor language distance – L1 and dialect match			Kim et al. (2011)
	High phonetic talent			Lewandowski (2011)
	Low word frequency	Goldinger and Azuma (2004), Nielsen (2011)	Goldinger (1998)	
	High number of repetition	Goldinger and Azuma (2004)	Goldinger (1998)	
	Extended VOT	Nielsen (2011)		
Social/ Psychological	Gender - Female participant		Namy et al. (2002)	
	Gender - Male pair			Pardo (2006)
	Close interlocutor relationship			Pardo et al. (2012)
	Positive attitude towards model talker		Babel (2009, 2010)	
	High attractiveness for gender mismatch		Babel (2009, 2012)	
	High social desirability			Natale (1975)
Situational	Talker role - Giver			Pardo (2006)
	Instruction to receiver to imitate giver			Pardo et al. (2010)
	Visual prompt for model talker voice		Babel (2009, 2012)	
Social x Situational	Woman giver, man receiver			Pardo (2006)

1.1.2.1 Linguistic factors

Some phonetic “distance” between interlocutors in conversations or between model talkers and participants in passive exposure or shadowing tasks is a prerequisite condition for phonetic accommodation to occur. In other words, some perceptible phonetic difference between

interlocutors before any contact is a prerequisite for speech styles accommodation which can then vary in magnitude and direction depending on various other conditions. Most previous studies utilized idiosyncratic or experimenter-manipulated phonetic distances within native English talker communities as the source of variation for phonetic accommodation to occur (Babel, 2009, 2012; Fowler, Brown, Sabadini, & Weihing, 2003; Goldinger, 1998; Goldinger & Azuma, 2004; Honorof et al., 2011; Namy et al., 2002; Natale, 1975; Nielsen, 2011; Pardo, 2006; Pardo, Jay, & Krauss, 2010; Shockley et al., 2004; Tilsen, 2009), but a few studies used naturally occurring dialectal variations (Alshangiti & Evans, 2011; Babel, 2010; Bourhis & Giles, 1977; Delvaux & Soquet, 2007; Kim et al., 2011; Kraljic et al., 2008) and native status differences (Kim et al., 2011; Lewandowski, 2011) as the source of phonetic variation. In most of the previous studies, English was used as the target language, while phonetic accommodation has also been observed in Korean (Kim et al., 2011), German (Lewandowski, 2011), and Belgium French (Delvaux & Soquet, 2007).

Of particular relevance to the present study, is the study by Kim et al. (2011) in which we tested the effect of linguistic distance between interlocutors as indexed by the same native language and the same dialects (closest distance), the same native language but different dialects (middle distance), and different native languages (farthest distance) on the degree of phonetic convergence. This study found that the degree of perceived phonetic convergence was negatively proportional to interlocutor language distance. An important feature of Kim et al. (2011) was the task used to set up the conditions for phonetic accommodation to occur. This study involved a task-oriented conversation in which two interlocutors were paired to perform a cooperative, conversation-based picture-matching task, the diapix task (Hazan & Baker, 2011; Lewandowski, 2011; Van Engen et al., 2010). The pair types were controlled to vary regarding native status

(native or nonnative speaker of the language of the interaction) and dialect match (if both speakers were native speakers of the target language): native-native with dialect match, native-native with dialect mismatch, and native-nonnative. Additionally, there was variation in English proficiency amongst the nonnative interlocutors. The results showed that the pairs where talkers shared their native status and the same dialects showed significant degrees of phonetic convergence, while the pairs with shared native status but with different dialects showed phonetic maintenance or divergence from their interlocutors. Also, in conversations where a native talker and a nonnative talker were paired, the native talkers generally did not converge to their nonnative partners, while some nonnative talkers converged to their native partners. Overall, these results indicated that interlocutor language distance negatively influenced the degree of phonetic convergence: the greater the linguistic distance, the less the convergence.

One exception to the native-to-nonnative maintenance pattern in Kim et al. (2011) was when the nonnative partner exhibited very high English proficiency. While this finding is hard to generalize beyond the data from Kim et al. (2011) because of the limited number of observations, it raised an important question regarding the interaction of two different aspects of linguistic distance. That is, can native talkers converge towards nonnative talkers with high target language proficiency? By investigating this question, we will be able to see whether nonnative talkers' high proficiency would be able to overcome the potential nonnative barrier and facilitate native talkers' phonetic convergence to a nonnative interlocutor.

Additionally, nonnative talkers' phonetic talent was found to facilitate their phonetic convergence towards native talkers (Lewandowski, 2011). Low word frequency also facilitated phonetic convergence (Goldinger, 1998; Goldinger & Azuma, 2004; Nielsen, 2011). The number of repetition of target linguistic items in a shadowing or passive exposure setting positively

influenced the degree of listeners' imitation of the items (Goldinger, 1998; Goldinger & Azuma, 2004). Moreover, native English talkers imitated extended VOTs but not reduced VOTs (Nielsen, 2011).

In sum, regarding linguistic factors, previous studies found that phonetic accommodation was facilitated by closer interlocutor language distance, phonetic talent, lower word frequency, higher number of repetition, and the increasing direction of change for VOT. However, we still do not know how native status mismatch and nonnative proficiency influence phonetic accommodation.

1.1.2.2 Social and Psychological factors

Social and psychological factors can also influence phonetic accommodation. Such factors are of importance in investigating accommodation in speech, as it is suggested by CAT that phonetic accommodation is a highly social behavior which talkers intentionally utilize to adjust their social distance to each other (Giles et al., 1991; Shepard et al., 2001), and by the interactive alignment account (Pickering & Garrod, 2004), alignment at one level facilitates alignment at different levels in interactive communications (for example, alignment at the psychological level can percolate to alignment at the phonetic level, or vice versa).

For example, closer personal relationship enhanced phonetic convergence to a roommate (Pardo, Gibbons, Suppes, & Krauss, 2012). Positive attitudes towards the model talker and social desirability were also found to facilitate phonetic convergence (Babel, 2009, 2010; Natale, 1975). Regarding gender, opposite results have been reported. That is, women were found to converge more to a model talker than men in a shadowing task in a study (Namy et al., 2002), while, in another study, male pairs showed higher rates of phonetic convergence (Pardo, 2006). Moreover,

high attractiveness of a male model talker facilitated female participants' convergence and inhibited male participants' convergence (Babel, 2009, 2012).

Among these, Babel (2009, 2010) are of particular interest for the current study. These studies measured participants' implicit attitudes towards the model talker identity (white participants' attitudes to a black model talker or New Zealand participants' attitudes to an Australian model talker, respectively), using the Implicit Association Task (Greenwald, McGhee, & Schwartz, 1998). In both studies, it was found that there was a significant relation between participants' degree of phonetic convergence and their social attitudes towards the model talkers. That is, the more positive attitudes participants had towards the model talkers, the larger the degree of convergence was. In other words, closer psychological attitudes between participants and the model talker can enhance their phonetic convergence towards the model talker.

While Babel's studies found a relationship between phonetic accommodation and attitudes towards different races or different cultures within a native talker community, in the current study we ask a similar question regarding native talkers' accommodation towards a nonnative model talker. That is, will native talkers exhibit different degrees of phonetic accommodation towards a nonnative model talker depending on their attitudes towards foreigners?

1.1.2.3 Situational factors

Different settings for the contact between interlocutors or between a model talker and the participants might also contribute variation to phonetic accommodation. Phonetic accommodation has been observed in many previous studies with two types of settings: conversation between interlocutors and auditory exposure to a model talker. The former line of

research typically involves a task-oriented conversational setting (e.g. Alshangiti & Evans, 2011; Gregory, Green, Carrothers, Dagan, & Webster, 2001; Kim et al., 2011; Pardo, 2006; Pardo et al., 2012; Pardo et al., 2010), while the latter line involves either shadowing of a model talker (e.g. Babel, 2009; Babel, 2012; Goldinger, 1998; Jungers & Hupp, 2009; Jungers, Palmer, & Speer, 2002; Namy et al., 2002; Shockley et al., 2004), or perceptual learning through training with feedback (e.g. Bradlow, Pisoni, Akahans Yamada, & Tohkura, 1997; Kraljic et al., 2008), or a passive auditory exposure task to a model talker (e.g. Delvaux & Soquet, 2007; Goldinger & Azuma, 2004; Jungers & Hupp, 2009; Nielsen, 2011). These different experimental settings by themselves reveal different possible situational factors for phonetic accommodation: communication between interlocutors in a social setting, shadowing of a model talker, being trained to imitate a model talker, and simply hearing a model talker.

One important difference among these different settings for phonetic accommodation might be the existence of socially intervening factors. In conversations, there is a common goal between interlocutors, that is, to perform a joint action of communication (Clark, 1996), and this can enhance phonetic convergence between the talkers. However, as we discussed above, this joint action can also be facilitated or inhibited by the interlocutors' attitudes to each other or their personal relationship. In a shadowing task, talkers have to read materials right after they hear the model talker. This removes the potential social influences from interactions with the model talker, but the requirement of production immediately following perceptual exposure may give participants a strong explicit motivation to accommodate to the model talker. On the contrary, passive auditory exposure to a model talker without any production practice or training may minimize any possible forcing or mediating factors to phonetic accommodation based purely on the situational aspect (i.e. the setting of the interaction). In other words, passive auditory

exposure allows us to explore phonetic accommodation based on a more fundamental speech perception-production connection in the absence of factors that come into play with direct interpersonal interaction.

Within these different experiment settings, there can be additional situational variations both in the conversation and passive exposure settings. For example, talker roles between interlocutors in a conversation can vary. Information-givers tended to converge to information-receivers more than the receivers to the givers (Pardo, 2006; Pardo et al., 2010). Furthermore, in female pairs, receivers converged more to the givers than givers to the receivers, while there was the opposite tendency for male pairs (Pardo, 2006). Instruction to a receiver to imitate the giver facilitated phonetic convergence in the conversation (Pardo et al., 2010). Additionally, a more social setting in a shadowing task, namely, displaying model talkers' face photos with the voices, increased convergence rates of the participants (Babel, 2009, 2012).

Summing up, researchers have found that phonetic accommodation is generally facilitated by closer linguistic and social or psychological distances between talker pairs, in terms of dialect match, positive attitudes, and closer relationship. However, we still lack solid evidence for the possibility of native talkers' accommodation towards nonnative model talkers and for the possible influence of native talkers' general attitudes towards foreigners on the native talkers' accommodation towards a nonnative model talker. Moreover, passive auditory exposure seems to be a good setting to ask these questions without any other socially intervening factors.

1.1.3. Generalizability of phonetic accommodation

For phonetic accommodation to lead to persistent language learning or accent change, the model-talker induced change should generalize to previously unexposed (i.e. new) linguistic items.

Nielsen (2011) directly investigated this question with VOT manipulations for the English phoneme /p/. After passively hearing English words that started with /p/ with an extended VOT, participants produced significantly longer VOTs in the posttest than in the pretest. Importantly, they did so not only for the words that they heard during the exposure phase, but also for new words with the target phoneme /p/ and even for the new phoneme /k/. Therefore, the VOT imitation could be said to have generalized to new words within the same phonemic category and also to a new phoneme within the same voicing category (voiceless) and with the same manner of articulation (stop consonant). This suggests that the learning through this phonetic accommodation procedure became a part of the participants' linguistic system.

While Nielsen (2011) tested generalizability of phonetic accommodation only on a single acoustic cue in isolated words and only with native English talkers, in the current study, we extend our examination to more linguistic levels and to talker-pairs that involve both native and non-native model talkers. That is, we ask whether phonetic accommodation generalizes to novel items not only with respect to VOT of voiceless, word-initial stops but also with respect to other acoustic cues in words and sentences. Also, can native talkers accommodate to both another native talker's and a nonnative talker's accents and generalize this accommodation to new linguistic items?

1.1.4. Measurements on phonetic accommodation

Now, how do we know whether phonetic accommodation occurs in speech? There have been various ways to analyze and measure phonetic accommodation patterns in speech. One way to do so is by using global judgments by human listeners. In an XAB or AXB perception test, researchers have asked listeners to select between pretest and posttest tokens (A or B), recorded

before and after perceptual exposure to or shadowing of the model talker, as a better match to model talker samples (X) (Goldinger, 1998; Goldinger & Azuma, 2004; Namy et al., 2002; Shockley et al., 2004). In studies of phonetic accommodation during recorded conversations with no pretest or posttest recordings, the X, A and B samples all come from conversational snippets extracted from early or late points in the conversation (Alshangiti & Evans, 2011; Kim et al., 2011; Pardo, 2006; Pardo et al., 2012; Pardo et al., 2010). In other studies, listeners were asked to rate regional accent change or the degree of regional accentedness of the talkers on gradient scales (Alshangiti & Evans, 2011; Giles, 1973). Because these tests are based on human listeners' holistic perceptions that are assumed to be based on many different factors at the same time, they have provided us with reliable answers on the matter of which of phonetic convergence, maintenance, and divergence occurs in speech, and what the degree of that phonetic accommodation pattern is.

However, these perceptual judgments do not suggest which features of the speech have actually been affected in the phonetic accommodation process. In Kim et al. (2011), as the speech materials investigated were taken from unscripted conversations, systematic acoustic measurement on linguistic features could not be conducted, rather judgments on phonetic accommodation patterns were obtained only through separate XAB perception tests. Therefore, although we found reliable evidence for phonetic accommodation in native-native or native-nonnative pairs and the positive influence of dialect match on phonetic convergence, we could not track down which acoustic features were involved in the phenomena. That is, what exactly changed acoustically or phonetically when talkers accommodated their speech to their interlocutors? What are the linguistic features that listeners pick up on when they judge phonetic accommodation?

Some researchers have also performed various acoustic analyses to parse out the various acoustic-phonetic dimensions along which phonetic accommodation occurs. Acoustic-phonetic features of words that have been analyzed for phonetic accommodation include VOT of the initial consonant of English words (Fowler et al., 2003; Nielsen, 2011; Shockley et al., 2004), allophonic variations of /l/ in American English (Honorof et al., 2011), the first and second formants, F1 and F2, of vowels in English words (Babel, 2009, 2010, 2012; Delvaux & Soquet, 2007; Pardo, 2010; Pardo et al., 2012; Pardo et al., 2010; Tilsen, 2009), the fundamental frequency, F0, of vowels in English words (Babel & Bulatov, 2011; Goldinger, 1997; Pardo, 2010), vowel duration of English words (Pardo, 2010; Pardo et al., 2012), and articulation rate (Pardo et al., 2010). For a broader linguistic range, researchers have found that the F0 band beneath 500 Hz in recordings of conversations (Gregory, Dagan, & Webster, 1997; Gregory et al., 2001; Gregory & Webster, 1996; Gregory, Webster, & Huang, 1993) and average speech rate of scripted and unscripted paragraph recordings (Jungers & Hupp, 2009; Jungers et al., 2002) are acoustic targets of phonetic accommodation.

One problem that arises from these previous studies is that most of them measured a single or a small number of acoustic cues, and such acoustic measurements often exhibited varying results across different acoustic cues, different talkers, or different linguistic items. Therefore, it is hard to generalize such findings beyond the given situations. Pardo et al. (2012) measured duration and vowel formants of English words and found that the measurements did not converge on the same overall conclusion for accommodation, and different roommate pairs used different sets of acoustic cues for phonetic accommodation to their roommates. Babel (2010) and Babel (2012) found that, in a shadowing task, only the formants of /æ/ and /ɑ/ in the English vowel system were imitated by native English talkers, but not those of /i/, /o/, and /u/. Most

importantly, when related with human listeners' holistic judgments of accommodation, these acoustic analyses often showed no significant correlation to the perceived accommodation patterns. For example, Pardo (2010) revealed that F0 and vowel duration were not related to perceived convergence patterns found in Pardo (2006). Pardo et al. (2010) and Pardo et al. (2012) reiterated the conclusion that acoustic measurements, namely, articulation rate or word duration and vowel formants, did not adequately explain listeners' judgments of accommodation. Babel and Bulatov (2011) also found that F0 convergence patterns were not correlated with perception of phonetic accommodation.

What would be a possible reason for the inconsistency between accommodation patterns found by acoustic analyses and those found by listeners' perceptual judgments? One point to note in this regard is an emphasis in reported research on segmental features of English words, (for example, consonants and vowels of English words), with some lesser amount of attention in the literature to global features of sentences, paragraphs, or conversations. Actually, possible acoustic features of speech that are influenced by phonetic accommodation might vary from words to sentences, and might include both segmental features and suprasegmental features of speech.

Because phonetic accommodation has been analyzed mostly either with human holistic perceptual judgments or with isolated segment-level acoustic features of isolated word productions in the previous studies (including Kim et al. (2011)), it has been difficult to reliably identify the acoustic parameters along which phonetic accommodation operates. These limitations of previous studies motivated the current study where we hoped to gain further insight into three main aspects of phonetic accommodation: first, the general relationship between phonetic accommodation and talkers' linguistic distances and psychological attitudes to

model talkers, second, generalizability of phonetic accommodation with such intervening factors, and finally, acoustic realizations of phonetic accommodation and their relationship to human holistic judgments.

1.2. The current study and the outline of the paper

As a follow-up study of Kim et al. (2011), the current research investigated phonetic accommodation by female native English talkers following passive auditory exposure to either native or nonnative female model talkers with new measurements. The passive auditory exposure setting was chosen to overcome one of the methodological limitations of Kim et al. (2011), namely that phonetic accommodation could not be measured acoustically with the highly variable spontaneous, conversational data. Therefore, in the current study, instead of active interaction with another talker, participants were exposed to recordings of the model's speech. Their own speech productions were recorded before and after the auditory exposure phase. To capture acoustic and phonetic characteristics of speech accommodation at various linguistic levels, English monosyllabic and disyllabic words and sentences were chosen as the speech materials for the experiment. In this way, we could measure various acoustic features in controlled linguistic items from low-level segmental features to high-level suprasegmental features of speech accommodation. Human listeners' judgments on accommodation were also gathered on part of the sentence data.

We asked the three questions mentioned above with these measurements. The first question is about two potential intervening factors on phonetic accommodation: linguistic talker distances and psychological implicit attitudes. The influence of linguistic talker distances on phonetic accommodation was investigated with both group level linguistic distances and item

level linguistic distances. For the group level linguistic distances between talkers and their model talkers, three types of linguistic distance variations were developed: *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1*. The *same-L1-same-dialect* distance was made when the native participants were exposed to a native model talker with the same dialectal background as themselves (a US Northern dialect). The *same-L1-different-dialect* distance was made with the native participants who had different dialects from their native model talkers. The *different-L1* distance was between the native participants and their nonnative model talkers. Importantly, the nonnative model talkers with very high English proficiency were chosen, following the hint found in Kim et al. (2011) that nonnative talkers' high proficiency might enhance the possibility of native talkers' convergence to them. For the item level talker linguistic distance, preexisting acoustic and phonetic differences between participants and their model talkers (the model talker value – the pretest participant value) on individual linguistic items were measured along various acoustic-phonetic dimensions. Note that the item level talker linguistic distances would have polarity, showing the direction of model talker-participant differences. Therefore, we can ask an additional question about the influence of the direction of linguistic talker differences on phonetic accommodation at the level of particular acoustic-phonetic dimensions (e.g. does F0 accommodation vary depending on whether the model talker has a higher or lower mean F0 than the participant?).

Generally, it is predicted that mismatch of native status or dialects between participants and their model talkers would inhibit phonetic convergence. This is because language processing can be less automatic in a farther language distance situation than in a closer language distance situation (Costa et al., 2008). This slower and more effortful language processing can make it hard for participants to accommodate their speech styles. Specifically, participants who have a

different L1 or different dialect from their model talker might adopt a clear speech strategy, which might inhibit phonetic convergence (Kim et al., 2011). This might be either to increase their speech intelligibility for the model talker or to emphasize their linguistic identity against their model talker. On the contrary, participants who share the same L1 and dialect with their model talker would experience more automatic language processing and might display a larger degree of phonetic convergence.

However, within a certain group level language distance condition, the preexisting acoustic and phonetic distance between talkers and their model talkers for individual acoustic dimensions might be positively correlated with the degree of phonetic convergence. This is because, unlike the large group level talker-model talker distances, at the finer-grained acoustic dimensions, a certain distance between talkers and their model talkers might actually be needed for any significant phonetic changes to occur. Therefore, when the distance is larger, there's more room for acoustic change, and it might be more likely that phonetic convergence occurs.

Importantly, we predict that participants would exhibit phonetic convergence in both directions of preexisting model talker-participant linguistic distances at the individual acoustic dimensions. That is, phonetic convergence can either involve increasing or decreasing values along the acoustic dimension, depending on where the model talker value is located compared to a participant's pretest value. However, we also predict that the degree of phonetic convergence might be higher in the direction of decrease for some acoustic dimensions. For example, durations such as VOT and vowel duration might be more likely to converge towards the decreasing direction. This is because the general direction of speech change when speech items are produced with repetition is reduction in duration and intelligibility (Baker & Bradlow, 2009; Fowler & Housum, 1987). Therefore, when a model talker value is smaller than a participant

value, convergence in this case can be enhanced, as it is in the same direction as the general repetition effect on speech.

The second potential intervening factor we tested is psychological distance between talkers. Here, we asked whether native talkers who were exposed to a nonnative model talker (the *different-L1* condition) would show more evidence of phonetic convergence when they have more positive social attitude towards foreign identity, in other words, when they had positive implicit attitudes towards their nonnative model talker. For this question, participants' implicit attitudes towards foreigners were analyzed, following the application of an implicit association task in Babel (2010) and Babel (2012). Then the relation between phonetic accommodation in the *different-L1* condition and their implicit attitudes towards foreigners was investigated. We predict that native talkers' positive attitude towards a nonnative model talker would positively affect their phonetic accommodation, since talkers' linguistic and psychological functions would be connected in the general cognitive system, and speech accommodation in one level might be permeated to the other level, according to Pickering and Garrod (2004).

Secondly, we tested whether phonetic accommodation patterns obtained through exposure to certain items can be generalized to new items. Nielsen (2011) found that participants generalized their production changes along the VOT dimension to words they did not hear during the auditory exposure phase. Similarly, we established two sets of speech materials, so that participants heard one of the two sets during the perceptual exposure phase and tested on their production changes for both of the two sets. If participants acquire robust learning through phonetic accommodation, their phonetic changes on new items would not be significantly different from those on old items, or the changes on new items would be in the same direction as those on old items.

Importantly, we tested the interaction between the generalization effect and the effects of linguistic distances and psychological attitudes on phonetic accommodation. That is, we investigate whether participants would exhibit their generalization effects differently in the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions. In addition, we asked whether participants would differentiate their generalization patterns depending on their attitudes from the model talkers. It is expected that, if any, the generalization effect would decrease as linguistic and psychological distances increase. This is because it would be hard to experience phonetic convergence in large linguistic and psychological distances, and it would be even harder to generalize the small degree of convergence to new items.

Lastly, we examined how various acoustic and perceptual measurements at different linguistic levels reveal phonetic accommodation. From the previous literature on phonetic convergence, we still do not know at which linguistic or acoustic levels phonetic convergence can be found, and how results from different measurements vary. Moreover, it is unknown whether perceived phonetic convergence patterns converge with acoustically measured phonetic convergence. For this research purpose, material sets with monosyllabic and disyllabic words and sentences were established. Specifically, segmental features such as VOT, vowel duration, F0, and formants of vowels were measured from monosyllabic words. Disyllabic words provided us with measurements of word-level stress pattern. On sentences, we applied two global analyses to measure phonetic accommodation: the dynamic time warping technique and perceived phonetic convergence patterns through an XAB perception test. This allowed for an additional investigation of the relationship between perceived phonetic accommodation and acoustically judged phonetic accommodation.

The current paper consists of seven chapters. First, in Chapter 1, the critical issues for the current research, relevant previous literature, important points of the experiment design, and the research questions were introduced. In Chapter 2, the general methodology for the whole experimental process is described. In Chapters 3, 4, and 5, detailed methods and results for monosyllabic words, disyllabic words, and sentences are fully explained. Chapter 6 concludes the dissertation with a summary and discussion of the results and their implications.

2. GENERAL METHODOLOGY

2.1. Definition of phonetic accommodation: revisited and controls

Previous studies that used AXB or XAB perception tests to investigate phonetic accommodation (Babel & Bulatov, 2011; Goldinger & Azuma, 2004; Kim et al., 2011; Namy et al., 2002; Pardo, 2006; Pardo et al., 2012; Pardo et al., 2010; Shockley et al., 2004) have defined phonetic convergence as decrease of the absolute distance between interlocutors or between the model and the participant at posttest compared to the model-participant distance at pretest.

Applying the same logic, some studies that introduced acoustic measurements to investigate phonetic accommodation (Babel, 2009, 2010; Pardo, 2010; Pardo et al., 2010) defined phonetic convergence as the difference between interlocutors or between the pretest-to-model absolute difference and the posttest-to-model absolute difference.

Figure 1 describes possible results of a phonetic accommodation experiment with acoustic measurements in a schematic manner. In this figure, phonetic convergence is determined in the area of B. When the posttest value is located between the positive and negative addition of the absolute difference of the model and pretest values, it is phonetic convergence. Otherwise, the result is divergence, as in the areas of A and C. That is, the absolute distance towards the model has increased in the posttest, compared to the pretest, in A and C.

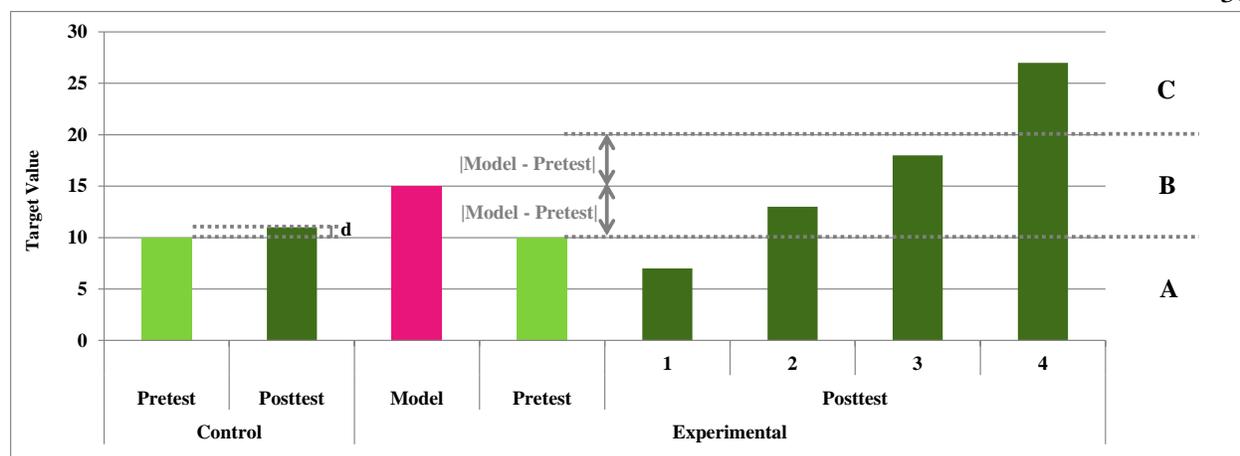


Figure 1. Schematic description of acoustic measurements from the phonetic accommodation experiment.

Note. d: the pretest-posttest difference in the control group

The pretest-posttest differences in the experimental conditions are all larger than d.

A: the area where the posttest value diverges from the model value compared to the pretest value
 B: the area where the absolute distance towards the model decreases in the posttest compared to the pretest

C: the area where the absolute distance towards the model increases in the posttest compared to the pretest but the pretest-to-posttest change is in the same direction as the pretest-to-model difference

However, it is questionable whether it should be defined as phonetic divergence when a posttest value is located in the area of C. When a posttest value is in the area of C, it means that the participant moved her production towards the model to an extent that exceeds the absolute distance between the pretest and model's production. In other words, this can be viewed as convergence with overshooting. In the current study, we redefine phonetic convergence to include the area of C in the definition of phonetic convergence, because a participant is unlikely to control the degree of convergence to always avoid any overshoot.

Additionally, none of the previous studies introduced control conditions. That is, there were no participants that were put under the condition of no model to imitate. Instead, all participants performed a phonetic accommodation task, and the data were interpreted by

comparing only the pretest and posttest responses in terms of their distances towards the model, either through an XAB perception test or through acoustic analyses. The problem of this setting is that it does not consider the potential repetition effect of the target linguistic items during the task. In other words, the pretest-posttest change could be drawn from simple repetition effects. This problem can be resolved by introducing a control condition where participants are exposed to the linguistic items the same number of times as in the experimental conditions but not auditorily. In the current study, participants in the control group were exposed to the materials visually by viewing the orthographies of the words and sentences while participants in the experimental groups were exposed to the materials auditorily. For any pretest-posttest changes in the experimental conditions to be taken as significant, they need to be significantly different from the pretest-posttest changes of the given measurement in the control group, as in Figure 1.

Therefore, the revisited definition of phonetic convergence with acoustic measurements has to fulfill the three conditions below:

1. The pretest is significantly different from the model.
2. The pretest-posttest difference in an experimental group is significantly different from that in the control group.
3. The direction of the pretest-posttest change is the same as the pretest-model difference.

The first condition is the prerequisite condition for any phonetic changes to occur. If there is no difference between the model and the pretest values, the pretest and the model are already aligned and there is no room for convergence or divergence to occur. If the first

condition is not met, that is, if the pretest and the model values are exactly the same, the data were not considered in the following analyses. The second condition is required to show that the effect in an experimental group is not merely a repetition effect of the same items during the task, but an actual effect of the exposure condition. If this condition is met, it is convergence or divergence, and if not, it is maintenance. Finally, the third condition determines whether a significant difference between pretest and posttest that passed the second condition is convergence towards the model or divergence from the model. If the pretest-posttest change is in the direction towards the model, it is convergence, and if not, it is divergence. Importantly, if the third condition is satisfied, this covers the case of overshoot for either convergence or divergence.

Based on the schematic definition of phonetic convergence above, a formula for a single dependent measure that can represent phonetic accommodation patterns of a given measurement was developed as below:

$$\text{Adjusted phonetic change} = ((\text{posttest} - \text{pretest})_{\text{expr}} - \text{average}(\text{posttest} - \text{pretest})_{\text{control}}) \\ \times (|\text{model} - \text{pretest}|_{\text{expr}} / (\text{model} - \text{pretest})_{\text{expr}})$$

Note. $_{\text{expr}}$ = values from an experimental group, $_{\text{control}}$ = values for control group, $_{\text{average}}$ = average of the following group of values

For this index of phonetic accommodation, first, posttest-pretest differences for each item for each talker in each experimental group and the control group were calculated $((\text{posttest} - \text{pretest})_{\text{expr}}$ and $(\text{posttest} - \text{pretest})_{\text{control}}$). Then the posttest-pretest differences for the talkers in the control group were averaged over each item (word or sentence) for each exposure condition (Set 1 or Set 2, to be explained further below) $(\text{average}(\text{posttest} - \text{pretest})_{\text{control}})$. Next, the averaged

posttest-pretest differences for each item and each exposure condition in the control group were subtracted from those in the experimental conditions ($(posttest - pretest)_{\text{expr}} - \text{average}(posttest - pretest)_{\text{control}}$). The steps up to this point were taken to adjust posttest-pretest differences in an experimental group, taking into consideration the averaged values in the control group. The resultant value was then multiplied by the polarity of the model-pretest difference of the experimental group ($((posttest - pretest)_{\text{expr}} - \text{average}(posttest - pretest)_{\text{control}}) \times (|model - pretest|_{\text{expr}} / (model - pretest)_{\text{expr}})$). This was to reflect the direction of the model-pretest difference in the final value. In this formula, if a final value is positive, it would mean that the change from pretest to posttest is towards the model, therefore, convergence, while if negative, it would mean that the pretest-posttest change is away from the model, therefore, divergence. Additionally, if the model-pretest difference is exactly 0, the formula above would not function. This condition also does not fulfill the first condition of the definition of phonetic accommodation given above, namely, the prerequisite participant-model talker distance. Therefore, such data points would have to be excluded from the following analyses. However, there were no model-pretest difference that was 0 in the acoustic analyses in Chapter 3 and 4, and no data points were excluded from the final datasets.

2.2. Experiments

All participants performed three tasks, namely, a phonetic accommodation experiment, an implicit association task, and a questionnaire. First, the phonetic accommodation experiment was conducted with auditory or visual exposure to native and nonnative speech before and after production tasks with the same speech materials. There were three levels of speech materials, namely, monosyllabic words, disyllabic words, and sentences. Second, all participants in the

phonetic accommodation experiment performed an implicit association task to measure their social attitudes towards native and foreign people in the United States. Lastly, all participants completed a questionnaire where they were asked whether they noticed the native status of their model talkers. It took approximately 2 hours for each participant to finish all the tasks.

Additionally, an XAB perception test was conducted with a separate set of participants on part of the sentence data taken from the phonetic accommodation experiment. Details of the XAB perception test and the results will be discussed in Chapter 5.

2.2.1. Phonetic Accommodation Experiment

2.2.1.1 Materials

Two sets of English words and sentences (Set 1 and Set 2) were established for the phonetic accommodation experiment (two sets of English words (W1 and W2) and two sets of English sentences (S1 and S2)). Set 1 consisted of W1 and S1, while Set 2 consists of W2 and S2. Within these sets, W1 consisted of MW1 and DW1 with 32 monosyllabic and 32 disyllabic words that start with bilabial stops (/b/ or /p/), while W2 consisted of MW2 and DW2 with a separate set of 31 monosyllabic and 31 disyllabic words that start with alveolar stops (/d/ or /t/). Each of S1 and S2 consisted of 32 sentences. Half of the 32 sentences for each set were high probability SVO sentences where the verb started with a bilabial stop (/b/ or /p/) in S1 and an alveolar stop (/d/ or /t/) in S2, and the verb and object shared coherent meanings. The other half of the 32 sentences for each set were high probability sentences that were selected from the Speech Perception in Noise test (Kalikow, Stevens, & Elliott, 1977). In total, 126 words and 64 sentences were used as materials for the phonetic accommodation experiment. The lists and further details of the

monosyllabic words, disyllabic words, and sentences will be discussed in Chapters 3, 4, and 5, respectively.

Two female monolingual native American-English talkers (N1 and N2) and two female nonnative English talkers whose native languages were Korean (NN1 and NN2) were recorded as model talkers for the phonetic accommodation experiment. The dialect of both of the two native talkers, N1 and N2, was classified as the US Northern dialect, according to Labov, Ash, and Boberg (2006). The two nonnative talkers were tested on their English proficiency, using The Versant™ English Test (www.versanttest.com). Both nonnative talkers showed high English proficiency, as their overall scores were 68 and 54 out of 80, respectively. Specifically, the nonnative model talkers had high scores on phonetically related subareas, namely, fluency (NN1: 69, NN2: 49, out of 80) and pronunciation (NN1: 64, NN2: 50, out of 80). As mentioned in Chapter 1, Kim et al. (2011) found that, out of eight native interlocutors who were paired with a nonnative talker for an English conversation task, only one native interlocutor who was paired with a high proficiency nonnative partner converged towards the partner. Therefore, by selecting these high proficiency nonnative talkers as model talkers, we expected a greater likelihood that native participants would converge towards them than towards lower proficiency nonnative talkers. All four model talkers were in their 20s with an average age of 22.75 years.

The model talkers read the complete set of words and sentences in random order in a sound booth. The words and sentences were presented in separate blocks on a computer monitor. The recordings were made using a computer through a Shure SM81 condenser handheld microphone with a sampling rate of 48000 Hz. The words and sentences were sliced from the recordings with the help of TriggerWave (Chan, 2009) and manually corrected, and finally normalized to have the same overall RMS value (1.0 Pa).

2.2.1.2 Participants

Sixty-seven female monolingual native American-English talkers with normal speech and hearing participated in the phonetic accommodation experiment. All participants were undergraduate or graduate students at Northwestern University. Their age ranged from 18 to 28 years with an average of 20.58 years. Out of these sixty seven talkers, twenty talkers were assigned to the control group, and forty seven talkers to the experimental groups. Out of the forty-seven talkers in experimental groups, thirteen were assigned to one native model talker, N1, ten to the other native model talker, N2, fourteen to one nonnative talker, NN1, and ten to the other nonnative talker, NN2. These subgroups were then each sub-divided into two conditions for perceptual exposure to different material sets (Set 1 or Set 2). In the control group, nine talkers were exposed to Set 1, and eleven talkers to Set 2. Among the thirteen talkers who were exposed to N1, seven talkers were exposed to Set 1, and six talkers to Set 2. Among the ten talkers who were exposed to N2, five talkers were exposed to Set 1, and five talkers to Set 2. Among the fourteen talkers who were exposed to NN1, six talkers were exposed to Set 1, and eight talkers to Set 2. Among the ten talkers who were exposed to NN2, five talkers were exposed to Set 1, and the other five talkers to Set 2.

Three groups of linguistic distance were made with these participants: *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1*. First, dialects of participants who were assigned to a native model talker were classified as a US Northern dialect or a non-Northern dialect, based on the regions they reported having lived in until 18 years of age. When a participant heard a native model talker, if she had a Northern dialect, this would make a condition of *same-L1-same-dialect*, while a participant with a non-Northern dialect would make

a condition of *same-L1-different-dialect*. Participants who were assigned to a nonnative model talker made the condition of *different-L1*.

In sum, there were one control group and four experimental groups, and within each group, there were two conditions for perceptual exposure to different material sets. Table 2 summarizes the numbers of participants for all conditions and groups for the phonetic accommodation experiment.

Table 2. Numbers of talkers for all conditions and groups for the phonetic accommodation experiment

Group	Linguistic distance	Stimuli set	Number of talkers (per model talker)
Control (n = 20)	(no model talker) (n = 20)	Set 1	9
		Set 2	11
Experimental (n = 47)	<i>Same-L1-Same-Dialect</i> (n = 13)	Set 1	7 (N1: 4, N2: 3)
		Set 2	6 (N1: 3, N2: 3)
	<i>Same-L1-Different-Dialect</i> (n = 10)	Set 1	5 (N1: 3, N2: 2)
		Set 2	5 (N1: 3, N2: 2)
	<i>Different-L1</i> (n = 24)	Set 1	11 (NN1: 6, NN2: 5)
		Set 2	13 (NN1: 8, NN2: 5)

Note. N = native model talker, NN = nonnative model talker.

2.2.1.3 Procedure

To rigorously examine the effect of auditory exposure on phonetic change, a control group where participants were exposed to stimuli in a visual manner not an auditory manner was included.

Participants were divided into two kinds of groups for different manners of perceptual exposure to the materials: 1) four experimental groups for auditory exposure to stimuli read by the four model talkers, and 2) a control group for visual exposure to stimuli written on a computer screen.

In other words, participants in experimental groups heard recordings of the words and sentences

read by one of the four model talkers (N1, N2, NN1, and NN2), while participants in the control group saw written forms of the words and sentences on the monitor for perceptual exposure. In this way, the number of linguistic exposures to the stimulus items was the same for both the experimental groups and the control group, either through auditory exposure or visual exposure. Also, to test whether any phonetic accommodation generalized to unexposed items, participants in each group were then divided into two conditions: 1) the condition where they were exposed to Set 1 and tested on all materials, and 2) the condition where they were exposed to Set 2 and tested on all materials.

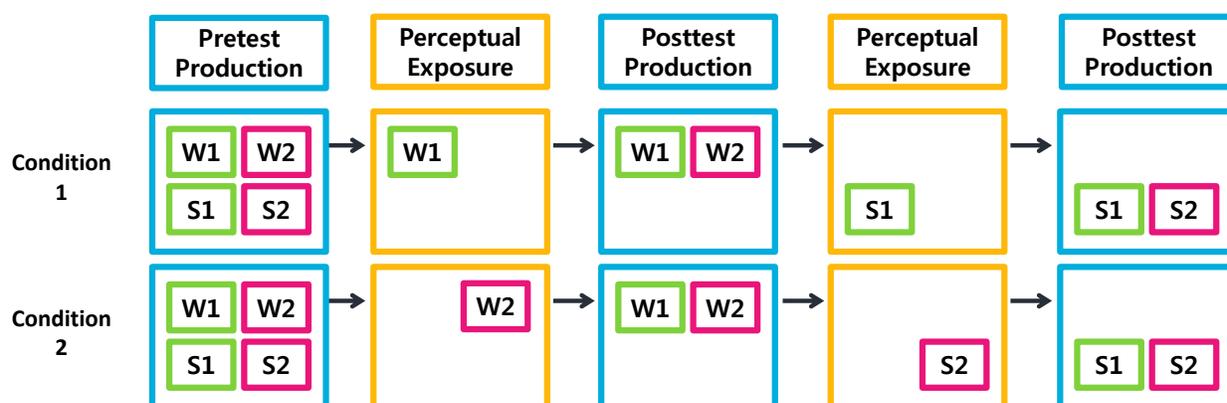


Figure 2. Schematic description on the experiment procedure for each group.

Note. 1. W1: word set 1, W2: word set 2, S1: sentence set 1, S2: sentence set 2.

2. Perceptual exposure is auditory exposure for experimental groups and visual exposure for control groups.

All participants followed five phases for the phonetic accommodation experiment (see Figure 2): 1) pretest production of all word and sentence items (Set 1 and Set 2), 2) auditory or visual exposure to either the word sets that start with bilabial stops (W1) or the word sets that start with alveolar stops (W2), 3) posttest production of all word sets (W1 and W2), 4) auditory

or visual exposure to one of the two sentence sets (S1 or S2), and 5) posttest production of the two sentence sets (S1 and S2). In detail, 1) in the pretest production phase, all participants in all groups were recorded reading all words and sentences out loud. Words and sentences were presented in different blocks with randomized orders on a computer monitor. 2) Then during the first exposure phase, participants in experimental conditions heard recordings of W1 or W2, read by one of the four model talkers (N1, N2, NN1, or NN2), with 9 repetitions of each word in random order. The inter-sample interval was 100 ms. On each trial, the participants heard a word and selected the critical item written in standard English orthography on a computer display that included the target item plus seven alternatives. The seven alternative words were chosen from the same set where the target word belonged, sharing the same place of articulation for the initial consonant and similar vowels with the target word. This item-identifying task was intended to encourage participants to focus on listening to the stimuli, but they were not given any direct task training or any feedback. Participants in the control conditions viewed orthographic representations of 9 repetitions of words taken from either W1 or W2, and did the same item-identifying task during the exposure phase, with no auditory stimulation and no feedback. 3) In the first posttest phase, all participants in all groups read all words in W1 and W2 in random order from a computer monitor again. 4) During the second exposure phase, participants were exposed to one of the two sentence sets (S1 or S2). Participants in experimental groups heard 9 repetitions of recordings of the sentences from a sentence set read by one of the four model talkers (N1, N2, NN1, or NN2) in random order, while participants in control groups viewed 9 repetitions of the sentences from a set in written forms. Again, the inter-sample interval was 100 ms. On each trial, participants did the same item-identifying task explained above with sentence items instead of words. In each trial, first, participants heard a sentence through headphones.

Then, the latter parts of a target sentence and seven alternatives excluding the first few words of the sentences were displayed on the computer monitor. This was to avoid the case where participants might listen to the first words of the sentence stimulus and not concentrate for the rest of the sentence. Then participants were instructed to click on the item that matched with the sentence they just heard. The seven alternatives were chosen from the same sentence set where the target sentence belonged. Participants in the control group, instead, viewed a target sentence first. Then, in the same way from the experimental groups, the latter parts of eight sentences excluding the a few first words were displayed on the monitor, and participants were instructed to select an item that matched the sentence they just viewed. Again, there was no explicit task training or any feedback. 5) Finally, all participants in all groups read all sentences for both sentence sets, S1 and S2 out loud in random order from a computer monitor. In all reading phases, 1), 3), and 5), all readings were recorded to a separate computer through a Shure SM81 Condenser Handheld microphone with the sampling rate of 48000 Hz. In the hearing phases, 2) and 4), participants in the experimental groups heard stimuli through Sony MDRV700 headphones. For the display of the audio and visual stimuli during the hearing phases, Millisecond Inquisit 3.0.4.0. was used.

2.2.1.4 Analyses

All recordings were first automatically segmented at the phoneme and word levels using NUALigner 2 (Chan, 2010) which was based on the Penn Phonetics Lab Forced Aligner (Yuan & Liberman, 2008). Automatic segmentations were then manually corrected, using Praat (Boersma & Weenink, 2010). Then, target acoustic measurements were obtained by running Praat (Boersma & Weenink, 2010) scripts. Additionally, dynamic time warping analyses using

MATLAB (2010) codes (Ellis, 2003, 2005) on the sentence dataset from the experimental groups and XAB perception tests on part of the sentence data were conducted. The details of acoustic measurements and data analyses for monosyllabic words, disyllabic words, and sentences will be discussed in Chapters 4, 5, and 6, respectively.

2.2.2. Implicit Association Task (IAT)

First described in Greenwald et al. (1998), the IAT is a standard measure for implicit biases in social psychology. In the task, participants are instructed to rapidly classify words or pictures into four categories (two for concept discrimination and two for attribute discrimination) by pressing either of a right key or a left key on the keyboard. Critically, the two attribute categories represent “good” and “bad” or “positive” and “negative”. The two concept categories share the same response key with different attribute categories in different blocks. For example, for concepts A and B, in one block, concept A and “good” share the same response key, while concept B and “bad” share the other response key. In the next block, in turn, concept B and “good” share the same response key, and concept A and “bad” share the other response key. Then participants’ response latencies are calculated for the two blocks and calculated for the participants’ final implicit attitude scores towards the two concepts. The basic logic is that when the average response latency for an association is shorter than that for the other association, it suggests the participant’s more automatic processing towards the quicker association than towards the slower association. Babel (2010) and Babel (2012) applied this measure to phonetic accommodation research and found that talkers with more positive implicit attitudes towards the model talker exhibited a larger degree of phonetic convergence to the model talker. In the current study, we compared participants’ response latencies for two different associations, “native-

positive/foreign-negative” and “foreign-positive/native-negative”, aiming at measuring participant’s implicit attitudes towards foreigners, compared to their attitudes towards native people.

2.2.2.1 Stimuli

Twenty four names that would likely to represent a typical US native person were selected from lists of the twenty five most popular American male names and twenty five most popular American Female names, based on the US Census 1990. Another set of twenty four names that would likely to represent a foreign person were selected from European, Middle East, African, and Asian names. For both of the US native and foreign name sets, equal numbers of male and female names were selected. The complete list of US native and US foreign names are given in Table 3.

Table 3. Native and foreign names used in the IAT

Gender	Native	Foreign	Area
Male	James	Pierre	European
	John	Klaus	
	Robert	Luciano	
	Michael	Bjorn	
	William	Igor	
	David	Hasan	Middle
	Richard	Yonatan	East
	Charles	Aditya	African
	Joseph	Azizi	
	Thomas	Wei	Asian
Christopher	Daichi		
Daniel	Minjun		
Female	Mary	Aurelie	European
	Patricia	Claudia	
	Linda	Giovanna	
	Barbara	Ingrid	
	Elizabeth	Ivanna	
	Jennifer	Khadija	Middle
	Maria	Yardena	East
	Susan	Ekta	African
	Margaret	Zahra	
	Dorothy	Fang	Asian
Lisa	Chihiro		
Nancy	Mikyong		

Also, twenty four semantically positive adjectives and twenty five negative adjectives for people's personal characters were chosen for the IAT by the author. All adjectives are given in Table 4.

Table 4. Positive and negative adjectives used in the IAT

Positive	Negative
Sweet	Unfriendly
Friendly	Rule-bound
Intelligent	Unsociable
Sociable	Passive
Active	Personality-less
Creative	Dull
Comfortable	Dogmatic
Open-minded	Mean
Flexible	Uninteresting
Interesting	Uncreative
Relaxed	Isolated
Loyal	Sour
Honest	Ill-tempered
Accomplished	Selfish
Smart	Cruel
Successful	Boring
Funny	Thoughtless
Clear	Uncomfortable
Easy-going	Intimidating
Generous	Struggling
Approachable	Arrogant
Kind	Irritable
Passionate	Bitter
Bright	Ignorant

2.2.2.2 Participants

All of the sixty seven participants for the phonetic accommodation experiment participated in the IAT immediately after the phonetic accommodation experiment. This decision was made to make the experimental conditions as similar as possible for all participants in the control group and with native and nonnative model talkers. However, IAT scores were of interest only for the data from participants to a nonnative model talker. Therefore, only the IAT scores of participants in the *different-L1* condition were analyzed for the current study.

2.2.2.3 Procedure

To measure participants' implicit attitudes towards US native and foreign people, the most recent procedure for IAT, introduced in Greenwald, Nosek, and Banaji (2003), was implemented in the current study, using Millisecond Inquisit 3.0.4.0. Participants sat in front of a computer monitor and were instructed to press a left key ("E") or a right key ("I") as quickly as possible and with as few mistakes as possible. This task was to categorize an item that was located at the center of the monitor with the category shown at the upper left side of the monitor or with the category shown at the upper right side of the monitor. "X" was shown at the center of the monitor whenever a participant made an error all through the task. The task consisted of seven blocks: Block 1 for attribute practice, Blocks 2, 3, 4, for one practice block and two test blocks for one of the "native-positive/foreign-negative" association or "native-negative/foreign-positive" association, and Blocks 5, 6, and 7 for one practice block and two test blocks for the other of the two associations. The order that the two kinds of associations were tested was counterbalanced across participants.

For example, let's see the case where the "native-positive/foreign-negative" association was tested before the "native-negative/foreign-positive" association. Block 1 was a practice block with twenty trials for side-attribute associations. An adjective was displayed at the center of the monitor, while the category "positive" was displayed at the upper left side of the monitor, and the category "negative" at the upper right side. Participants categorized a given adjective on the monitor either as "positive" or "negative". Ten positive adjectives and ten negative adjectives were randomly selected from the complete adjective list (see Table 4) and given in a random order. Block 2 was a practice block with twenty trials for side-name associations. A name was displayed at the center, while the "native" category was displayed at the left upper corner of the

monitor, and the “foreign” category at the right upper corner. Participants categorized a given name either as “native” or “foreign”. Ten US native names and ten US foreign names out of the complete name list (see Table 3) were randomly selected and given in a random order. Then Block 3 was a short test block with twenty trials for side-name-attribute associations. That is, either a name or an adjective was displayed at the center of the monitor, while “native” was displayed with “positive” at the left upper corner, and “foreign” was displayed with “negative” at the right upper corner. Participants categorized a given name as “native” or “foreign” and a given adjective as “positive” or “negative”. Five US native names and five US foreign names were randomly selected from the complete name list, and five positive adjectives and five negative adjectives were randomly selected from the complete adjective list. The names and adjectives were given every other time in random orders for each set. Block 4 was a long test block for side-name-attribute associations. The same procedure in Block 3 was conducted on forty trials. Ten US native names, ten US foreign names, ten positive adjectives, and ten negative adjectives were randomly selected from the lists. Block 5 was a practice block with twenty blocks for the other side-name associations. Here, a name was displayed at the center, and now the “foreign” category at the left upper corner, and the “native” category at the right upper corner. Participants categorized a given name as “native” or “positive”. Twenty randomly picked US native and foreign names were used as stimuli. Block 6 is a long test block with forty trials for the “foreign-positive/native-negative” associations. The “foreign” category was displayed with “positive” at the left upper corner, and the “native” category was displayed with “negative” at the right upper corner. Participants categorized a given name at the center of the monitor as either “native” or “foreign”, and a given adjective as either “positive” or “negative”. Ten native names, ten foreign names, ten positive adjectives, and ten negative adjectives were randomly picked from the lists.

Names and adjectives were given every other time. Finally, Block 7 is a short test block with twenty trials for the same associations and procedure used in Block 6. Five US names, five foreign names, five positive adjectives, and five negative adjectives were randomly selected from the lists. In the other case where the “foreign-positive/native-negative” associations were tested first, the orders of Blocks 2, 3, 4, and Blocks 5, 6, 7 were switched.

2.2.2.4 Analyses

IAT scores were calculated for the 24 participants to a nonnative model talker with the most recent IAT formula introduced in Greenwald et al. (2003). Response latency data gathered from test blocks (Blocks 3, 4, 6, and 7) were used to determine IAT scores for each participant. Trials with latencies over 10000 ms were eliminated from the dataset. Then in the four blocks, each error latency value within a block was replaced with the value of 600 ms added to the mean of correct latency values in the block. After this error correction, the resulting latency values for each of the four blocks were averaged. Then two differences were calculated: the average latency value of the short test block for the “foreign-positive/native-negative” associations minus the average latency value of the short test block for the other set of associations, and the average latency value of the long test block for the “foreign-positive/native-negative” associations minus the average latency value of the long test block for the other set of associations. Each of the two differences was then divided by their associated pooled standard deviation. By averaging these two quotients, the final IAT score for a certain participant was calculated. If the latencies in the blocks for the “foreign-positive/native-negative” associations are smaller than the latencies in the blocks for the “native-positive/foreign-positive” associations, the final score would be negative, and vice versa.

Therefore, a final IAT score can be interpreted to be negatively proportional to the participant's positive attitude towards US foreign people. That is, when a participant's IAT score is over 0, her attitudes towards US foreign people is assumed to be generally negative, while an IAT score below 0 would indicate that her attitude towards US foreign people is positive. Also, when a participant's IAT score is higher than others, her attitude towards US foreign people is assumed to be more negative than others, and vice versa.

The final IAT scores from the total 67 participants ranged from -0.17 to 1.49 with the average of 0.57 and the standard deviation of 0.38. Within the total set, the final IAT scores for the 24 participants to the nonnative model talkers (the *different-L1* condition) ranged from 0.05 to 1.42 with the average of 0.68 and the standard deviation of 0.37. This indicates that none of the participants with the nonnative model talkers had positive attitudes towards foreigners, while the degree of negative attitudes varied across participants. The IAT scores for each participant in the *different-L1* condition were associated with the data from the phonetic accommodation experiment as an estimate of the participants' implicit social attitude towards foreigners.

2.2.3. Final questionnaire

After finishing the phonetic accommodation experiment and the implicit association task, all participants in the experimental groups answered the final questionnaire. There were two questions regarding the participants' guess on the model talkers' identity: 1) first, the native status of the model talker, and 2) second, the dialect or native region of the model talker. All 23 participants for the two native model talkers, N1 and N2, identified the model talkers as native, and all of them identified the native model talkers' dialects as a US Northern dialect. 20 out of 24 participants for the two nonnative model talkers identified the model talkers as nonnative. 19

out of the 20 participants who identified the model talkers as nonnative responded that the nonnative model talkers would be from an Asian country, while one of the 20 responded that her model talker was from a South American country. The other 4 participants for the two nonnative model talkers responded that they had no idea about their model talkers' native status or native region.

2.2.4. Statistical analyses

The complete dataset from the phonetic accommodation experiment consists of 22,806 words (20 control participants x 126 words x 2 timings + 47 experimental participants x 126 words x 3 timings) and 11,584 sentences (20 control participants x 64 sentences x 2 timings + 47 experimental participants x 64 sentences x 3 timings) for each type of acoustic measurement.

The acoustic measurements were modified to “adjusted phonetic change”, following the formula described in Chapter 2.1. Then the adjusted phonetic change data were submitted to linear mixed effects regression models using the lme4 package (Bates, Maechler, & Bolker, 2011) and the languageR package (Baayen, 2011). The adjusted phonetic change indicates phonetic convergence, when the value is over 0, divergence, when the value is below 0, and maintenance, when the value is 0. Therefore, in the regression results, if the coefficient of a fixed effect factor is positive, it indicates that the factor leads to phonetic convergence. If the coefficient is negative, it indicates that the factor leads phonetic divergence.

Additionally, the data from dynamic time warping analyses and XAB perception tests with sentences were analyzed separately. The details will be mentioned in Chapter 6.

2.2.4.1 Linguistic talker distance and generalizability

First we asked whether participants differentiated their phonetic accommodation patterns depending on linguistic distances from the model talker. We tested two levels of linguistic distance, namely, the group level linguistic distance (*same-L1-same-dialect*, *same-L1-different-dialect*, *different-L1*) and the item level linguistic distance (preexisting acoustic distances between participants and their model talkers based on their productions of individual items within the experiment). Importantly, the item level linguistic distances were added not only with a linear function, but also with a quadratic function, when the quadratic function improved the model fit. This was to test whether the direction of change would influence the degree of phonetic convergence. Additionally, we asked whether participants generalized their phonetic changes from exposed words to unexposed words with different degrees depending on such linguistic distances between the participants and their model talkers.

To test these two questions, two types of statistical analyses were conducted on the full dataset for each acoustic measurement. First, a single *t*-test was run for each of the six subsets, namely, exposed and unexposed words in the three group level linguistic distances (*same-L1-same-dialect*, *same-L1-different-dialect*, *different-L1*), to see whether the adjusted phonetic changes were significantly different from zero (convergence or divergence) in each subset. Importantly, the purpose of the single *t*-tests was to judge the accommodation pattern of each of the six subsets separately, not to make any comparisons across the subsets. Therefore, Bonferroni correction was not applied to the single *t*-test results. Second, a linear mixed effects regression model was built with adjusted phonetic changes as the dependent measure to find intervening factors in the accommodation patterns found in the single *t*-test results. In each model, the fixed effect factors included group level linguistic distance (*same-L1-same-dialect*, *same-L1-different-*

dialect, and *different-LI*), linear and quadratic functions of item level linguistic distance, stimulus exposure (exposed, unexposed), and the two-way and three-way interactions of the three factors. Stimulus exposure was contrast coded. To make full comparisons among the three group level linguistic distances, the same regression model was run twice with different reference levels for the group level linguistic distance factor. Accordingly, the significance level was adjusted from 0.05 to 0.025 by Bonferroni correction. Other fixed effect factors such as exposed material set, word frequency, voicing of the initial consonant, and vowel frontness were added to each of the regression models, when the fixed effect factors improved the fits of the models. Categorical factors with two levels, such as stimulus exposure or vowel frontness, were contrast coded. Model talkers, participants, and items (words or sentences) were added as random effect factors. Further details of building the regression models will be discussed in Chapters 3, 4, and 5.

2.2.4.2 Implicit attitudes and generalizability

To investigate the influences of implicit attitudes of participants towards their model talkers on phonetic accommodation, we analyzed the data in the *different-LI* condition with the participants' IAT scores. For each acoustic measurement, we performed a linear mixed effects regression model analysis with the adjusted phonetic changes as the dependent measure. The fixed effect factors included IAT of the participants, the polynomial functions of item level acoustic talker distance, stimulus exposure, and their two-way and three-way interactions. Note that we added the polynomial functions of item level acoustic talker distances and stimulus exposure and their interactions with IAT scores, which are not directly related for the current research question, as fixed effect factors. This was because we found that such factors were

significant in the analyses on the full data. Thus, adding the fixed effect factors would allow us to see the marginal effect of implicit attitudes on phonetic accommodation. Other fixed effect factors were included when they significantly improved the model fit. All categorical fixed effect factors that were included in the final regression models with two levels were contrast coded. Model talkers, participants, and items (words or sentences) were added as random effect factors. Further details of building the regression models will be discussed in Chapters 3, 4, and 5.

2.2.4.3 Model validation

For model validation, we performed two types of robustness check on each of the linear mixed effects regression models. First, to check for the possibility of data overfitting by polynomial functions, we applied restricted cubic splines instead of polynomial (quadratic and linear) functions to item level acoustic talker distances. Importantly, restricted cubic splines did not provide significantly different model fits from the polynomial functions. That is, in each regression model, restricted cubic splines with three knots, thus combination of two lines, explained the tendency of the data the best. This resulted in similar model fits with the quadratic and linear terms of polynomial functions. Therefore, we could confirm that quadratic functions were indeed a good fit for the data. Second, using the method suggested by Baayen (2008), we attempted excluding outliers that made the distribution of regression residuals non-normal. Specifically, while checking quantile-quantile plots of regression residuals, we tried excluding data whose residuals were off from the normal line with various criteria. Then with the new datasets, we ran the same regression models. We found that the regression results were robust even after excluding the outliers. In other words, we could confirm that the outliers did not significantly affect the original regression models.

2.2.4.4 Graphical representation

In Chapters 3 and 4, for each acoustic measurement, the results are presented in figures with 6 panels: panels 1 and 2 for exposed items and unexposed items in the *same-L1-same-dialect* condition, panels 3 and 4 for those in the *same-L1-different-dialect* condition, and panels 5 and 6 in the *different-L1* condition. Within each panel, two types of data points are shown in a scatterplot with the x-axis representing the preexisting acoustic talker distance and the y-axis representing the adjusted phonetic changes. The two types of data points are the original adjusted phonetic changes and the fitted values yielded by the regression analyses of the adjusted phonetic changes. In these figures, grey dots represent the original adjusted phonetic changes of the given acoustic dimension, and black dots represent the fitted values of the regression model. The fitted values were taken from the results of the regression analysis and represent adjusted phonetic change as a function of preexisting acoustic talker distances with the interaction of group level linguistic distance and stimulus exposure. Although the fitted values (black dots) are shown as a function of preexisting acoustic talker distance, they also reflect regression coefficients of the other fixed effect factors, such as material sets, voicing of the initial or final consonants in words, and vowel frontness, height, and tenseness. With the effects of these other fixed factors reflected, the fitted values (black dots) on the figures represent the full results of the mixed effects regression models. It is for this reason that the fitted values are shown as dots for individual data points without a smoothed curves for each fixed factor.

3. MONOSYLLABIC WORDS

Segmental features of words are the most studied in the phonetic accommodation literature. As mentioned in Chapter 1, phonetic accommodation patterns of VOT (i.e. Nielsen, 2011; Shockley et al., 2004), vowel duration (Pardo, 2010), vowel F0 (i.e. Babel & Bulatov, 2011), and vowel F1 and F2 (i.e. Babel, 2009, 2010, 2012; Delvaux & Soquet, 2007) of words have been acoustically analyzed in the previous studies. We measured these five acoustic correlates in our monosyllabic word dataset.

3.1. Methods

3.1.1. Materials

63 English monosyllabic words were chosen to examine phonetic accommodation at the segmental level (initial and final consonants, and the vowel). Voicing and place of articulation of the initial and final consonants, frontness, height, and tenseness of the vowel were manipulated systematically across words. The list of monosyllabic words is given in Table 5.

Table 5. Two sets of English monosyllabic words

Word set				MW1				MW2				
Initial C				b		p		d		t		
Final C Voicing				+	-	+	-	+	-	+	-	
V	high	front	tense	i	bead	beak	pease	peak	deed	deep	tease	teak
			lax	ɪ	bib	bisque	pimp	pitch	din	ditch	tin	tiff
		back	tense	u	booze	boot	pool	poop	dune	deuce	tune	toot
			lax	ʊ	bull	bush	pull	putsch	dour	•	tour	took
	low	front	tense	æ	badge	batch	pad	patch	dab	dash	tab	tat
			lax	ɛ	bent	beck	pent	peck	deb	debt	tense	text
		back	tense	ɑ	bog	bock	pod	pock	dodge	dot	tog	tot
			lax	ʌ	buzz	buff	pug	puff	dud	dutch	tub	tusk

To test generalization of phonetic accommodation to unexposed items, two sets of monosyllabic words, MW1 and MW2, were established. Specific conditions for the actual word sets are as follows: 1) the two sets differ in the place of articulation of the initial consonant. In MW1, the words start with a bilabial stop (/b/ or /p/), and in MW2, with an alveolar stop (/d/ or /t/). 2) In each word set, half of the words have voiced initial stops (/b/ or /d/), and the other half, voiceless initial stops (/p/ or /t/). 3) Likewise, in each word set, half of the words have voiced final consonant, and the other half of the words, voiceless final consonant. 4) The vowels, /æ, ε, i, ɪ, α, ʌ, u, ʊ/, were controlled to be the same over the two sets. 5) Following Goldinger (1998) and Goldinger and Azuma (2004)'s finding that low frequency words were most successfully imitated, the criterion for word frequency was set to be under 30 per million words in SUBTLEXus (Brysbaert & New, 2009). 59 of the 63 monosyllabic words chosen fulfilled this condition, while exceptions were 31 ("tour"), 47 ("pool"), 76 ("deep"), and 342 ("took") per million words. The average frequency of all 63 monosyllabic words was 13 per million words. Considering all these conditions, 32 words were chosen for MW1 (words with bilabial initial stops), and 31 words were chosen for MW2 (words with alveolar initial stops).

3.1.2. Participants

All 67 participants described in Chapter 2 participated in the phonetic accommodation experiment that included monosyllabic words as part of the materials.

3.1.3. Procedure

All participants followed the general procedure described in Chapter 2, the General Methodology.

3.1.4. Analyses

Praat (Boersma & Weenink, 2010) was used for acoustic analyses of the monosyllabic words read by the model talkers and participants (pretest and posttest readings). Five acoustic measurements that were identified in previous studies on phonetic accommodation were performed for the monosyllabic words. First, voice onset time (VOT) was measured from the burst of the initial consonant until the start of periodicity for the vowel. Second, the vowel duration was measured from the onset of periodicity until the end of clear formant structure. Then a Praat script was run to measure the maximum F0 value of the vowel period. Any maximum F0 values measured with the script that were above 370 Hz or below 85 Hz were manually checked and corrected in case of a pitch-halving or -doubling error. The maximum F0 value of a vowel was selected instead of the average to find and correct errors in the Praat F0 measurement that might have happened during the automatic measurement through this manual checking. This is because, when values are averaged over a period, it is hard to check by the extreme error values (above 370 Hz, below 85 Hz) whether any errors are included in the average. The maximum F0 can be representative for the vowel of a monosyllabic word, because the variation of F0 values is minimal in a list reading format. Lastly, the Praat script also measured the average values of the F1 and F2 values of the vowel over the middle 50% of the vowel. Finally, adjusted phonetic change values were calculated for all measurements according to the formula described in Chapter 2.1.

3.2. Results

The complete dataset of monosyllabic words consists of 8,694 words (4 model talkers x 63 monosyllabic words + 20 control participants x 63 monosyllabic words x 2 timings + 47

experimental participants x 63 monosyllabic words x 2 timings) for each type of the acoustic measurements, namely, VOT, vowel duration, F0-max, F1, and F2 of the vowel (44,820 measurements in total = 8,694 words x 5 measurement types). Among these, in Table 6, the average values of the model talkers for each acoustic dimension are given depending on their group level linguistic distances to their participants.

Table 6. Model talker average values for each acoustic measurement type

Model talker	N1	N2	NN1	NN2
VOT (ms)	40.98	38.65	41.71	58.09
Vowel duration (ms)	248.02	187.94	166.88	214.52
F0-max (Hz)	266	222	220	249
F1 (Hz)	659	690	647	620
F2 (Hz)	1794	1901	1773	1785

Note. 1. N = native model talker, NN = nonnative model talker.
 2. N1 and N2 served as model talkers in the *same-L1-same-dialect* and *same-L1-different-dialect* conditions, and NN1 and NN2 in the different-L1 condition.

The dependent measures of phonetic accommodation, “adjusted phonetic changes”, were calculated by the formula described in Chapter 2.1 with each acoustic measurement for the 8,694 words, resulting in 2,961 data points for monosyllabic words (47 experimental participants x 63 monosyllabic words) for each measurement type (14,805 data points in total = 2,961 x 5 measurement types). There were no data where the preexisting talker distances were zero, therefore, no data points were excluded from the analyses.

3.2.1. Linguistic talker distance and generalizability for phonetic accommodation

To test the influence of linguistic talker distance on phonetic accommodation and its generalizability, we built linear mixed effects regression models, following the general methods

for statistical analyses described in Chapter 2.2.4.1. In all regression models, the quadratic functions of item level acoustic distance improved the model fit, so were included in the final regression models. Additional fixed effect factors were added when they improved the model fit. The condition numbers for multicollinearity of the fixed effect factors included to each of the final regression models were small to moderate ($8.6 < value < 12.9$), according to Baayen (2008, p. 200). Table 7 shows the additional fixed effect factors and the condition number for multicollinearity of all fixed effect factors for each of the regression models.

Table 7. Additional fixed effect factors and multicollinearity condition number for each linear mixed effects regression model

Acoustic measurement	Additional fixed effect factors	Multicollinearity condition number
VOT	Exposed word set, word set, vowel tenseness, vowel frontness, voicing of the final consonant	12.82
Vowel duration	Exposed word set, word set, voicing of the initial consonant, vowel frontness, voicing of the final consonant	10.39
F0-max	Exposed word set, word frequency, voicing of the initial consonant, vowel tenseness, voicing of the final consonant	11.73
F1	Exposed word set, voicing of the initial consonant, vowel frontness, voicing of the final consonant	8.67
F2	Exposed word set, word set, vowel tenseness, vowel height, voicing of the final consonant	11.74

- Note.* 1. The formula of a model for each acoustic measurement: (adjusted phonetic change) ~ group level linguistic distance x item level phonetic distance x stimulus exposure + (linear combinations of the additional fixed effect factors) + (1|model talker) + (1|participant) + (1|word)
2. The fixed effect factor, word set, represents the two word sets (MW1 and MW2) that were established for the phonetic accommodation experiment. Participants read both word sets in the pretest and posttest reading phases. The fixed effect factor, exposed word set, represents one of the two word sets to which participants were exposed during the perceptual exposure phase. Half of the participants were exposed to MW1, and the other half of the participants were exposed to MW2.

3.2.1.1 VOT

Figure 3 describes the adjusted VOT changes for exposed and unexposed monosyllabic words in the three group level linguistic distances. First, single *t*-tests showed that the average adjusted VOT changes for exposed and unexposed words in the *same-L1-same-dialect* condition (exposed: $M = 4.77$ ms, $t(409) = 5.76$, $p < 0.05$, unexposed: $M = 5.06$ ms, $t(408) = 5.94$, $p < 0.05$) and in *different-L1* condition were significantly different from zero (exposed: $M = 3.76$ ms, $t(754) = 5.18$, $p < 0.05$, unexposed: $M = 4.42$ ms, $t(756) = 6.40$, $p < 0.05$), suggesting VOT convergence in such conditions. However, the adjusted VOT changes for exposed and unexposed words in the *same-L1-different-dialect* condition were not significantly different from zero (exposed: $M = 1.38$ ms, unexposed: $M = 0.66$ ms), suggesting VOT maintenance. Therefore, we can see that participants converged towards their native model talkers with the same dialectal background and towards the nonnative model talkers, but not towards the model talkers with different dialects.

An ANOVA summary of the regression model results confirmed significant fixed effects of group level linguistic distance ($F(2, 2938) = 6.76$, $p < 0.025$) and the polynomial functions of item level linguistic distance ($F(2, 2938) = 171.24$, $p < 0.025$) on adjusted VOT changes. However, stimulus exposure was not significant. This suggests that participants generalized their VOT accommodation patterns from exposed words to unexposed words. The two-way interaction between group level linguistic distance and polynomial functions of item level linguistic distance was significant ($F(4, 2938) = 20.26$, $p < 0.025$), while the other two-way interactions and the three-way interaction with group level linguistic distance, polynomial functions of item level linguistic distance, and stimulus exposure were not significant. Regarding the insignificant interactions with stimulus exposure, the results suggest that participants

generalized their VOT accommodation patterns to new words in all of the three group level linguistic distance conditions, and also showed the same polynomial functions relating preexisting VOT talker distance and their VOT accommodation for old words and new words. The other fixed effects, namely, exposed word set, word set, vowel tenseness, vowel frontness, and voicing of the final consonant, were not significant.

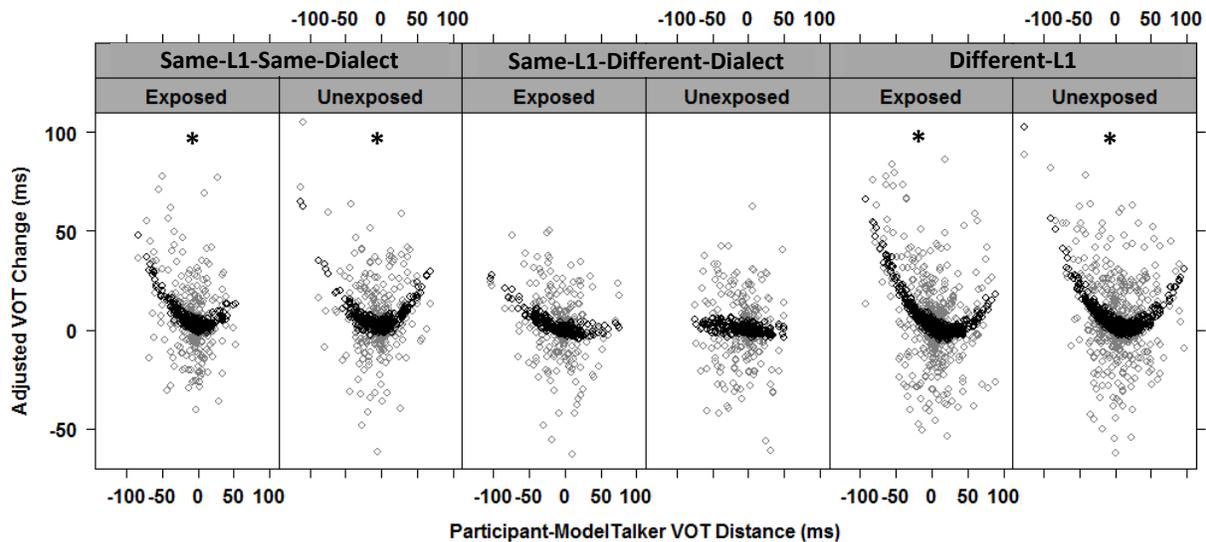


Figure 3. VOT accommodation patterns for exposed and unexposed monosyllabic words in *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions

Note. 1. The grey dots represent adjusted VOT change data, while the black dots represent the fitted values from the linear mixed effects regression model.

2. * denotes that the given dataset of adjusted VOT changes (the grey dots) is significantly different from zero.

3. For additional details, see section 2.2.4.4.

In detail, adjusted VOT changes were significantly lower in the *same-L1-different-dialect* condition than both in the *same-L1-same-dialect* condition ($\hat{\beta} = -4.85, p < 0.025$) and in the *different-L1* condition ($\hat{\beta} = -4.04, p < 0.025$). The quadratic function was significantly positive for the *same-L1-same-dialect* condition ($\hat{\beta} = 370.97, p < 0.025$), suggesting a U-shaped curve

pattern for its relation to adjusted VOT changes. This pattern was not significantly different in the *different-L1* conditions, while it was significantly flatter in the *same-L1-different-dialect* condition than in the *same-L1-same-dialect* condition ($\hat{\beta} = -302.79, p < 0.025$). The linear function of item level linguistic distance was not significant in the *same-L1-same-dialect* condition. This was not significantly different in the *same-L1-different-dialect* condition, but the linear function in the *different-L1* condition was significantly more negative than in the *same-L1-same-dialect* condition ($\hat{\beta} = -244.60, p < 0.025$).

Taken together, the results confirmed the intervening effects of group level VOT talker distance and item level VOT talker distance on VOT accommodation, and the generalization of VOT accommodation from exposed words to unexposed words. First, VOT convergence occurred for participants who had a native model talker with the same dialectal background and participants who had a nonnative model talker, but not for participants who had a model talker with a different dialectal background. In other words, VOT convergence was inhibited by dialect mismatch but not by L1 mismatch. This confirms the possibility we raised in Kim et al. (2011) about native talkers' phonetic convergence towards a high proficiency nonnative model talker. That is, the native participants in the current study converged towards their nonnative model talkers with high English proficiency in terms of VOT of monosyllabic words. However, it is still questionable why VOT convergence was inhibited in the middle distance, the *same-L1-different-dialect* condition, while it was not in the farthest distance, the *different-L1* condition.

Second, when VOT convergence occurred (*same-L1-same-dialect, different-L1*), the quadratic function of item level VOT difference between participants and their model talkers was significantly positive. This suggests two points. First, VOT convergence occurred in both the VOT increasing and decreasing directions. Second, the degree of VOT convergence was

positively proportional to the absolute preexisting VOT difference between participants and their model talkers. This opposite tendency of VOT convergence regarding two different types of linguistic distances, namely, the group level linguistic distance and the item level VOT distance confirms our prediction. That is, in the larger group level distance, a closer distance would facilitate phonetic convergence, as it would enhance the intelligibility between talkers. On the other hand, in the item level acoustic distance, the farther the distance, the larger the degree of convergence, as participants would get more room for accommodation.

In addition, when a model talker was a nonnative talker (*different-L1*), the degree of VOT convergence was larger in the decreasing direction. In fact, the insignificant coefficients of the linear function of preexisting VOT talker distance in the other group level linguistic distances were also negative (*same-L1-same-dialect*: $\hat{\beta} = -55.69$, *same-L1-different-dialect*: $\hat{\beta} = -30.16$, when the reference level was *same-L1-same-dialect*). Therefore, it might be that the general negative linear function became significant in the *different-L1* condition ($SD = 18.64$ ms) because of the larger variance of preexisting VOT distances than that of the other conditions (*same-L1-same-dialect*: $SD = 16.32$ ms, *same-L1-different-dialect*: $SD = 15.18$ ms).

Lastly, we found that VOT accommodation patterns generalized from old words to new words. This was true both for the degree of VOT convergence and the polynomial functions of item level VOT talker distance on VOT convergence in each of the group level linguistic distance conditions. This suggests that the VOT accommodation through passively hearing a native or nonnative model talker would not be limited to the exposed lexical items, but can result in systematic change in the talkers' speech production.

3.2.1.2 Vowel duration

Results from the single t -tests exhibited that the adjusted vowel duration changes were significantly over zero for both exposed and unexposed words in the *same-L1-same-dialect* condition (exposed: $M = 5.02$ ms, $t(409) = 2.57$, $p < 0.05$, unexposed: $M = 5.36$ ms, $t(408) = 2.98$, $p < 0.05$) and in the *different-L1* condition (exposed: $M = 7.02$ ms, $t(754) = 4.49$, $p < 0.05$, unexposed: $M = 6.92$ ms, $t(756) = 4.47$, $p < 0.05$), suggesting vowel duration convergence. But the adjusted vowel duration changes in the exposed and unexposed words in the *same-L1-different-dialect* condition were not significantly different from zero, suggesting vowel duration maintenance (exposed: $M = 3.38$ ms, unexposed: $M = 3.66$ ms). This suggests that participants converged towards native model talkers with the same dialect background and nonnative model talkers, but not towards native model talkers with different dialectal backgrounds. Figure 4 summarizes vowel duration accommodation patterns of exposed and unexposed monosyllabic words in the three group level linguistic distance conditions.

The ANOVA summary of the regression results exhibited that the group level linguistic distance did not significantly affect adjusted vowel duration changes. That is, unlike the single t -test results, the adjusted vowel duration changes did not significantly differ among the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions when we consider other fixed and random effect factors in the regression model. However, the polynomial functions of item level vowel duration talker differences were significant ($F(2, 2938) = 61.43$, $p < 0.025$). The interaction between group level linguistic distance and the polynomial functions of vowel duration talker differences was also significant ($F(4, 2938) = 14.38$, $p < 0.025$). Stimulus exposure was not significant, suggesting that participants generalized their vowel duration accommodation patterns from exposed items to unexposed items. None of two-way and three-

way interactions with stimulus exposure, group level linguistic distance, and the item level vowel duration talker distance were significant. This indicates that participants exhibited the same vowel duration accommodation on old and new words, not only for the degree of convergence, but also for the quadratic and linear functions relating preexisting vowel duration talker differences to adjusted vowel duration changes. Additionally, voicing of the initial consonant was significant in predicting adjusted vowel duration changes ($F(1, 2938) = 5.60, p < 0.025$). None of the other fixed effect factors, namely, exposed word set, word set, vowel frontness, and voicing of the final consonant, were significant.

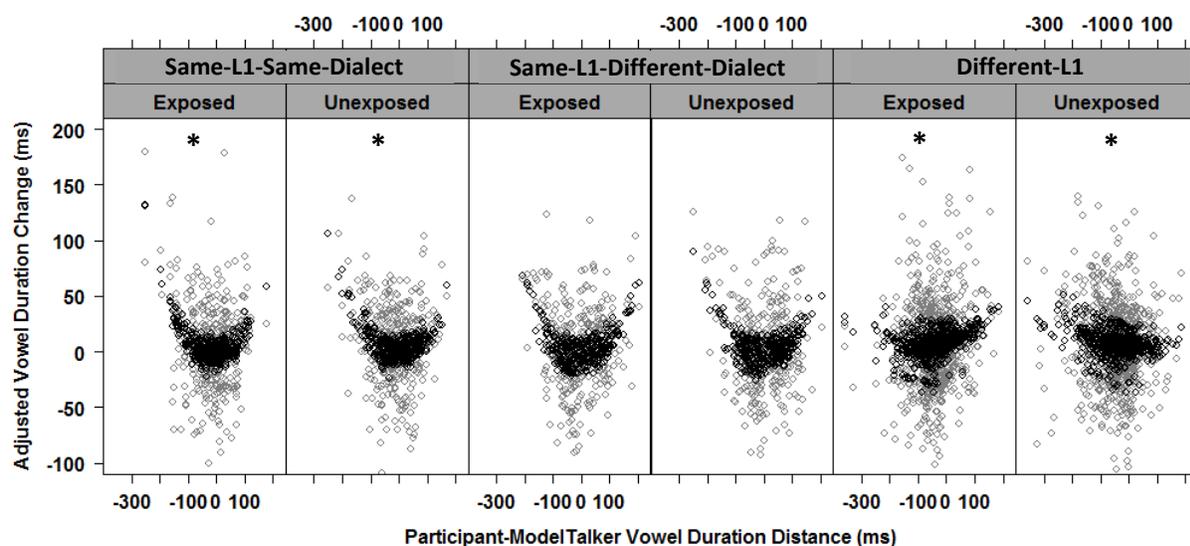


Figure 4. Vowel duration accommodation patterns on exposed and unexposed monosyllabic words in the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions

- Note.*
1. The grey dots represent adjusted vowel duration change data, while the black dots represent the fitted values from the linear mixed effects regression model.
 2. * denotes that the given dataset of adjusted vowel duration changes (the grey dots) is significantly different from zero.
 3. For additional details, see section 2.2.4.4.

In detail, the quadratic function of adjusted vowel duration changes in the *same-L1-same-dialect* condition was significantly positive ($\hat{\beta} = 972.99, p < 0.025$). This was not significantly different in the *same-L1-different-dialect* condition, but significantly flatter in the *different-L1* condition ($\hat{\beta} = -672.30, p < 0.025$). The linear function was significantly negative in the *same-L1-same-dialect* condition ($\hat{\beta} = -557.55, p < 0.025$). Again, this pattern was not significantly different in the *same-L1-different-dialect* condition, but significantly more positive in the *different-L1* condition ($\hat{\beta} = 500.58, p < 0.025$). In addition, adjusted vowel duration changes were significantly higher in words with a voiceless initial consonant than in words with a voiced initial consonant ($\hat{\beta} = 3.71, p < 0.025$).

In sum, participants converged towards their native or nonnative model talkers in terms of vowel duration. Dialect mismatch and L1 mismatch did not inhibit vowel duration convergence. When the model talker was a native model talker, vowel duration convergence occurred in both of increasing and decreasing directions, and the farther the preexisting vowel duration talker distances, the larger the degree of vowel duration convergence was. Additionally, the degree of vowel duration convergence was even larger in the decreasing direction. This result is in line with our prediction that durations might be imitated more in the decreasing direction, facilitated by the repetition reduction effect. However, these quadratic and linear functions of preexisting vowel duration talker distances became significantly weaker with a nonnative model talker. Importantly, participants generalized their vowel duration accommodation patterns from exposed to unexposed words, both for the degree of vowel duration convergence and for the function of preexisting vowel duration talker distance. Lastly, vowel durations in monosyllabic words with a voiceless initial consonant were imitated with a larger degree than those in monosyllabic words with a voiced initial consonant.

3.2.1.3 F0-max

Single *t*-test results suggest that the adjusted F0-max changes for exposed and unexposed words in the *same-L1-same-dialect* condition (exposed: $M = 4.84$ Hz, $t(409) = 2.49$, $p < 0.05$, unexposed: $M = 15.01$ Hz, $t(408) = 6.83$, $p < 0.05$) and in the *different-L1* condition (exposed: $M = 13.09$ Hz, $t(754) = 8.23$, $p < 0.05$, unexposed: $M = 15.94$ Hz, $t(756) = 8.99$, $p < 0.05$), and for unexposed words in the *same-L1-different-dialect* condition ($M = 4.99$ Hz, $t(314) = 2.75$, $p < 0.05$) were significantly above zero, suggesting F0-max convergence. Therefore, we can see that participants converged towards their model talkers in all of the three group level linguistic distances and both for exposed and unexposed words in terms of vowel F0-max of monosyllabic words, except for unexposed words in the *same-L1-different-dialect* condition. Figure 5 describes the F0-max accommodation patterns of exposed and unexposed words in the three group level linguistic distances.

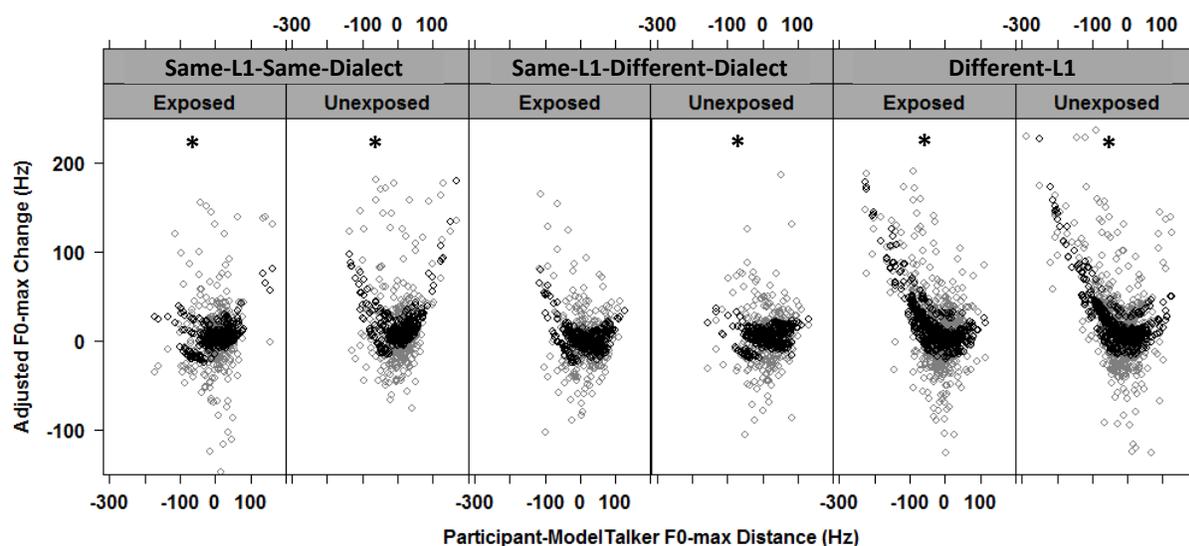


Figure 5. F0-max accommodation patterns on exposed and unexposed monosyllabic words in the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions

- Note.*
1. The grey dots represent adjusted F0-max change data, while the black dots represent the fitted values from the linear mixed effects regression model.
 2. * denotes that the given dataset of adjusted F0-max changes (the grey dots) is significantly different from zero.
 3. For additional details, see section 2.2.4.4.

The first part of the detailed regression results is about the insignificant interaction between stimulus exposure and group level linguistic distance. The adjusted F0-max changes were significantly larger for unexposed words than for exposed words in the *same-L1-same-dialect* condition ($\beta = 11.82, p < 0.025$). This means that participants converged towards their model talker for unexposed words even more than for exposed words. This generalization effect was not significantly different in the other group level linguistic distances. Second, the quadratic function of item level F0-max talker distances was significantly positive for exposed words in the *same-L1-same-dialect* condition ($\hat{\beta} = 951.20, p < 0.025$). The positive quadratic function was larger for unexposed words than for exposed words in the same condition ($\hat{\beta} = 800.00, p < 0.025$), but decreased for exposed words in the *same-L1-different-dialect* condition ($\hat{\beta} = -352.80, p < 0.025$), and decreased more for unexposed words in the *same-L1-different-dialect* condition ($\hat{\beta} = -1355.00, p < 0.025$). The positive quadratic function of item level F0-max talker distance for exposed words in the *different-L1* condition was not significantly different from that for exposed words in the *same-L1-same-dialect* condition. However, it was significantly flatter for unexposed words than for exposed words in the *different-L1* condition ($\hat{\beta} = -773.50, p < 0.025$). The linear function of item level F0-max talker distances was significantly negative for exposed words in the *same-L1-same-dialect* condition ($\hat{\beta} = -276.40, p < 0.025$). The negative linear function was more negative for unexposed words in the *same-L1-same-dialect* condition ($\hat{\beta} = -444.90, p < 0.025$), and for exposed words in the *same-L1-different-dialect* condition ($\hat{\beta} = -440.90, p < 0.025$) and in the *different-L1* condition ($\hat{\beta} = -379.80, p < 0.025$), than for exposed words in the *same-L1-same-dialect* condition. in the *same-L1-different-dialect* condition ($\hat{\beta} = -440.90, p < 0.025$), and in the *different-L1* condition ($\hat{\beta} = -379.80, p < 0.025$). However, the negative linear function of item level F0-max talker distance was significantly more positive or

less negative for unexposed words in the *same-L1-different-dialect* condition ($\hat{\beta} = 1108.00, p < 0.025$), and for unexposed words in the *different-L1* condition ($\hat{\beta} = 477.70, p < 0.025$).

Additionally, low vowels showed a larger degree of adjusted F0-max changes than high vowels ($\hat{\beta} = 3.72, p < 0.025$).

Taken together, participants exhibited F0-max convergence to their model talker in all of the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions. Therefore, again, we found evidence that native talkers converged to both native and nonnative model talkers phonetically. The preexisting F0-max talker differences were positively correlated with the degree of F0-max convergence in both of F0-max increasing and decreasing directions in all group level linguistic distances, but the degree of the positive quadratic function decreased with dialect mismatch and L1 mismatch and for unexposed words in such conditions. Additionally, the degree of F0-max convergence was larger in the F0-max decreasing direction in all of the group level linguistic distances, and the degree of the negative linear function decreased for unexposed words, compared to exposed words. Importantly, participants generalized their F0-max convergence from exposed words to unexposed words with even larger degree of convergence for unexposed words. Moreover, participants also generalized the positive quadratic and negative linear functions relating preexisting F0-max talker differences to the degree of F0-max convergence from exposed words to unexposed words, with smaller degrees of the quadratic and linear functions for unexposed words in most of the cases. Finally, F0 of a low vowel was imitated with a larger degree than that of a high vowel.

3.2.1.4 F1

Results of the single t-tests revealed that the adjusted F1 changes were over zero for exposed and unexposed words in the *same-L1-same-dialect* (exposed: $M = 10.33$ Hz, $t(409) = 3.11$, $p < 0.05$, unexposed: $M = 12.43$ ms, $t(408) = 3.47$, $p < 0.05$) and *differet-L1* conditions (exposed: $M = 15.55$ Hz, $t(754) = 6.77$, $p < 0.05$, unexposed: $M = 8.92$ Hz, $t(756) = 3.61$, $p < 0.05$) and for exposed words in the *same-L1-different-dialect* condition ($M = 17.26$ Hz, $t(314) = 5.72$ $p < 0.05$). However, the adjusted F1 changes were not significantly different from zero for unexposed words in the *same-L1-different-dialect* condition ($M = 5.82$ Hz). Figure 6 summarizes the F1 accommodation patterns for exposed and unexposed words in the three group level linguistic distances.

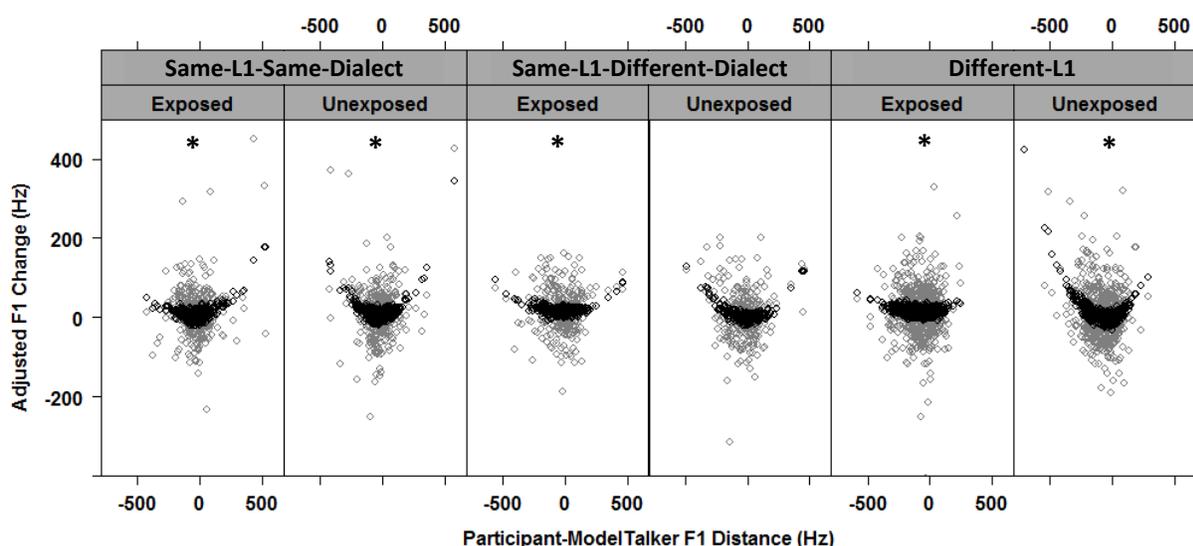


Figure 6. F1 accommodation patterns on exposed and unexposed monosyllabic words in the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions

- Note.*
1. The grey dots represent adjusted F1 change data, while the black dots represent the fitted values from the linear mixed effects regression model.
 2. * denotes that the given dataset of adjusted F1 changes (the grey dots) is significantly different from zero.
 3. For additional details, see section 2.2.4.4.

The ANOVA summary of the linear mixed effects regression model results exhibited that group level linguistic distance did not significantly affect adjusted F1 changes. This suggests that dialect mismatch and L1 mismatch did not inhibit F1 convergence. However, the polynomial functions of item level F1 distances were significant in predicting adjusted F1 changes ($F(2, 2938) = 107.91, p < 0.025$). The two-way interaction between group level linguistic distance and polynomial functions of item level F1 talker distance was not significant, indicating that the polynomial functions in the three group level talker distances were not significantly different from one another. Stimulus exposure was not significant, suggesting that participants generalized their F1 convergence from exposed words to unexposed words. The two-way interaction between the polynomial functions of item level F1 talker distance and stimulus exposure was significant ($F(2, 2938) = 18.75, p < 0.025$). However, the two-way interaction between stimulus exposure and group level linguistic distance and the three-way interaction among stimulus exposure, group level linguistic distance, and the polynomial functions of item level F1 talker distance were not significant. Thus, we can see that there were no significant differences in the F1 convergence generalization effect among the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions, in terms of the degree of generalization and the difference in the polynomial functions. Additionally, voicing of the onset was significant ($F(1, 2938) = 6.20, p < 0.025$). None of the other fixed effect factors, namely, exposed word set, word set, vowel frontness, and voicing of the final consonant, were significant.

The details of the significant regression results are as follows. First, the quadratic function of item level F1 talker distance was significantly positive ($\hat{\beta} = 1014.78, p < 0.025$). However, the linear function was not significant. Second, the positive quadratic function of item level F1 talker distance on adjusted F1 changes was significantly stronger for unexposed words

than for exposed words ($\hat{\beta} = 623.51, p < 0.025$). Lastly, monosyllabic words with a voiceless initial consonant exhibited significantly larger adjusted F1 changes ($\hat{\beta} = 6.69, p < 0.025$).

In sum, we found that there was significant F1 convergence to all model talkers, thus dialect mismatch and L1 mismatch did not inhibit F1 convergence. F1 convergence occurred in both of F1 increasing and decreasing directions, and the preexisting F1 differences between participants and their model talkers were positively correlated with the degree of F1 convergence. Moreover, F1 convergence was generalized from exposed words to unexposed words both for the degree of convergence and the quadratic function relating the preexisting F1 talker distance to F1 convergence. The quadratic function was even stronger for unexposed words than for exposed words. In addition, participants exhibited a larger degree of F1 convergence for monosyllabic words with a voiceless initial consonant.

3.2.1.5 F2

Results from the single *t*-tests revealed that the adjusted F2 changes were significantly over zero for exposed and unexposed words in all of the *same-L1-same-dialect* (exposed: $M = 26.52$ Hz, $t(409) = 3.46, p < 0.05$, unexposed: $M = 21.03$ Hz, $t(408) = 2.87, p < 0.05$), *same-L1-different-dialect* (exposed: $M = 22.13$ Hz, $t(314) = 2.61, p < 0.05$, unexposed: $M = 23.17$ Hz, $t(314) = 2.63, p < 0.05$), *different-L1* (exposed: $M = 19.80$ Hz, $t(754) = 3.50, p < 0.05$, unexposed: $M = 23.8$ Hz, $t(756) = 4.04, p < 0.05$) conditions. These results suggest that participants in all group level linguistic distances converged towards their model talkers in terms of F2, and they generalized their F2 convergence from exposed words to unexposed words. Figure 7 describes the F2 accommodation patterns for exposed and unexposed words in the three group level linguistic distances.

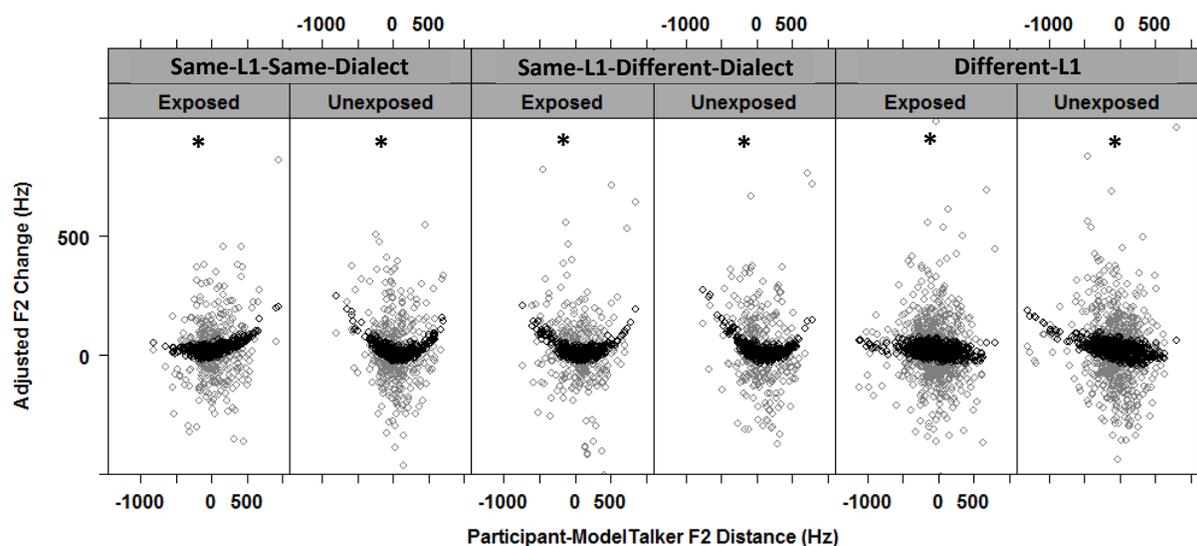


Figure 7. F2 accommodation patterns on exposed and unexposed monosyllabic words in the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions

- Note.*
1. The grey dots represent adjusted F2 change data, while the black dots represent the fitted values from the linear mixed effects regression model.
 2. * denotes that the given dataset of adjusted F2 changes (the grey dots) is significantly different from zero.
 3. For additional details, see section 2.2.4.4.

The ANOVA summary of regression results confirmed the single *t*-test results. The group level linguistic distance was not significant, and neither were the stimulus exposure nor the interaction between the two fixed effect factors. Therefore, dialect mismatch and L1 mismatch between participants and their model talkers did not inhibit F2 convergence. Moreover, participants applied their F2 convergence to unexposed words in all of the group level linguistic distance conditions. Additionally, the polynomial functions of item level F2 talker distances were significant ($F(2, 2938) = 20.33, p < 0.025$). The two-way interactions with the polynomial functions of item level F2 talker distance, namely, the one with group level linguistic distance ($F(4, 2938) = 6.52, p < 0.025$) and the one with stimulus exposure ($F(2, 2938) = 4.01, p < 0.025$),

were significant in predicting adjusted F2 changes. The three-way interaction among group level linguistic distance, item level F2 talker distance, and stimulus exposure was not significant. In addition, vowel height was significant ($F(1, 2938) = 5.12, p < 0.025$). None of the other fixed effect factors, namely, exposed word set, word set, vowel tenseness, and voicing of the final consonant, were significant.

Details of the significant regression results are as follows. First, the quadratic function of item level F2 talker distance was significantly positive in the *same-L1-same-dialect* condition ($\hat{\beta} = 1704.56, p < 0.025$), while the linear function was not significant. Second, in the *same-L1-different-dialect* condition, the quadratic function was not significantly different from that in the *same-L1-same-dialect* condition, while the linear function was significantly more negative than in the *same-L1-same-dialect* condition ($\hat{\beta} = -1195.21, p < 0.025$). Third, the quadratic function in the *different-L1* condition was significantly flatter than in the *same-L1-same-dialect* condition ($\hat{\beta} = -1402.31, p < 0.025$). However, the linear function in the *different-L1* condition was not significantly different from that in the *same-L1-same-dialect* condition. Fourth, unexposed words exhibited significantly more negative linear function of item level F2 talker distances than exposed words ($\hat{\beta} = -1788.66, p < 0.025$). Finally, low vowels exhibited significantly lower adjusted F2 changes than high vowels ($\hat{\beta} = -15.61, p < 0.025$).

Taken together, we found that participants converged to all model talkers in terms of F2. This indicates that dialect mismatch and L1 mismatch did not inhibit F2 convergence. Additionally, F2 convergence occurred in both of F2 increasing and decreasing directions, and the degree of F2 convergence was positively proportional to the model talker – participant F2 distance before perceptual exposure. However, this tendency was weaker in the L1 mismatch condition. Moreover, participants generalized their F2 convergence patterns from exposed words

to unexposed words. For unexposed words, the degree of F2 convergence was larger in the direction of F2 decrease than in the F2 increasing direction. Finally, F2 of high vowels were imitated more than that of low vowels.

Summing up, we found that participants generally converged to all model talkers in most of the acoustic measurements on monosyllabic words. Specifically, we found three tendencies regarding our research questions on linguistic talker distances and phonetic accommodation as below:

1. Regarding group level linguistic distance, dialect mismatch inhibited VOT and vowel duration convergence. However, convergence of F0, F1, and F2 was not affected by group level linguistic distance at all. Importantly, L1 mismatch did not inhibit phonetic convergence of monosyllabic words on any of the acoustic correlates.
2. For all acoustic correlates for monosyllabic words, the quadratic function of preexisting acoustic distance between participants and their model talker was positively proportional to the degree of phonetic convergence. That is, the farther the acoustic distance at the item level, the larger the degree of convergence was. This also means that phonetic convergence occurred in both of the decreasing and increasing directions along the acoustic correlates. Additionally, the degree of convergence was larger in the decreasing direction for vowel duration and F0-max in all group level linguistic distance conditions. Regarding vowel duration, this confirms our prediction that lexical repetition effects of reducing duration and intelligibility would facilitate phonetic convergence.
3. Finally, participants generalized phonetic convergence on monosyllabic words from exposed words to unexposed words for all of the acoustic correlates in all group level

linguistic distance conditions. Moreover, the quadratic function of preexisting acoustic talker distances was also generalized to unexposed words in most of the cases. These indicate that phonetic convergence on monosyllabic words is not a transient change. Rather, it might lead to solid language learning with more training or experiences.

3.2.2. Implicit attitudes and generalizability of phonetic accommodation

Now we look at the influence of participants' implicit attitudes towards foreigners on their phonetic accommodation towards a nonnative model talker. We also tested how the generalizability of phonetic accommodation from exposed words to unexposed words was controlled by this implicit attitudes factor. To investigate these questions, linear mixed effects regression models were built, following the general methods described in Chapter 2.2.4.2. Additional fixed effect factors were included in the regression model, when they improved the model fit. The condition numbers for multicollinearity of fixed effect factors in the final regression models were small ($5.3 < \text{value} < 8.7$), according to Baayen (2008, p. 200). Table 8 lists additional fixed effect factors and the multicollinearity condition number for each regression model.

Table 8. Additional fixed effect factors and multicollinearity condition number for each regression model

Acoustic measurement	Additional fixed effect factors	Multicollinearity condition number
VOT	Word frequency, voicing of the initial consonant, vowel height, vowel frontness	7.31
Vowel duration	Exposed word set, word set, vowel tenseness, voicing of the final consonant	6.92
F0-max	Word frequency, voicing of the initial consonant, vowel height	8.17
F1	Word frequency, word set, voicing of the initial consonant, vowel frontness	8.67
F2	Exposed word set, word set, voicing of the initial consonant, vowel tenseness, voicing of the final consonant	5.33

Note. 1. The formula of a model for each acoustic measurement: (adjusted phonetic change) ~ IAT x item level phonetic distance x stimulus exposure + (linear combinations of the additional fixed effect factors) + (1|model talker) + (1|participant) + (1|word)

2. The fixed effect factor, word set, represents the two word sets (MW1 and MW2) that were established for the phonetic accommodation experiment. Participants read both word sets in the pretest and posttest reading phases. The fixed effect factor, exposed word set, represents one of the two word sets, which was exposed to participants during the perceptual exposure phase. Half of the participants were exposed to MW1, and the other half of the participants were exposed to MW2.

3.2.2.1 VOT

The ANOVA summary of the results from a linear mixed effects regression model on adjusted VOT changes suggest that participants' IAT was not significant. However, polynomial functions of the item level VOT talker distances were significant ($F(2, 1498) = 138.94, p < 0.05$), as well as the interaction between IAT and the polynomial functions ($F(2, 1498) = 5.56, p < 0.05$).

Stimulus exposure was not significant, and none of the two-way and three-way interactions with stimulus exposure was significant, either. None of the other fixed effect factors, namely, word frequency, voicing of the initial consonant, vowel height and frontness was significant in predicting adjusted VOT changes in the *different-L1* condition. Figure 8 summarizes VOT

accommodation patterns in the *different-L1* condition as a function of stimulus exposure, IAT, and item level VOT talker distance.

In the detailed regression results, both the quadratic function ($\hat{\beta} = 387.1, p < 0.05$) and the linear function ($\hat{\beta} = -221.5, p < 0.05$) of the item level VOT talker distances were significant. The interaction between IAT and the quadratic function of item level VOT talker distances was significant ($\hat{\beta} = -200.7, p < 0.05$), while the interaction between IAT and the linear function was not.

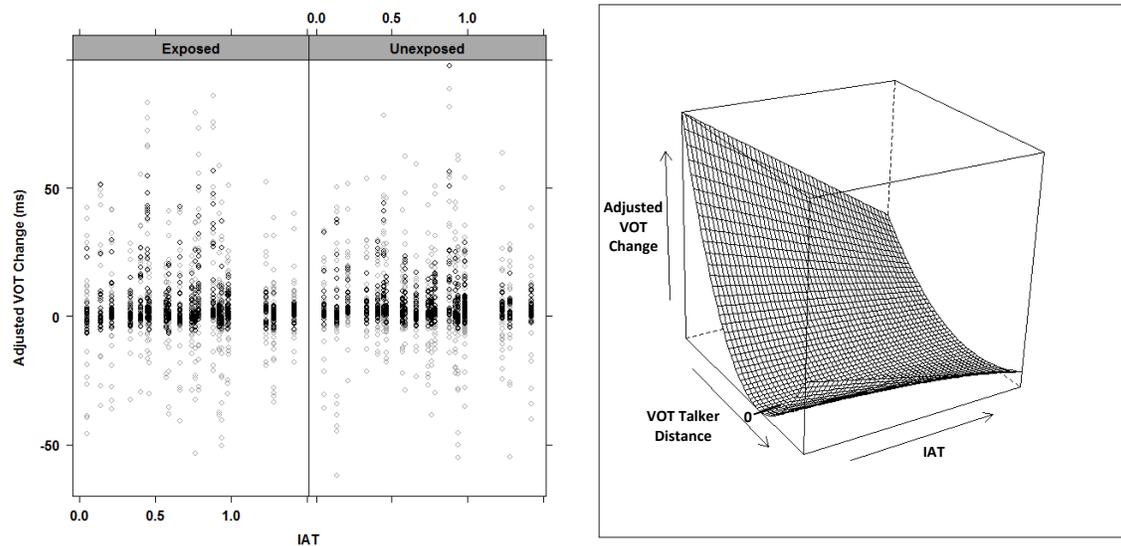


Figure 8. VOT accommodation patterns in the *different-L1* condition

- Note.*
1. In the left plot, the grey dots represent adjusted VOT change data, while the black dots represent the fitted values from the linear mixed effects regression model. For additional details, see section 2.2.4.4.
 2. The right plot is a schematic three-dimension plot with adjusted VOT changes, preexisting VOT talker distance, and IAT scores.
 3. A positive IAT score indicates a negative implicit attitude towards foreigners, and the higher a score is, the more negative the attitude is for foreigners.

Therefore, participants' implicit attitudes towards foreigners did not significantly affect their degree of VOT convergence. Instead, the IAT scores negatively affected the quadratic function of item level VOT talker distance. That is, the more negative participants' implicit attitudes towards foreigners, the weaker the quadratic function was, thus participants with negative attitudes to foreigners might need a larger VOT distance from their model talkers for VOT convergence. Moreover, participants generalized these behaviors from words they heard during the exposure phase to words they did not hear.

3.2.2.2 Vowel duration

The ANOVA summary of the linear mixed effects regression model results exhibited that the implicit attitudes, namely, participants' IAT scores, were not significant. The polynomial functions of preexisting vowel duration talker distances were significant ($F(2, 1498) = 16.86, p < 0.05$), as well as their interaction with IAT ($F(2, 1498) = 4.86, p < 0.05$). Stimulus exposure was not significant, nor were the two-way and three-way interactions with stimulus exposure significant. None of the other fixed effect factors, namely, exposed words set, word set, vowel tenseness, voicing of the final consonant, were significant. Figure 9 displays plots for vowel duration accommodation with regards to IAT and preexisting vowel duration talker distances.

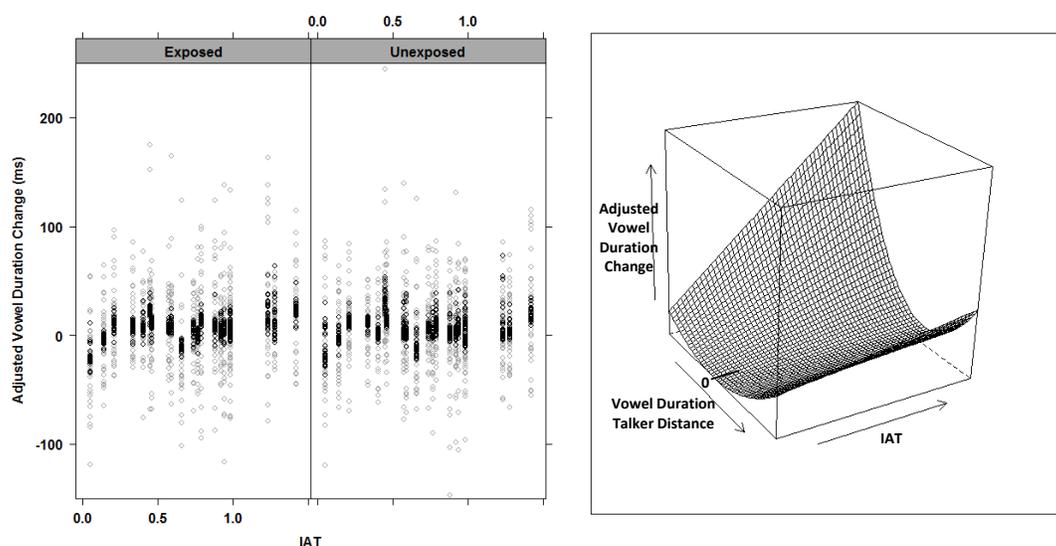


Figure 9. Vowel duration accommodation patterns in the *different-L1* condition

- Note.*
1. In the left plot, the grey dots represent adjusted vowel duration change data, while the black dots represent the fitted values from the linear mixed effects regression model. For additional details, see section 2.2.4.4.
 2. The right plot is a schematic three-dimension plot with adjusted vowel duration changes, preexisting vowel duration talker distance, and IAT scores.
 3. A positive IAT score indicates a negative implicit attitude towards foreigners, and the higher a score is, the more negative the attitude is for foreigners.

In detail, the quadratic function of vowel duration talker distance was significant ($\hat{\beta} = 148.3, p < 0.05$), while the linear function was not. This quadratic function was larger when IAT scores were higher ($\hat{\beta} = 227.17, p < 0.05$).

Taken together, we found that participants' implicit attitudes towards foreigners did not significantly influence the degree of vowel duration convergence. However, they affected the relation between preexisting vowel duration talker distance and the degree of vowel duration convergence. That is, participants with more negative attitudes towards foreigners needed smaller preexisting vowel duration talker distance to converge towards their nonnative model

talkers. This is the opposite tendency from that of VOT accommodation in the *different-L1* condition. The source of this variation is unknown yet.

3.2.2.3 F0

Results from the linear mixed effects regression model suggest that participants IAT scores did not significantly affect adjusted F0 changes in the *different-L1* condition. However, the linguistic distance measure, namely, the polynomial functions of preexisting F0 talker distances were significant ($F(2, 1498) = 257.89, p < 0.05$). The interaction between IAT scores and the polynomial functions of preexisting F0 talker distances was also significant ($F(2, 1498) = 12.05, p < 0.05$). Additionally, vowel height was significant in predicting adjusted F0 changes by participants in the *different-L1* group ($F(1, 1498) = 8.78, p < 0.05$). None of the other fixed effect factors, namely, word frequency and voicing of the initial consonant, were significant. Figure 10 summarizes the F0 accommodation patterns in the *different-L1* condition.

In detail, both the quadratic function ($\hat{\beta} = 473.43, p < 0.05$) and the linear function ($\hat{\beta} = -339.97, p < 0.05$) of preexisting F0 talker distance were significant. The two-way interactions between the quadratic function and IAT scores ($\hat{\beta} = 332.82, p < 0.05$) and between the linear function and IAT scores ($\hat{\beta} = 227.17, p < 0.05$) were also significant. Moreover, low vowels were imitated with a larger degree than high vowels ($\hat{\beta} = -680.40, p < 0.05$).

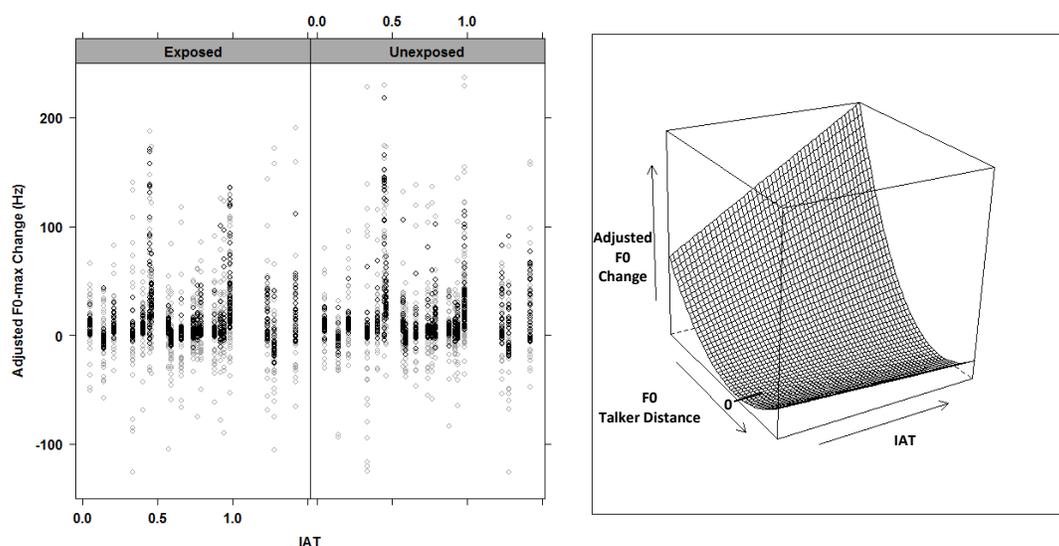


Figure 10. F0 accommodation patterns in the *different-L1* condition

- Note.*
1. In the left plot, the grey dots represent adjusted F0 change data, while the black dots represent the fitted values from the linear mixed effects regression model. For additional details, see section 2.2.4.4.
 2. The right plot is a schematic three-dimension plot with adjusted F0 changes, preexisting F0 talker distance, and IAT scores.
 3. A positive IAT score indicates a negative implicit attitude towards foreigners, and the higher a score is, the more negative the attitude is for foreigners.

These results indicate that participants' implicit attitudes towards foreigners did not directly affect their degree of F0 convergence, rather, participants with more negative attitudes towards foreigners needed smaller preexisting F0 distances from their model talkers for F0 convergence to occur. Again, we see that the effect of the interaction between IAT and preexisting acoustic talker distance might vary depending on the acoustic measurement type. In order for convergence to occur, participants with more negative attitudes needed less VOT talker distances, more vowel duration talker distances, and more F0 talker distances than participants with less negative attitudes. Additionally, for F0 convergence, participants generalized these interactions from heard words to unheard words.

3.2.2.4 F1

Results of the linear mixed effects regression model on adjusted F1 changes in the *different-L1* condition suggest that participants' IAT scores were not significant. However, polynomial functions of preexisting F1 talker distance were significant ($F(2, 1498) = 52.49, p < 0.05$). Moreover, stimulus exposure was significant ($F(1, 1498) = 4.90, p < 0.05$). Two-way interactions between the polynomial functions and stimulus exposure ($F(2, 1498) = 12.04, p < 0.05$) and between IAT and the polynomial functions ($F(2, 1498) = 7.23, p < 0.05$) were also significant. Additionally, voicing of the initial consonant was also significant ($F(1, 1498) = 5.42, p < 0.05$). The other fixed effect factors, word frequency and word set, were not significant. Figure 11 exhibits F1 accommodation patterns in the *different-L1* condition, regarding stimulus exposure, IAT, and preexisting F1 talker distance.

In detail, both of the quadratic ($\hat{\beta} = 636.11, p < 0.05$) and linear functions ($\hat{\beta} = -247.58, p < 0.05$) of F1 talker distance were significant. Although IAT was not significant, its interaction with the quadratic function of F1 talker distance was significant ($\hat{\beta} = -369.52, p < 0.05$), while its interaction with the linear function was not. Unexposed words showed significantly lower adjusted F1 changes than exposed words ($\hat{\beta} = -17.13, p < 0.05$). However, unexposed words had significantly stronger quadratic function of F1 talker distance than exposed words ($\hat{\beta} = 565.81, p < 0.05$). Additionally, F1 of words with a voiceless initial consonant exhibited significantly higher adjusted F1 changes ($\hat{\beta} = 8.05, p < 0.05$).

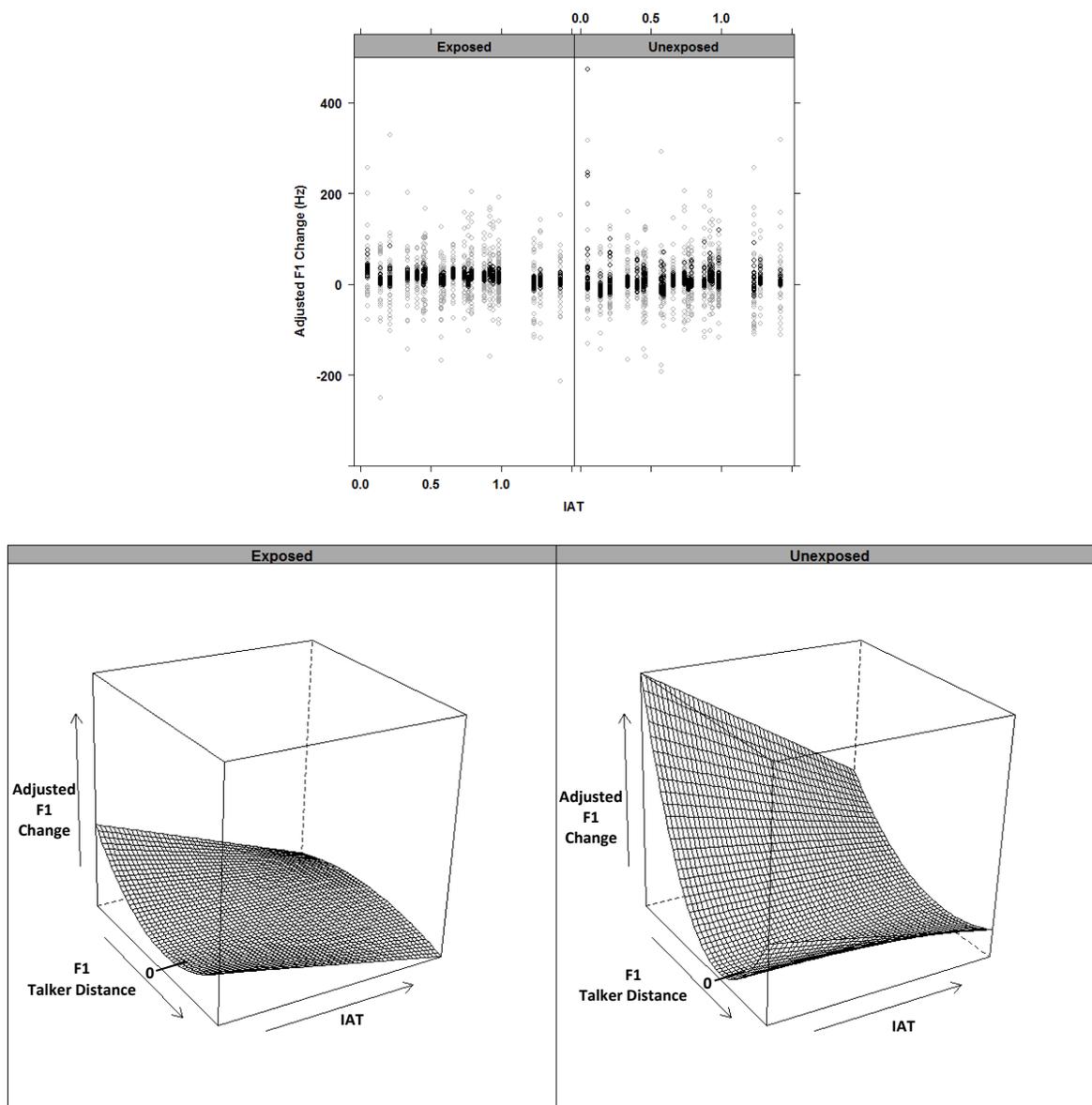


Figure 11. F1 accommodation patterns in the *different-L1* condition

- Note.*
1. In the top plot, the grey dots represent adjusted F1 change data, while the black dots represent the fitted values from the linear mixed effects regression model. For additional details, see section 2.2.4.4.
 2. The bottom plot is a schematic three-dimension plot with adjusted F1 changes, preexisting F1 talker distance, and IAT scores.
 3. A positive IAT score indicates a negative implicit attitude towards foreigners, and the higher a score is, the more negative the attitude is for foreigners.

Taken together, participants' implicit attitudes towards foreigners did not significantly affect the degree of F1 convergence. However, the more negative their attitudes towards foreigners, the weaker the positive influence of preexisting F1 talker distance on F1 convergence. Interestingly, this pattern was stronger for unexposed words than for exposed words, although unexposed words experienced a lower degree of F1 convergence.

3.2.2.5 F2

Results from the linear mixed effects regression model suggest that participants' IAT scores were not significant in predicting adjusted F2 changes. However, polynomial functions of F2 talker distances were significant ($F(2, 1498) = 8.53, p < 0.05$), as was the interaction with IAT ($F = 2.38, p < 0.05$). Stimulus exposure was not significant, suggesting that participants generalized their F2 accommodation patterns from exposed to unexposed words. Additionally, vowel tenseness was significant ($F(1, 1498) = 6.87, p < 0.05$). None of the other two-way and three-way interactions among IAT, polynomial functions of F1 talker distance, and stimulus exposure were significant. None of the other fixed effect factors, exposed word set, word set, voicing of the initial and final consonants, were significant. Figure 12 displays the F2 accommodation patterns in the *different-L1* condition, regarding IAT, stimulus exposure, and F2 talker distance.

Specifically, when IAT was closer to zero, neither of the polynomial functions of F2 talker distance, namely, the quadratic and linear functions, was significant. However, when IAT became higher, the linear function was significantly negative ($\hat{\beta} = -845.4, p < 0.05$), while the quadratic function was still not significant. Additionally, tense vowels exhibited significantly higher adjusted F2 changes than lax vowels ($\hat{\beta} = 25.6, p < 0.05$).

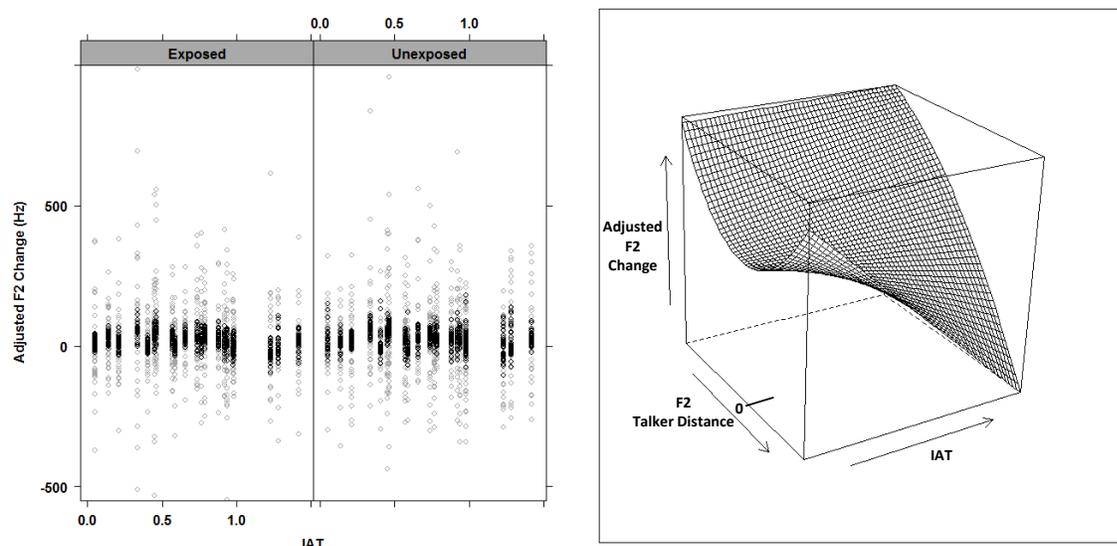


Figure 12. F2 accommodation patterns in the *different-L1* condition

- Note.*
1. In the left plot, the grey dots represent adjusted F2 change data, while the black dots represent the fitted values from the linear mixed effects regression model. For additional details, see section 2.2.4.4.
 2. The right plot is a schematic three-dimension plot with adjusted F2 changes, preexisting F2 talker distance, and IAT scores.
 3. A positive IAT score indicates a negative implicit attitude towards foreigners, and the higher a score is, the more negative the attitude is for foreigners.

In sum, these results indicate that participants' implicit attitudes towards foreigners did not directly impact their degree of F2 convergence. Rather, their negative attitude towards foreigners made the linear function of preexisting F2 talker distance significantly negative. That is, the more negative attitudes participants have towards foreigners, the more likely they would be to show F2 convergence in the direction of F2 decrease.

Now we summarize the results on the influence of implicit attitudes (measured by implicit attitudes towards foreigners) on phonetic accommodation on monosyllabic words in the *different-L1* condition. First, implicit attitudes did not directly affect phonetic accommodation of monosyllabic words. None of the adjusted phonetic changes for each of the acoustic

measurement types in the *different-L1* condition were influenced by participants' IAT scores. Second, however, for all acoustic measurements, it modified the degree of the positive relation between preexisting acoustic talker distance and phonetic convergence. The patterns of the interaction varied depending on the acoustic measurement types. Lastly, with one exception for F1, participants generalized their phonetic accommodation patterns from exposed words to unexposed words both for the degree of convergence and the interaction of implicit attitudes and polynomial functions of acoustic talker distance. This, again, confirms the solidness of phonetic convergence after perceptual exposure as a basis for long-term language learning.

Altogether, the results from monosyllabic words suggest the following:

1. Phonetic convergence on monosyllabic words was observed with all acoustic measurements, namely, VOT, vowel duration, F0-max, F1, and F2. This confirms results from the previous studies that investigated phonetic convergence along these acoustic measurements.
2. As for group level linguistic distance, dialect mismatch inhibited convergence of VOT and vowel duration, but not of F0-max, F1, and F2.
3. Preexisting item level linguistic talker distance facilitated phonetic convergence (i.e. greater distance led to greater convergence) for all acoustic measurements in both the increasing and decreasing directions.
4. For vowel duration and F0-max, the decreasing direction of the given acoustic measurement facilitated phonetic convergence in all group level linguistic distance conditions.
5. Participants' implicit attitude towards foreigners did not impact the degree of phonetic convergence by participants of a nonnative model talker.

6. Rather, attitudes towards foreigners controlled the tendency in 3. For VOT and F1, the more negative the attitudes towards foreigners, the weaker the positive relation between preexisting acoustic distance and phonetic convergence was, while for vowel duration and F0-max, the stronger the positive relation was.
7. Importantly, in most of the cases, participants generalized their degrees of phonetic convergence and the interaction patterns of participants' implicit attitudes towards foreigners and preexisting talker distance from exposed items to new items.

4. DISYLLABIC WORDS

English disyllabic words carry lexical stress patterns of English, and previous studies have measured the ratio of vowel duration, F0, and amplitude between the two syllables (Beckman, 1986; Fry, 1955, 1958; Lieberman, 1960) as acoustic correlates of lexical stress. Specifically, in word pairs with different stress patterns distinguishing a noun and a verb, the stressed syllables were longer in duration, higher in F0, and louder in amplitude. In the current study, we measure these three acoustic correlates of lexical stress in our disyllabic word dataset to see if lexical stress patterns of English realized by the native and nonnative model talkers were a target of phonetic convergence for the participants.

4.1. Methods

4.1.1. Materials

63 English disyllabic words with the stress on the first syllable were chosen to test phonetic accommodation effects on word-level stress. Following the pattern for the initial consonants of monosyllabic words described in Chapter 3, two sets of disyllabic words that started with either bilabial stops (DW1) or alveolar stops (DW2) were made. Table 9 lists all 63 disyllabic words for the phonetic accommodation experiment.

Table 9. Two sets of disyllabic words

Word set	DW1				DW2				
C1 of S1	b		p		d		t		
Voicing of C1 of S2	voiced	voiceless	voiced	voiceless	voiced	voiceless	voiced	voiceless	
V of S1	æ	badger	batter	paddy	packer	dagger	dapper	tabby	tacky
	ɛ	beggar	beckon	pebble	pepper	debit	deafen	tether	tepid
	i	beaver	beaker	peevish	peacock	diva	decaf	teaser	teeter
	ɪ	bigot	bicker	pivot	pitcher	dither	differ	timid	titter
	ɑ	bawdy	bossy	posit	pauper	dodder	docket	toddy	toffee
	ʌ	buggy	buffer	pudgy	pucker	dully	duchy	tunny	touchy
	u	booby	booty	poodle	putto	duma	ducal	tuber	tutor
	ʊ	bully	butcher	pudding	pussy	dura	•	tourist	tootsy

- Note.* 1. DW1 = first set of disyllabic words, DW2 = second set of disyllabic words
 2. C1 of S1 = the initial consonant of the first syllable of disyllabic words
 3. C1 of S2 = the initial consonant of the second syllable of disyllabic words
 4. V of S1 = the vowel of the first syllable of disyllabic words

The 63 disyllabic words also followed the vowel condition of monosyllabic words for the vowel of the first syllable. That is, the vowels of the first syllable of the disyllabic words are /æ, ɛ, i, ɪ, ɑ, ʌ, u, ʊ/ in both of DW1 and DW2. Also, in both of the sets, half of the words had voiced initial stops (/b/ or /d/) as the initial consonant of the first syllable, and the other half had voiceless initial stops as the initial consonant of the first syllable (/p/ or /t/). The first consonants of the second syllable of the disyllabic words were either voiced or voiceless in both sets. Again, following the finding of Goldinger and Azuma (2004) and Goldinger (1998), low frequency words (frequency under 30 per million words in SUBTLEXus (Brysbaert & New, 2009)) were selected. All disyllabic words fulfilled the criterion, with the average frequency of 2.2 per million.

4.1.2. Participants

All 67 participants participated in the phonetic accommodation experiment that included disyllabic words as part of the materials.

4.1.3. Procedure

Participants followed the general procedure described in Chapter 2.

4.1.4. Analyses

Three types of acoustic measurements on the vowels (V1 and V2) of the first and second syllables of the disyllabic word recordings by the model talkers and the participants (pretest and posttest readings) were made using Praat: V2/V1 duration ratio, V2/V1 F0 ratio, and V2/V1 amplitude ratio. The V2/V1 duration ratio was made by dividing the vowel duration (ms) of the second syllable by the vowel duration of the first syllable. For the V2/V1 F0 ratio, first, the maximum F0 values (Hz) of the vowels of the first and second syllables were automatically measured on Pitch objects of Praat, using a script. Any F0-max measurements above 370 Hz or below 85 Hz were manually checked and corrected to avoid pitch-doubling or -halving. Then the maximum F0 of the second vowel was divided by the maximum F0 of the first vowel for the F0 ratio. The maximum F0 values of the first and second vowels were selected instead of the average values for the same reason as the F0-max values of the vowel for monosyllabic words. In detail, in Praat's automatic F0 measurements, there can be many pitch halving and doubling errors. If F0 measurements are averaged over a period, it is hard to check by the extreme error values (above 370 Hz, below 85 Hz) whether any errors are included in the final measurement value. The maximum F0 can be representative for the vowels of a disyllabic word, because the

variation of F0 values is minimal in a list reading format. The V2/V1 amplitude ratio was made in a similar way with the F0 ratio. First, the maximum amplitude values (dB) for the first and second vowels of disyllabic words were measured on Amplitude objects of Praat, and then the maximum amplitude of the second vowel was divided by the maximum amplitude of the first vowel. The maximum amplitude values were chosen instead of averages to make the amplitude measurement parallel with the F0 measurements. For any measurement type, a V2/V1 ratio over 1 indicates that the measure for the second syllable is larger than that for the first syllable, and vice versa. Additionally, a larger V2/V1 ratio would mean that the stress on the first syllable is weaker, while a smaller V2/V1 ratio would indicate that the stress on the first syllable is stronger.

Finally, adjusted posttest-pretest differences for all three types of acoustic measurements were calculated, following the formula described in Chapter 2.

4.2. Results

The complete dataset for disyllabic words consists of 8,694 words (4 model talkers x 63 disyllabic words + 20 control participants x 63 disyllabic words x 2 timings + 47 experimental participants x 63 disyllabic words x 2 timings) for each type of acoustic measurements, V2/V1 duration ratio, V2/V1 F0 ratio, and V2/V1 amplitude ratio (26,082 measurements in total = 8,694 words x 3 measurement types). Among these, in Table 10, the average values of the model talkers for each acoustic dimension are given depending on their group level linguistic distances to their participants.

Table 10. Model talker average values for each acoustic measurement type

Model talker	N1	N2	NN1	NN2
V2/V1 duration ratio	1.76	1.71	2.12	2.46
V2/V1 F0 ratio	0.75	1.1	0.87	0.81
V2/V1 amplitude ratio	0.95	0.94	0.95	0.97

Note. 1. N = native model talker, NN = nonnative model talker.
 2. N1 and N2 served as model talkers in the *same-L1-same-dialect* and *same-L1-different-dialect* conditions, and NN1 and NN2 in the *different-L1* condition.
 3. V1 = the vowel of the first syllable, V2 = the vowel of the second syllable.

In Table 10, we can see that the unstressed vowels of the disyllabic words, namely, V2, had longer durations than V1 for all model talkers. This might be because of domain-final lengthening (Cooper & Paccia-Cooper, 1980; Turk & Shattuck-Hufnagel, 2007) of the second syllables during the list readings without carrier sentences. Note also that NN1 and NN2, the Korean nonnative model talkers, showed larger V2/V1 duration ratios than N1 and N2, the native model talkers. This might be because the rhythmic structure of Korean is “syllable-timed” (J. P. Lee & Jang, 2004; O. Lee & Kim, 2005) or “mora-timed” (Cho, 2004), while that of English is “stress-timed” (Abercrombie, 1967; Pike, 1945; Ramus, Nespor, & Mehler, 2000). In other words, the unstressed vowel in the English disyllabic words might have been less reduced by the Korean nonnative model talkers than by the native model talkers, resulting in their larger V2/V1 duration ratios (i.e. proportionally longer V2 durations) than the native talkers.

The dependent measures of phonetic accommodation, “adjusted phonetic changes”, were calculated by the formula described in Chapter 2.1 with each acoustic measurement for the 8,694 words, resulting in 2,961 data points for disyllabic words (47 experimental participants x 63 disyllabic words) for each measurement type (8,883 data points in total = 2,961 x 3 measurement types). There were no data where the preexisting talker distances were zero, therefore, no data points were excluded from the analyses.

4.2.1. Linguistic talker distance and generalizability of phonetic accommodation

The statistical analyses for linguistic talker distance and generalizability of phonetic accommodation for disyllabic words followed the general methods described in Chapter 2.2.4.1. In all regression models, the quadratic functions of item level acoustic distance improved the model fit, so were included in the final regression models. Additional fixed effect factors were added when they improved the model fit. The condition numbers for multicollinearity of the fixed effect factors included in each of the final regression models were small to moderate ($8.4 < \text{value} < 25.8$), according to Baayen (2008, p. 200). Table 11 shows the additional fixed effect factors and the condition number for multicollinearity of all fixed effect factors for each of the regression models.

Table 11. Additional fixed effect factors and multicollinearity condition number for each regression model

Acoustic measurement	Additional fixed effect factors	Multicollinearity condition number
V2/V1 duration ratio	Word set, voicing of the initial consonant, V1 tenseness	25.73
V2/V1 F0 ratio	Exposed word set, word set, vowel tenseness, V1 tenseness	7.12
V2/V1 amplitude ratio	Word frequency, voicing of the initial consonant, V1 tenseness	8.46

Note. 1. The formula of a model for each acoustic measurement: (adjusted phonetic change) ~ IAT x item level phonetic distance x stimulus exposure + (linear combinations of the additional fixed effect factors) + (1|model talker) + (1|participant) + (1|word)

2. The fixed effect factor, word set, represents the two word sets (DW1 and DW2) that were established for the phonetic accommodation experiment. Participants read both word sets in the pretest and posttest reading phases. The fixed effect factor, exposed word set, represents one of the two word sets, which was exposed to participants during the perceptual exposure phase. Half of the participants were exposed to DW1, and the other half of the participants were exposed to DW2.

4.2.1.1 V2/V1 duration ratio

Single t-test results revealed that adjusted V2/V1 duration ratio changes for exposed and unexposed words in all of the group level linguistic distances were significantly different from zero. Specifically, adjusted V2/V1 duration ratio changes for exposed and unexposed words in the *same-L1-same-dialect* condition were significantly over zero (exposed: $M = 0.18$, $t(409) = 6.22$, $p < 0.05$, unexposed: $M = 0.16$, $t(408) = 5.91$, $p < 0.05$), suggesting significant convergence towards their model talkers. Words in the *same-L1-different-dialect* condition (exposed: $M = 0.09$, $t(314) = 3.23$, $p < 0.05$, unexposed: $M = 0.13$, $t(314) = 5.04$, $p < 0.05$) and in the *different-L1* condition (exposed: $M = 0.09$, $t(754) = 3.52$, $p < 0.05$, unexposed: $M = 0.11$, $t(756) = 4.30$, $p < 0.05$) also showed adjusted V2/V1 duration ratio changes that were significantly over zero. Figure 13 summarizes the V2/V1 duration ratio accommodation patterns for exposed and unexposed words in the three group level linguistic distances.

The ANOVA summary of the linear mixed effects regression model results shows that group level linguistic distance did not significantly affect adjusted V2/V1 duration ratio changes. However, polynomial functions of item level V2/V1 duration ratio talker distance were significant ($F(2, 2938) = 90.14$, $p < 0.025$). Stimulus exposure was not significant, as well as the two-way and three-way interactions with stimulus exposure. This suggests that participants generalized their accommodation patterns regarding V2/V1 duration ratio from exposed words to unexposed words. The two-way interaction between group level linguistic distance and the polynomial functions of item level V2/V1 duration ratio talker distance was significant ($F(4, 2938) = 7.79$, $p < 0.025$). None of the other fixed effect factors, namely, word set and V1 tenseness, were significant in predicting adjusted V2/V1 duration ratio changes.

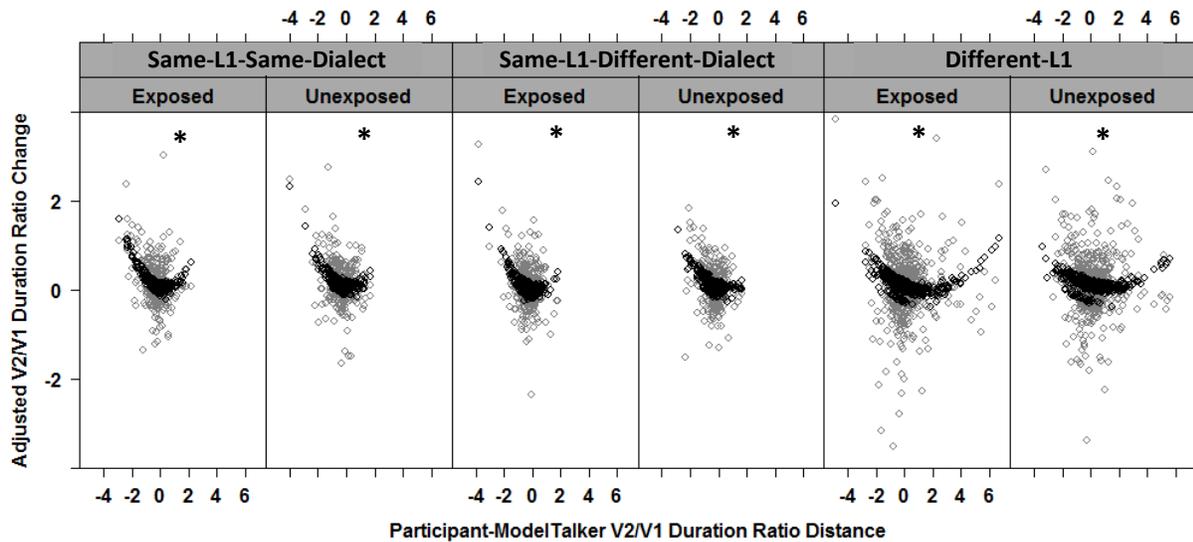


Figure 13. V2/V1 duration ratio accommodation for exposed and unexposed disyllabic words in the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions

- Note.*
1. The grey dots represent adjusted V2/V1 duration ratio change data, while the black dots represent the fitted values from the linear mixed effects regression model.
 2. * denotes that the given dataset of adjusted V2/V1 duration ratio changes (the grey dots) is significantly different from zero.
 3. For additional details, see section 2.2.4.4.

Details of the significant fixed effect factors are as follows. First, the quadratic function of item level V2/V1 duration ratio distance between participants and their model talkers was significantly positive in the *same-L1-same-dialect* condition ($\hat{\beta} = 19.13, p < 0.05$), while the linear function was not significant. In the *different-L1* condition, the quadratic function was significantly flatter ($\hat{\beta} = -13.51, p < 0.05$) and the linear function was significantly more negative ($\hat{\beta} = -8.48, p < 0.05$) than in the *same-L1-same-dialect* condition. The quadratic and linear functions of item level V2/V1 duration ratio in the *same-L1-different-dialect* condition were not significantly from those in the *same-L1-same-dialect* condition.

Taken together, we found that neither dialect mismatch nor L1 mismatch inhibited phonetic convergence of English word-level stress patterns measured with V2/V1 duration ratios. However, the phonetic convergence was controlled by preexisting talker distance along the V2/V1 duration ratios. That is, the farther the acoustic distance was, the larger the degree of convergence was. This also shows that V2/V1 duration ratio convergence occurred in both of increasing and decreasing directions. There was no preference in the direction of change for the degree of phonetic convergence when the model talker was a native model talker. However, when participants had a nonnative model talker, their sensitivity towards preexisting talker distance was lower, compared to participants who had a native model talker. Additionally, in the *different-L1* condition, it was easier to show phonetic convergence in the direction of V2/V1 duration ratio decrease, the direction where the stress of the first syllable would become stronger.

4.2.1.2 V2/V1 F0 ratio

Results from the single *t*-tests shows that adjusted V2/V1 F0 ratio changes for exposed and unexposed words in all of the group level linguistic distances were significantly different from zero. That is, adjusted V2/V1 F0 ratio changes in the *same-L1-same-dialect* condition were significantly over zero (exposed: $M = 0.11$, $t(409) = 6.27$, $p < 0.05$, unexposed: $M = 0.11$, $t(408) = 5.76$, $p < 0.05$), as well as those in the *same-L1-different-dialect* condition (exposed: $M = 0.13$, $t(314) = 7.56$, $p < 0.05$, unexposed: $M = 0.11$, $t(314) = 7.43$, $p < 0.05$) and in the *different-L1* condition (exposed: $M = 0.10$, $t(754) = 6.47$, $p < 0.05$, unexposed: $M = 0.13$, $t(756) = 10.63$, $p < 0.05$). Figure 14 displays the V2/V1 F0 ratio accommodation patterns for exposed and unexposed words in the three group level linguistic distances.

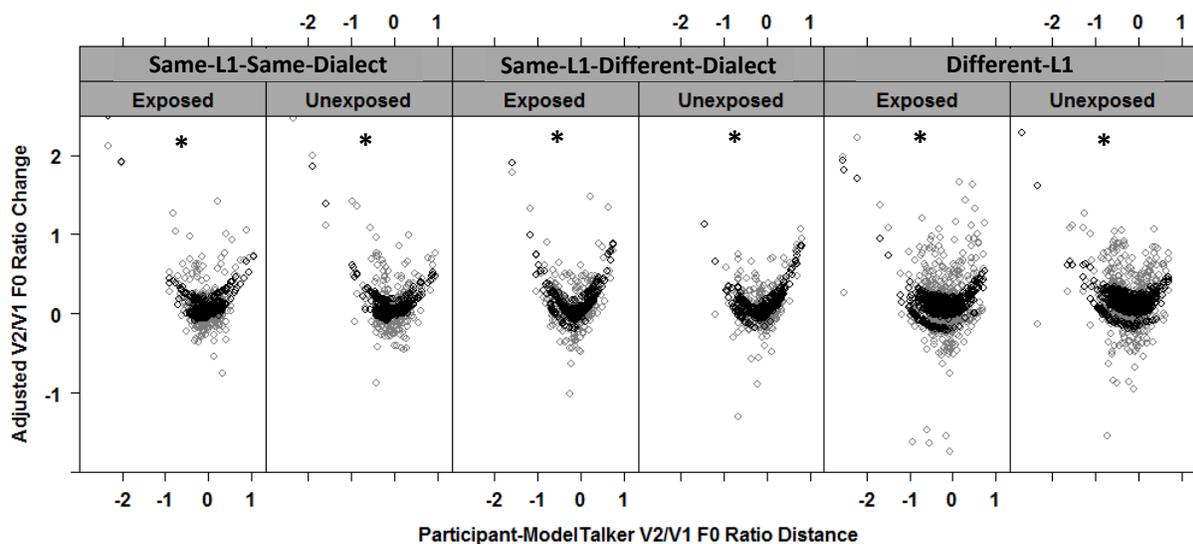


Figure 14. V2/V1 F0 ratio accommodation for exposed and unexposed disyllabic words in the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions

- Note.*
1. The grey dots represent adjusted V2/V1 F0 ratio changes with the grey lowess lines, while the black dots represent the fitted values from the linear mixed effects regression model.
 2. * denotes that the given dataset of adjusted V2/V1 duration ratio changes (the grey dots) is significantly different from zero.
 3. For additional details, see section 2.2.4.4.

The ANOVA summary of the linear mixed effects regression model shows that the polynomial functions of item level V2/V1 F0 ratio talker distance was significant ($F(2, 2938) = 350.25, p < 0.025$), while group level linguistic distance was not. However the interaction between the two factors were significant ($F(4, 2938) = 15.12, p < 0.025$). Stimulus exposure was not significant, suggesting generalization of V2/V1 F0 ratio from exposed words to unexposed words. The two-way and three-way interactions with stimulus exposure were also not significant. None of the other fixed effect factors, namely, exposed word set, word set, and V1 tenseness, were significant.

In detail, both the quadratic ($\hat{\beta} = 8.15, p < 0.05$) and linear functions ($\hat{\beta} = -4.1, p < 0.05$) of item level V2/V1 F0 ratio talker distance were significant in the *same-L1-same-dialect* condition. Additionally, the quadratic function was significantly flatter in the *different-L1* condition than in the *same-L1-same-dialect* ($\hat{\beta} = -2.08, p < 0.05$), and the linear function was significantly more positive in the *same-L1-different-dialect* condition in the *same-L1-same-dialect* ($\hat{\beta} = 5.37, p < 0.05$).

In sum, participants converged towards their model talkers in terms of V2/V1 F0 ratio, regardless of dialect mismatch or L1 mismatch. This phonetic convergence pattern was controlled by the preexisting distances between participants and their model talkers along V2/V1 F0 ratio. That is, the farther the acoustic distance, the larger the degree of V2/V1 F0 ratio convergence was, in both of increasing and decreasing directions. This pattern was weaker in the *different-L1* condition. Additionally, the decreasing direction facilitated V2/V1 F0 convergence in the *same-L1-same-dialect* condition and the *different-L1 condition*. Importantly, participants generalized not only the degree of V2/V1 F0 ratio convergence but also the relation between preexisting acoustic talker distance and convergence from exposed words to unexposed words.

4.2.1.3 V2/V1 amplitude ratio

Single *t*-test results show that adjusted V2/V1 amplitude ratio changes for exposed and unexposed words in all of the group level linguistic distances, namely, the *same-L1-same-dialect* (exposed: $M = 0.02, t(409) = 5.92, p < 0.05$, unexposed: $M = 0.02, t(408) = 6.78, p < 0.05$), *same-L1-different-dialect* (exposed: $M = 0.02, t(314) = 7.99, p < 0.05$, unexposed: $M = 0.02, t(314) = 7.37, p < 0.05$), and *different-L1* conditions (exposed: $M = 0.01, t(754) = 5.99, p < 0.05$, unexposed: $M = 0.02, t(756) = 8.72, p < 0.05$). Figure 15 summarizes the V2/V1 amplitude ratio

accommodation patterns for exposed and unexposed words in the three group level linguistic distances.

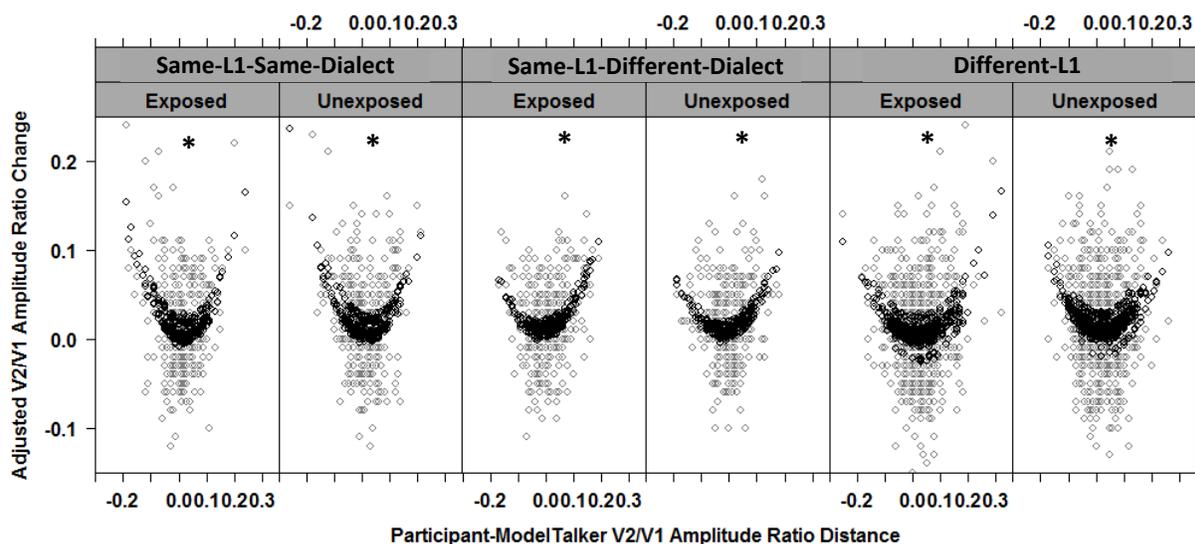


Figure 15. V2/V1 amplitude ratio accommodation for exposed and unexposed monosyllabic words in the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions

- Note.*
1. The grey dots represent adjusted V2/V1 duration ratio change data, while the black dots represent the fitted values from the linear mixed effects regression model.
 2. * denotes that the given dataset of adjusted V2/V1 duration ratio changes (the grey dots) is significantly different from zero.
 3. For additional details, see section 2.2.4.4.

The ANOVA summary of the linear mixed effects regression model with adjusted V2/V1 amplitude ratio shows that the polynomial functions of item level V2/V1 amplitude ratio talker distance were significant ($F(2, 2938) = 109.52, p < 0.025$), while group level linguistic distance was not significant. However, the interaction between the two was significant ($F(4, 2938) = 7.09, p < 0.025$). Stimulus exposure and two-way, three-way interactions with stimulus exposure were

not significant. None of the other fixed effect factors, word frequency, voicing of the initial consonant, and V1 tenseness, were significant.

In detail, the quadratic function of item level V2/V1 amplitude ratio was significantly positive in the *same-L1-same-dialect* condition ($\hat{\beta} = 1.06, p < 0.05$), while the linear function was not significant. The quadratic function was significantly flatter in the *different-L1* condition than in the *same-L1-same-dialect* condition ($\hat{\beta} = -0.38, p < 0.05$), and the linear function was significantly more positive in the *same-L1-different-dialect* condition than in the *same-L1-same-dialect* condition ($\hat{\beta} = 0.53, p < 0.05$). The linear function in the *different-L1* condition and the quadratic function in the *same-L1-different-dialect* condition were not significantly different from those in the *same-L1-same-dialect* condition.

Taken together, we found that participants were sensitive to their item level V2/V1 amplitude ratio distance from their model talkers, not to dialect mismatch or L1 mismatch regarding their V2/V1 amplitude ratio convergence. The farther the acoustic distance, the larger the degree of phonetic convergence was, in both of V2/V1 amplitude ratio increasing and decreasing directions. This pattern was weaker in the *different-L1* condition. Additionally, in the *same-L1-different-L1* condition, it was easier to phonetically converge towards a model talker in the direction of increase. Moreover, participants generalized the degree of V2/V1 amplitude ratio convergence and its interactions with group level and item level linguistic distances from exposed words to unexposed words.

Therefore, from phonetic accommodation patterns for disyllabic words, we found that neither dialect mismatch nor L1 mismatch inhibited phonetic convergence. Rather, the amount of phonetic convergence was controlled by preexisting acoustic talker distance. That is, the farther the acoustic distance, the larger the degree of phonetic convergence was. Importantly, both the

degree of phonetic convergence in each of the group level linguistic distances and the interactions with the polynomial functions of preexisting acoustic distances generalized from old words to new words.

4.2.2. Implicit attitudes and generalizability of phonetic accommodation

The statistical analyses for implicit attitudes and generalizability of phonetic accommodation for disyllabic words followed the general methods described in Chapter 2.2.4.2. In all regression models, the quadratic functions of item level acoustic distance improved the model fit, so were included in the final regression models. Additional fixed effect factors were added when they improved the model fit. The condition numbers for multicollinearity of the fixed effect factors included to each of the final regression models were very small ($6.6 < \text{value} < 6.8$), according to Baayen (2008, p. 200). Table 12 shows the additional fixed effect factors and the condition number for multicollinearity of all fixed effect factors for each of the regression models.

Table 12. Additional fixed effect factors and multicollinearity condition number for each regression model

Acoustic measurement	Additional fixed effect factors	Multicollinearity condition number
V2/V1 duration ratio	Word frequency, V1 tenseness	6.76
V2/V1 F0 ratio	Exposed word set, word set, V1 tenseness	6.70
V2/V1 amplitude ratio	Word frequency, word set, V1 tenseness	6.65

Note. 1. The formula of a model for each acoustic measurement: (adjusted phonetic change) \sim IAT x item level phonetic distance x stimulus exposure + (linear combinations of the additional fixed effect factors) + (1|model talker) + (1|participant) + (1|word)

2. The fixed effect factor, word set, represents the two word sets (DW1 and DW2) that were established for the phonetic accommodation experiment. Participants read both word sets in the pretest and posttest reading phases. The fixed effect factor, exposed word set, represents one of the two word sets, which was exposed to participants during the perceptual exposure phase. Half of the participants were exposed to DW1, and the other half of the participants were exposed to DW2.

4.2.2.1 V2/V1 duration ratio

The ANOVA summary of the linear mixed effects regression model results shows that the polynomial functions of V2/V1 duration ratio ($F(2, 1498) = 33.05, p < 0.025$) and its interaction with IAT scores ($F(2, 1498) = 13.38, p < 0.025$) were significant. However, IAT scores did not significantly affect the adjusted V2/V1 duration ratio changes. Stimulus exposure and its interactions with IAT and the polynomial functions were not significant. This indicates that participants applied what they learned during the passive exposure phase to new words both for the degree of convergence and its interactions with the other factors. The other fixed effect factors, word frequency and V1 tenseness were not significant. Figure 16 shows the V2/V1 duration ratio accommodation patterns regarding stimulus exposure, IAT, and preexisting acoustic talker distance.

In detail, the polynomial functions were not significant in the *same-L1-same-dialect* condition. However, they interacted significantly with IAT scores. When IAT scores were high, the quadratic function of item level V2/V1 duration ratio talker distance was significantly positive ($\hat{\beta} = 7.18, p < 0.05$), and the linear function was significantly negative ($\hat{\beta} = -4.38 p < 0.05$).

Taken together, these results suggest that participants' negative implicit attitudes towards foreigners did not directly affect their V2/V1 duration ratio convergence, rather, they influenced the polynomial functions of preexisting acoustic talker distances. That is, when participants' negative attitudes towards foreigners were higher, the positive influence of preexisting acoustic talker distances on the degree of phonetic convergence was stronger. Additionally, when participants had more negative attitudes towards foreigners, it was easier to converge towards model talkers in the V2/V1 duration ratio decreasing direction.

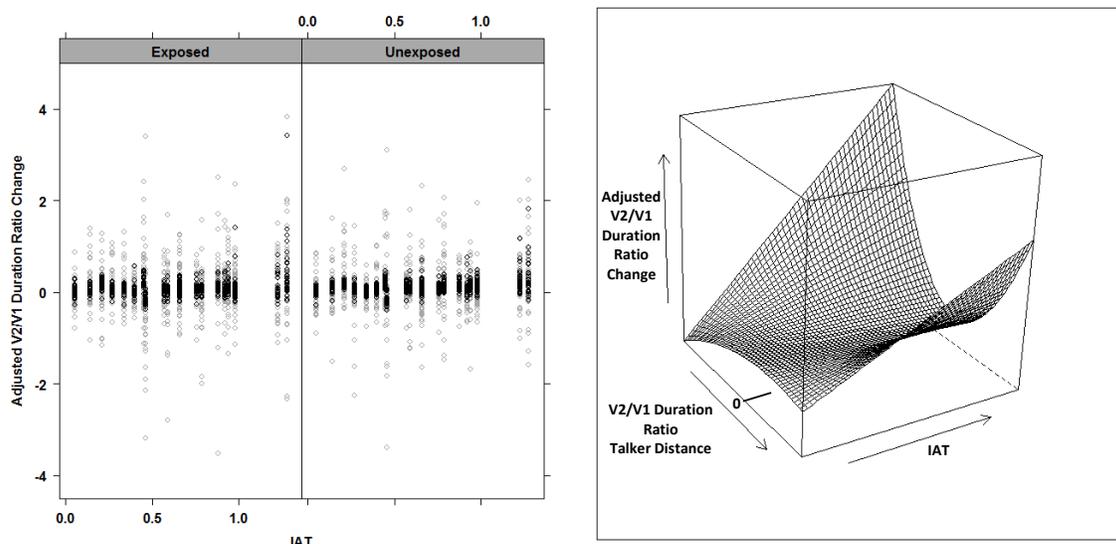


Figure 16. V2/V1 duration ratio accommodation patterns in the *different-L1* condition

- Note.*
1. In the left plot, the grey dots represent adjusted V2/V1 duration ratio change data, while the black dots represent the fitted values from the linear mixed effects regression model. For additional details, see section 2.2.4.4.
 2. The right plot is a schematic three-dimension plot with adjusted V2/V1 duration ratio changes, preexisting V2/V1 duration ratio talker distance, and IAT scores.
 3. A positive IAT score indicates a negative implicit attitude towards foreigners, and the higher a score is, the more negative the attitude is for foreigners.

4.2.2.2 V2/V1 F0 ratio

Results from the linear mixed effects regression model suggests that polynomial functions of item level V2/V1 F0 ratio distance between participants and their model talkers ($F(2, 1498) = 129.05, p < 0.025$) and their interaction with IAT scores ($F(2, 1498) = 11.33, p < 0.025$) were significant in predicting adjusted V2/V1 F0 ratio changes. However, IAT did not significantly affect adjusted V2/V1 F0 ratio changes. Stimulus exposure and its two-way, three-way interactions with IAT and the polynomial functions were not significant. The other fixed effect factors, exposed word set, word set, and V1 tenseness, were not significant. Figure 17 shows the

V2/V1 F0 ratio accommodation patterns in the *different-L1* condition, regarding stimulus exposure, IAT and the item level V2/V1 F0 ratio talker distance.

In detail, the quadratic function of item level V2/V1 F0 ratio talker distance was significantly positive ($\hat{\beta} = 2.80, p < 0.025$), while the linear function was not significant. When IAT scores were high, the quadratic function was significantly stronger ($\hat{\beta} = 3.73, p < 0.025$), and the linear function was significantly negative ($\hat{\beta} = -3.49, p < 0.025$).

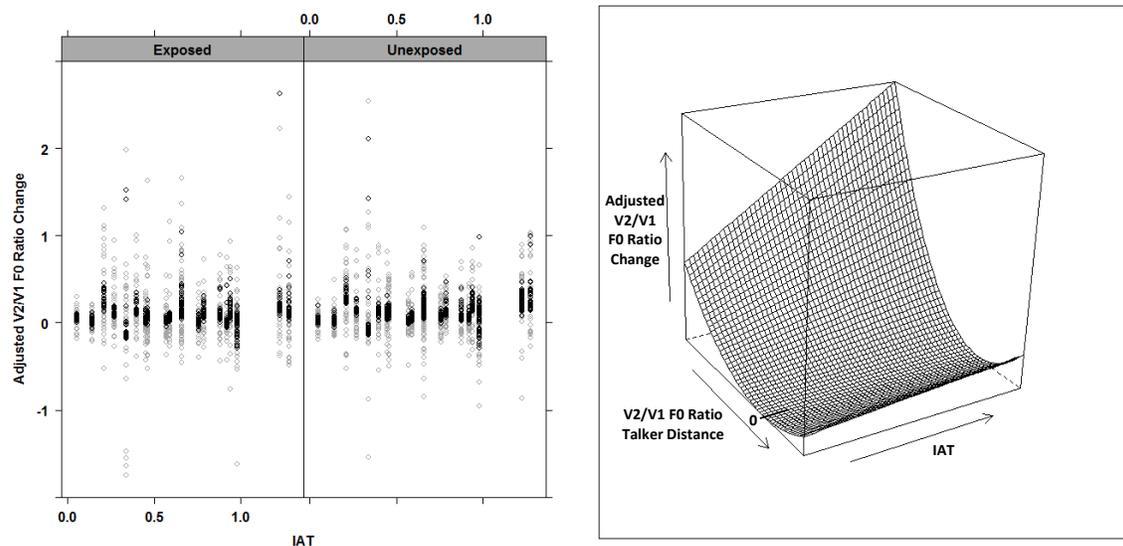


Figure 17. V2/V1 F0 ratio accommodation patterns in the *different-L1* condition

- Note.*
1. In the left plot, the grey dots represent adjusted V2/V1 F0 ratio changes with the grey loess lines, while the black dots represent the fitted values from the linear mixed effects regression model. For additional details, see section 2.2.4.4.
 2. The right plot is a schematic three-dimension plot with adjusted V2/V1 F0 ratio changes, preexisting V2/V1 F0 ratio talker distance, and IAT scores.
 3. A positive IAT score indicates a negative implicit attitude towards foreigners, and the higher a score is, the more negative the attitude is for foreigners.

In sum, participants' negative attitudes towards foreigners affected the relation between preexisting V2/V1 F0 ratio and the degree of convergence. That is, the more negative

participants' attitudes towards foreigners, the stronger the positive relationship between preexisting V2/V1 F0 ratio talker distance and their degree of convergence. Additionally, participants who had more negative attitudes towards foreigners than others were likely to show a larger degree of V2/V1 F0 ratio convergence in the decreasing direction.

4.2.2.3 V2/V1 amplitude ratio

The ANOVA summary of the linear mixed effects regression model with V2/V1 amplitude ratio suggests that polynomial functions of item level V2/V1 amplitude ratio talker distance ($F(2, 1498) = 43.57, p < 0.025$) and V1 tenseness ($F(1, 1498) = 5.40, p < 0.025$) significantly affected adjusted V2/V1 amplitude ratio changes. However, IAT scores and stimulus exposure were not significant. None of the two-way, three-way interactions with item level V2/V1 amplitude ratio talker distance, stimulus exposure, and IAT scores was significant. The other fixed effect factors, word frequency and word set were not significant. Figure 18 displays V2/V1 amplitude ratio accommodation patterns in the *different-L1* condition, regarding stimulus exposure, IAT, and the preexisting V2/V1 amplitude ratio talker distance.

In detail, the quadratic function of item level V2/V1 amplitude ratio talker distance was significantly positive ($\hat{\beta} = 0.59, p < 0.025$), while the linear function was not significant. Additionally, V2/V1 amplitude ratios of disyllabic words with a tense vowel for the first syllable showed significantly larger degree of phonetic convergence ($\hat{\beta} = 0.006, p < 0.025$).

In sum, participants were sensitive to preexisting V2/V1 amplitude ratio talker distance, regarding their phonetic convergence. That is, the farther the distance, the larger the degree of V2/V1 amplitude ratio, in both decreasing and increasing directions. Moreover, participants generalized the degree of convergence and the interaction between convergence and preexisting

V2/V1 amplitude ratio from exposed words to unexposed words. However, the IAT scores did not affect phonetic accommodation of V2/V1 amplitude ratio of disyllabic words in any matters.

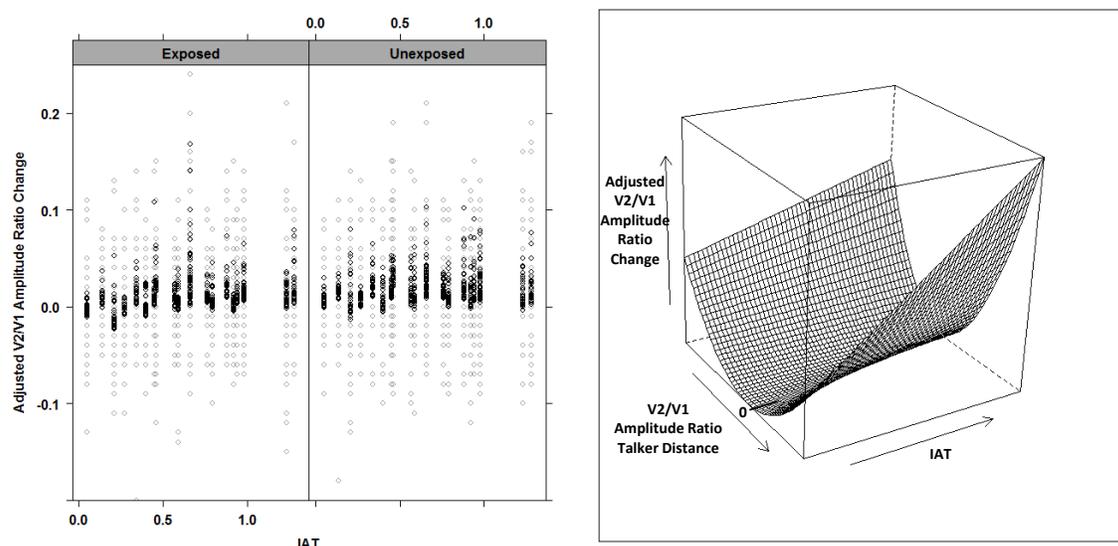


Figure 18. V2/V1 amplitude ratio accommodation patterns in the *different-L1* condition

- Note.*
1. In the left plot, the grey dots represent adjusted V2/V1 amplitude ratio change data, while the black dots represent the fitted values from the linear mixed effects regression model. For additional details, see section 2.2.4.4.
 2. The right plot is a schematic three-dimension plot with adjusted V2/V1 amplitude ratio changes, preexisting V2/V1 amplitude ratio talker distance, and IAT scores.
 3. A positive IAT score indicates a negative implicit attitude towards foreigners, and the higher a score is, the more negative the attitude is for foreigners.

In total, from various acoustic measurements of word-level stress realization in disyllabic words, we found that only preexisting acoustic talker distance affects phonetic accommodation directly. That is, the farther the acoustic distance is, the larger the degree of phonetic convergence is. Dialect mismatch, L1 mismatch, and participants' negative attitudes towards foreigners changed the influence of preexisting acoustic talker distance on the degree of phonetic

convergence in some cases. Lastly, both the degree of phonetic convergence and the interaction patterns with preexisting acoustic talker distance were generalized from old words to new words.

5. SENTENCES

As discussed in Chapter 2, using acoustic measurements to discover phonetic accommodation patterns has a benefit of clearly showing on which acoustic parameter the accommodation occurred. However, one limitation of this method is that accommodation patterns judged by different individual acoustic parameters are not necessarily consistent within the same speech materials (as evidenced in the analyses with mono- and dis-syllabic words in the preceding chapters). While it is still worthwhile to measure acoustic correlates of phonetic accommodation, an alternative is to examine phonetic accommodation with more “holistic” analyses that capture various features of speech at the same time. To do this, we attempted measuring phonetic accommodation with human perception and pattern-based rather than feature-based digital signal processing, namely, an XAB perception test and the dynamic time warping (DTW) technique. For these analyses we focused on the sentence materials which provide ample opportunity for longer-term characteristics of the utterances to exert an influence on these comparison techniques.

5.1. Methods

5.1.1. Materials

Two sets of 32 sentences, 64 sentences in total, were made as sentence materials for the phonetic accommodation experiment (see Table 13). In each set, half of the 32 sentences were newly made for this experiment, and the other half were selected from the Speech Perception in Noise test (Kalikow et al., 1977). As introduced in Chapter 2, the new sentences are sentences where the verb starts with a bilabial stop (/b/ or /p/) in S1 and an alveolar stop (/d/ or /t/) in S2. Also, the vowel of the verb in the sentences, /æ, ε, i, ɪ, α, ʌ, u, ʊ/, were controlled to be the same across S1 and S2.

Table 13. Two sets of sentences

	V of Verb	S1	S2
New sentences	æ	The man buzzed the door bell.	The guy dumped the old garbage.
	ɛ	The boys bent the formal rules.	The designer dressed the models.
	i	The developer built the new building.	The workers drilled the oil well.
	ɪ	The gangs bullied the little girl.	The woman dipped the dry biscuit.
	ɑ	The guys bought the big bed.	The gunman dodged the fast bullet.
	ʌ	The girl banged the glass window.	The boy dragged the big items.
	u	The boy beat the gray wall.	The mechanic detailed the red van.
	ʊ	The wind blew the fallen leaves.	The recession doomed the world market.
	æ	The man punched the bus driver.	The guy trusted the generous judge.
	ɛ	The bird pecked the wooden bridge.	The general terrified the enemy force.
	i	The director picked the young actor.	The lady tipped the young waiter.
	ɪ	The guy pulled the big weeds.	The archer took the big bow.
	ɑ	The baby popped the big balloon.	The girl tossed the white ball.
	ʌ	The buses passed the new buildings.	The dogs trapped the little mouse.
	u	The woman peeled the fresh oranges.	The man teased the angry boy.
	ʊ	The grads pooled the new datasets.	The driver tuned the old radio.
Selected from the Speech in Noise test		Bob was cut by the jackknife's blade.	A pigeon is a kind of bird.
		Our seats were in the second row.	The judge is sitting on the bench.
		He's employed by a large firm.	Football is a dangerous sport.
		The car drove off the steep cliff.	Keep your broken arm in a sling.
		A racecar can go very fast.	Please wipe your feet on the mat.
		The team was trained by their coach.	The swimmer's leg got a bad cramp.
		A spoiled child is a brat.	The cut on his knee formed a scab.
		The guests were welcomed by the host.	A wristwatch is used to tell the time.
		The mouse was caught in the trap.	The detectives searched for a clue.
		Spread some butter on your bread.	Elephants are big animals.
		A quarter is worth twenty-five cents.	Rain falls from clouds in the sky.
		Greet the heroes with loud cheers.	The color of a lemon is yellow.
		The scarf was made of shiny silk.	That job was an easy task.
		Monday is the first day of the week.	The cabin was made of logs.
		He was scared out of his wits.	Paul was arrested by the cops.
		Playing checkers can be fun.	The good boy is helping his mother and father.

The semantic probability of the new sentences was controlled. That is, the verb and object in each target sentence share coherent meanings, so that reading of the sentences can be facilitated by semantic expectation. The filler sentences were also high probability sentences in the Speech Perception in Noise test (Kalikow et al., 1977).

5.1.2. Participants

All 67 participants described in Chapter 2 participated in the phonetic accommodation experiment that included sentences as part of the materials.

5.1.3. Procedure

All participants followed the general procedure described in Chapter 2, the General Methodology.

5.1.4. Analyses

The complete sentence dataset from the phonetic accommodation experiment consisted of 11,584 sentences (20 control participants x 64 sentences x 2 timings + 47 experimental participants x 64 sentences x 3 timings).

5.1.4.1 Perceptual Assessment: XAB Perception Test

A traditional holistic approach in analyzing phonetic accommodation is conducting an XAB perception test. As mentioned in Chapter 1, XAB perception tests have been widely used in previous research on phonetic accommodation and have provided ample evidence of phonetic accommodation in different experimental settings and sociolinguistic variations (Goldinger, 1998; Goldinger & Azuma, 2004; Kim et al., 2011; Pardo, 2006; Pardo et al., 2010). We conducted an XAB perception test on part of our sentence dataset.

5.1.4.1.1 *Participants*

A separate set of 55 native English talkers participated in the XAB perception test as perceptual judges for the phonetic accommodation test with the sentences. All participants were undergraduate students at Northwestern University. Out of the 55 participants, 36 were female, 19 were male. Their ages ranged from 18 to 21 years, with an average of 20.1 years. As there were four model talkers, participants were randomly assigned to each model talker condition: 16 participants to the N1 condition, 13 participants to the N2 condition, 14 participants to the NN1 condition, and 12 participants to the NN2 condition.

5.1.4.1.2 *Stimuli*

From the total set of sentence recordings from the phonetic accommodation experiment, only the data from participants who were auditorily exposed to a model talker in the experimental groups were used for the XAB perception test. There were 47 participants in total in the four experimental groups, and only the data from 10 participants per experimental group (40 participants in total) were selected for the XAB perception test. This was to keep the same number of participants per experimental group for the stimuli in the XAB perception test, and also to limit the total XAB perception experiment time to under 2 hours. From each participant's data, only the pretest and posttest sentence readings that matched sentences that the participant heard in the model talker's voice during the exposure phase (exposed items = 32 sentences per participant) were selected. Also, their model talker recordings of these sentences were included in the test. No data from the control group or from unexposed items were included in the XAB perception test. Therefore, 3,840 sentence recordings in total (40 experimental participants x 32

sentences x 3 timings (pretest, model, posttest)) were used as stimuli in the XAB perception test. All recordings were normalized to have the same overall RMS value (1.0 Pa).

5.1.4.1.3 Procedure

Participants were seated in front of a computer monitor in a soundproof booth. In each trial, the participant heard a triplet of sentence recordings (XAB) played over headphones. The stimuli were presented from the computer running Millisecond Inquisit 3.0.4.0. In each trial, three letter boxes (X, A, and B) were displayed on the monitor, X at the top, A at the left bottom, and B at the right bottom. Participants were instructed to select the sentence recording (A or B) that sounded more similar to the first sentence recording (X) by clicking a letter box, A or B, with a mouse. The inter-sample interval was 100 ms.

The pretest and posttest recordings by the participants in the experimental groups in the phonetic accommodation experiment were used as stimuli for A or B in the XAB perception test. The order of the stimuli for A and B was counterbalanced. There were four different conditions for each of the four model talkers in the phonetic accommodation experiment. Within one condition, there were 10 blocks for 10 participants per model talker. Within one block, there were 64 trials for counterbalancing of the 32 sentences. Within each trial, one of the 32 sentences was presented three times, the model talker's recording (X) plus a participant's pretest, and posttest recordings (A and B). The inter-trial interval was 600 ms. Between each block, the participants were allowed to take a short break, if they wanted. In this way, a participant was presented with 640 trials in total (10 blocks x 32 sentences x 2 orders). The 10 blocks and the 64 trials within a block were each presented in random orders. It took approximately 8-10 minutes for each block and 1.5-2 hours for a participant to finish the XAB perception test.

5.1.4.2 Dynamic Time Warping (DTW)

Although it suggests phonetic accommodation patterns based on relatively robust human perception, an XAB perception test is time-consuming and requires a large number of participants in long and tedious test sessions. Another holistic measurement of phonetic accommodation is based on a computational comparisons of the temporal and spectral characteristics of two digital speech signals, namely, dynamic time warping (DTW) analyses (Berndt & Clifford, 1994; Turetsky & Ellis, 2003). In the current study, DTW analyses were conducted on the sentence data from the phonetic accommodation experiment, using Matlab code for DTW and Mel-frequency Cepstral Coefficients (MFCC) (Ellis, 2003, 2005). DTW is a signal recognition technique designed to quantify timing differences between two signals. Comparing spectral slices of the two signals with small steps in the time domain, it finds the path that maximizes the temporal alignment of the two signals and calculates the “cost” to achieve the alignment, the so-called “similarity cost”. This is a good measure to indicate the alignment distance between two speech signals in the temporal domain. If a similarity cost is small, the two signals would be similarly aligned to each other, and if large, the two signals would be distant from each other. MFCC slices were used instead of spectral slices to reflect human speech perception in the DTW analyses.

To examine phonetic accommodation patterns on sentence intonations, “hums” of the sentences were analyzed with DTW as well as the full (i.e. original) sentence recordings. The hums of the sentences were made in Praat with the Pitch objects described above. In this way, the F0 contours of the sentences were extracted from the sentence recordings and realized as sound signals, representing the intonation patterns of the sentences. As these hums included pure digital silence periods, white noise was added to the entire duration of each of the hum sound

signals. This process made the hum sounds sound more natural and suitable for analysis by the DTW Matlab code. The complete data for DTW consisted of two datasets, the raw sentence recordings from the experimental groups (9,024 sentences = 47 experimental participants x 64 sentences x 3 timings) and the hums of the sentences.

The Matlab DTW code was run on each sentence or hum twice: first with pretest recordings by participants and their model talker recordings and second with the participants' posttest recordings and the model talker recordings. Thus, two similarity costs were calculated for each sentence for each participant, the pretest-model distance and the posttest-model distance. Then DTW distance change was calculated by subtracting the posttest-model distance from the pretest-model distance to indicate the amount of phonetic accommodation on the sentence by the participant. If the DTW distance change is above 0, it would mean that the participant-model distance was smaller in the posttest than in the pretest, therefore, phonetic convergence. If the DTW distance change is below 0, it would mean phonetic divergence, as the participant-model distance was larger in the posttest than in the pretest. If the change value is 0, it would mean phonetic maintenance.

Note that the DTW analyses cannot follow the revisited definition of phonetic accommodation introduced in Chapter 2. This is because DTW similarity costs suggest only the absolute distance between two signals and do not show the direction of the difference or change. In other words, there is no consideration of the directions of pretest-model differences and pretest-posttest changes in DTW analyses, thus adjusted phonetic changes cannot be calculated with DTW similarity costs. Therefore, we applied a more traditional scheme for phonetic accommodation (Babel, 2012) on the DTW analyses: the difference between the absolute pretest-model distance and the absolute posttest-model distance. Data from the control group could not

be included in the DTW analyses, either, because there were no model talker values in the control group data.

5.2. Results

The complete dataset for sentence analyses consisted of 8,832 sentences (4 model talkers x 64 sentences + 20 control participants x 64 sentences x 2 timings + 47 experimental participants x 64 sentences x 2 timings). Data from control participants and pretest and posttest recordings on unexposed items were not used for the XAB perception test. Therefore, 2,688 sentence recordings in total (4 model talkers x 32 sentences + 40 experimental participants x 32 sentences x 2 timings) were used in the XAB perception test. For the DTW analyses, data from control participants were excluded, resulting in 6,272 sentences analyzed (4 model talkers x 64 sentences + 47 experimental participants x 64 sentences x 2 timings).

5.2.1. Perceived phonetic accommodation

First, single *t*-tests were performed on averaged posttest sample selection rates across talkers for each group level linguistic distance. In the *same-L1-same-dialect* condition, the posttest sample selection rate was significantly over zero ($M = 54.17\%$, $t(319) = 3.5$, $p < 0.05$). This was the same in the *same-L1-different-dialect* condition ($M = 55.27\%$, $t(319) = 4.82$, $p < 0.05$) and in the *different-L1* condition ($M = 54.64\%$, $t(607) = 6.13$, $p < 0.05$). That is, the XAB listeners selected posttest samples significantly more often than pretest samples as the better match to the model talker samples in all of the group level linguistic distances. Figure 17 summarizes the perceived phonetic convergence patterns in the three group level linguistic distances.

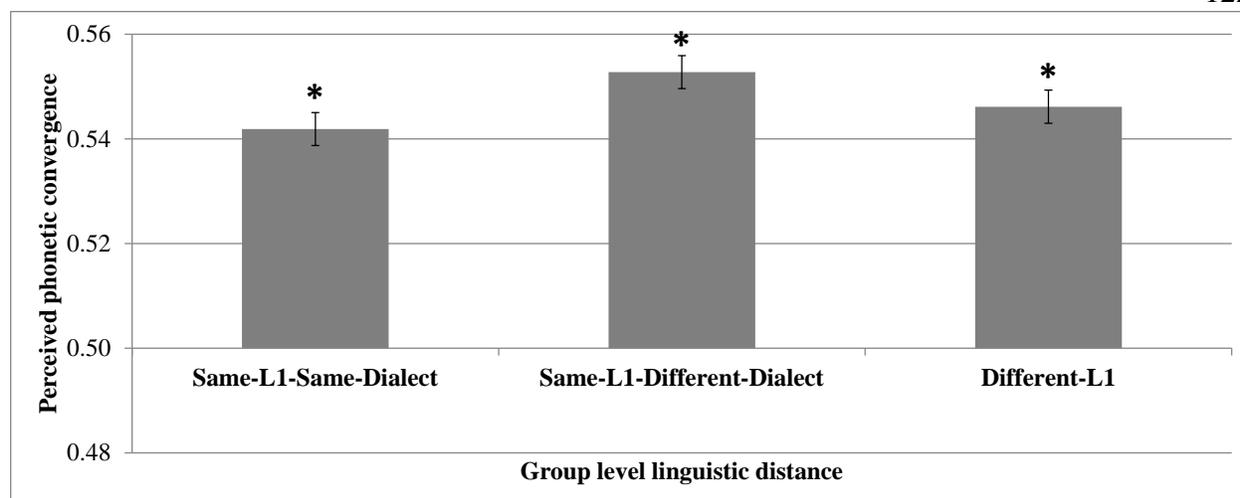


Figure 19. Results of XAB perception test

Note. * denotes the posttest item selection rate is significantly different from 0.5.

A generalized linear mixed effects regression model was built with the logit link function and binomial variance (Bates et al., 2011) to see whether there were significant difference in posttest sample selection rates across different group level linguistic distances. The dependent measure was the raw binary responses (pretest, posttest) from the XAB perception test. Group level linguistic distance (*same-L1-same-dialect*, *same-L1-different-dialect*, *different-L1*) was added as the fixed effect factor. For full comparisons among the three levels, the same regression model was run twice with different reference levels, once with *same-L1-same-dialect* and the other time with *same-L1-different-dialect* as the reference level. The significance level was adjusted from 0.05 to 0.025 by Bonferroni correction. Participants' age, exposed sentence set, and sentence type were not included in the final regression model, since they did not improve the model fit. Model talkers, participants, and sentences were added as random effect factors.

Results suggest that none of the group level linguistic distance levels were significantly different from one another in their posttest sample selection rates. Therefore, this confirms that dialect mismatch and L1 mismatch between participants and their model talkers did not inhibit their perceived phonetic convergence for sentences.

5.2.2. Dynamic time warping analyses

Single *t*-tests were performed on each dataset for exposed and unexposed sentences in the three group level linguistic distances for full sentence DTW and hum DTW. First, in the full sentence DTW results, the DTW distance changes in the exposed and unexposed sentences in the *same-L1-same-dialect* condition and in the *different-L1* condition were significantly different from zero. In the *same-L1-same-dialect* condition, the DTW distance changes were significantly under zero (exposed: $M = -0.0019$, $t(415) = -2.63$, $p < 0.05$, unexposed: $M = -0.0053$, $t(415) = -6.90$, $p < 0.05$), suggesting phonetic divergence. In the *different-L1* condition, the DTW distance changes were significantly over zero (exposed: $M = 0.0019$, $t(767) = 3.19$, $p < 0.05$, unexposed: $M = 0.0023$, $t(767) = 3.16$, $p < 0.05$), suggesting phonetic convergence. The DTW distance changes for exposed and unexposed words in the *same-L1-different-dialect* condition were not significantly different from zero, suggesting phonetic maintenance. Figure 20 summarizes the full sentence DTW accommodation patterns.

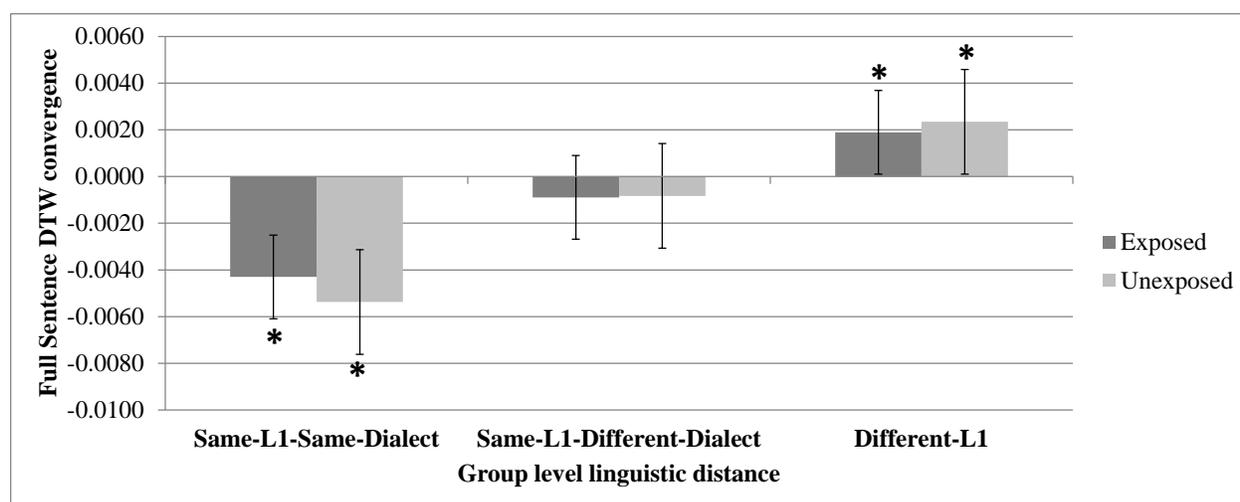


Figure 20. Full sentence DTW accommodation in the *same-L1-same-dialect*, *same-L1-different-dialect*, *different-L1* conditions

Note. * denotes the posttest item selection rate is significantly different from 0.

A linear mixed effects regression model was built with DTW distance changes from the full sentence DTW analyses as the dependent measure. Group level linguistic distance, stimulus exposure, and sentence type were included as fixed effect factors. The interaction of group level linguistic distance and stimulus exposure was also included. Stimulus exposure and sentence type was contrast coded. The same regression model was run twice with different reference levels for group level linguistic distance. This was to see the full comparison for group level linguistic distance. The significance level was adjusted from 0.05 to 0.025 by Bonferroni correction. Model talkers, participants, and sentences were added as random effect factors.

Results suggest that the DTW distance changes in the *different-L1* condition was significantly higher than those in the *same-L1-same-dialect* condition ($\hat{\beta} = 0.007, p < 0.025$). However, the *same-L1-different-dialect* condition was different from neither of the other group level linguistic distance conditions. Stimulus exposure was not significant. This suggests that participants generalized their phonetic accommodation patterns from exposed sentences to unexposed sentences. Sentence type was not significant, as well as the two-way interaction between group level linguistic distance and stimulus exposure.

These results suggest that L1 mismatch facilitated phonetic convergence, while L1 match inhibited phonetic convergence. On the other hand, dialect mismatch did not facilitate phonetic convergence. These are strikingly different results from the previous analyses. That is, in the other measures for phonetic accommodation, it was never the case that L1 match inhibited phonetic convergence or facilitated phonetic divergence.

Figure 21 shows the phonetic accommodation patterns of hums judged by DTW distance changes. In the hum DTW results, the DTW distance changes for exposed words in the *same-L1-same-dialect* condition and for exposed and unexposed words in the *different-L1* condition were

significantly different from zero. In the *same-L1-same-dialect* condition, the DTW distance changes for exposed words were significantly below zero ($M = -0.002$, $t(415) = -2.63$, $p < 0.05$), indicating phonetic divergence. In the *different-L1* condition, the DTW distance changes were significantly over zero both for exposed and unexposed words (exposed: $M = 0.0019$, $t(767) = 4.04$, $p < 0.05$, unexposed: $M = 0.0014$, $t(767) = 2.88$, $p < 0.05$), suggesting phonetic convergence. DTW distance changes for unexposed words in the *same-L1-same-dialect* condition and for exposed and unexposed words in the *same-L1-different-dialect* condition were not significantly different from zero, indicating phonetic maintenance.

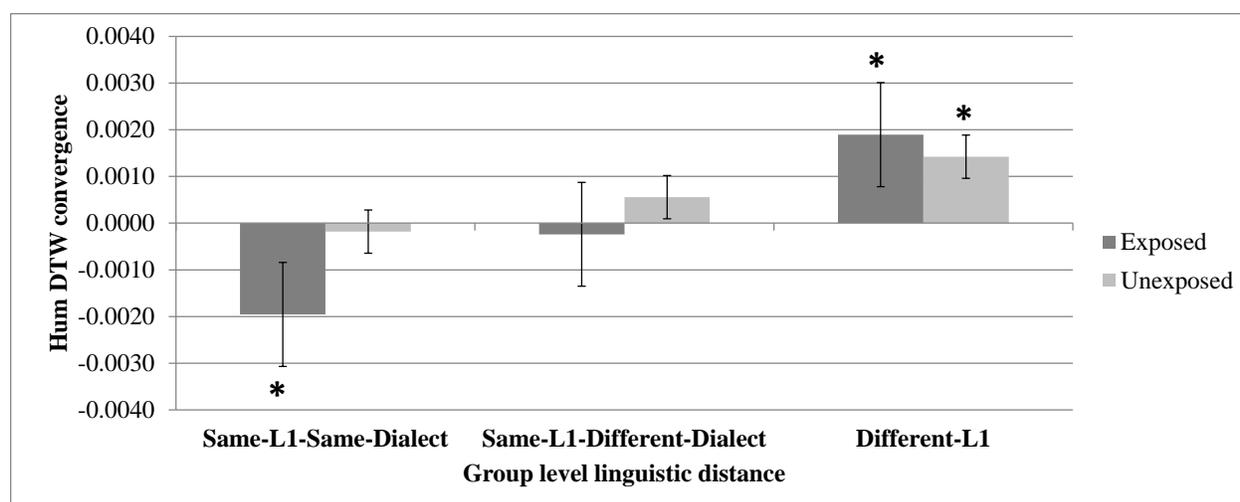


Figure 21. Hum DTW accommodation in the *same-L1-same-dialect*, *same-L1-different-dialect*, *different-L1* conditions

Note. * denotes the posttest item selection rate is significantly different from 0.

A linear mixed effects regression model was built with hum DTW distance changes as the dependent measure. Group level linguistic distance, stimulus exposure, exposed sentence set, and the two-way interaction between group level linguistic distance and stimulus exposure were included as fixed effect factors. Stimulus exposure and exposed sentence set were contrast coded.

For full comparisons among the three levels of group level linguistic distance, the same regression model was run twice with different reference levels. Accordingly, the significance level was adjusted from 0.05 to 0.025. Model talkers, participants, and sentences were added as random effect factors.

Results show that the hum DTW distance changes in the *different-L1* condition were significantly higher than those in the *same-L1-same-dialect* condition ($\hat{\beta} = 0.003, p < 0.025$). However, the hum DTW distance changes in the *same-L1-different-dialect* condition were not significantly different from those in the *same-L1-same-dialect condition*. Stimulus exposure was not significant, suggesting the generalization of hum DTW accommodation from exposed sentences to unexposed sentences. Exposed sentence set was not significant, as well as the interaction between stimulus exposure and group level linguistic distance.

Taken together, we found that L1 mismatch facilitated phonetic convergence, L1 match lead to phonetic divergence, and dialect mismatch resulted in phonetic maintenance in the hum DTW analyses. This is the same pattern with that in the full sentence DTW analyses. Again, we also found that participants generalized phonetic convergence from exposed sentences to unexposed sentences.

5.2.3. Prediction of perceived phonetic convergence with mechanically judged phonetic convergence

Although the DTW results suggested different patterns of phonetic accommodation from human perception for the group level linguistic distances, they might still contribute in explaining the perceived accommodation pattern. Figure 22 displays scatterplots of the XAB perception test results and the DTW analyses results in the *same-L1-same-dialect*, *same-L1-different-dialect*,

and *different-L1* conditions. Here, we see that there might be some relations between perceived phonetic convergence and the DTW analyses results.

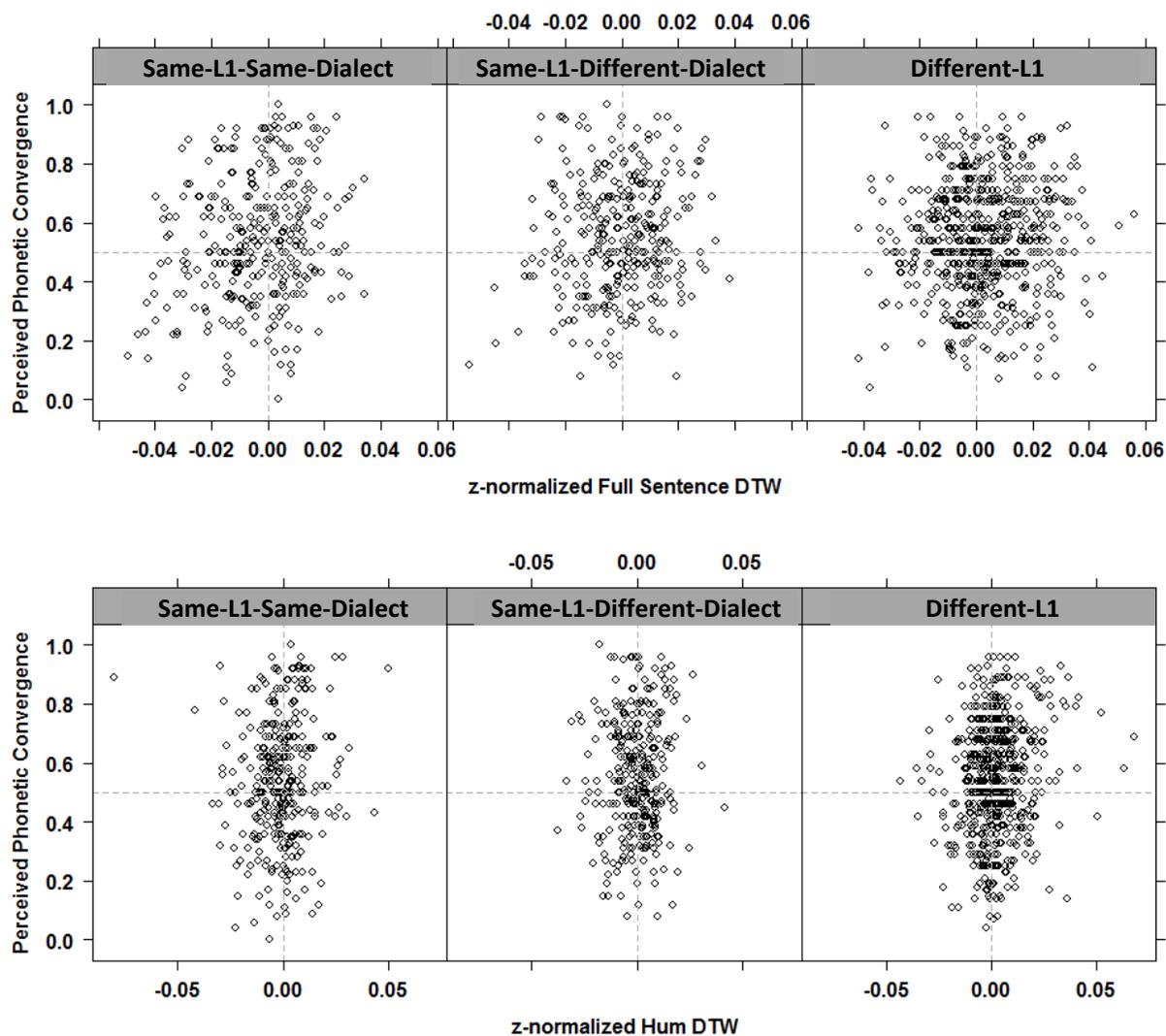


Figure 22. Scatterplots of z -normalized full sentence DTW and XAB perception test results (top) and scatterplots of z -normalized hum DTW and XAB perception test results (bottom) in the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions

Note. The dashed vertical and horizontal grey lines denote the criteria for phonetic convergence by the XAB perception test and the DTW analyses, respectively. Values that are larger than the criteria represent phonetic convergence in each dimension.

A generalized mixed effects regression model was built with the logit link function and binomial variance to see whether the DTW results and the factors of the model talker and participants could predict the perceived accommodation patterns. The dependent measure was the raw binary responses (pretest, posttest) from the XAB perception test. The reference level of the dependent measure was pretest. Therefore, a positive coefficient for a fixed effect factor would mean a better likelihood of posttest selections by listeners of the XAB perception test. Fixed effect factors were group level linguistic distance and z -normalized DTW distance changes on full sentences and hums. The two-way interactions between group level linguistic distance and full sentence DTW distance changes and between group level linguistic distance and hum DTW distance changes were also included. Participants' age, exposed set, sentence set, and sentence type did not change the model fit significantly, so were not included in the model. Critically, stimulus exposure was not added to the regression model, because, in the XAB perception test, all stimuli were from exposed items in the phonetic accommodation experiment. Model talkers, sentences, talkers, and XAB listeners were included as random effect factors.

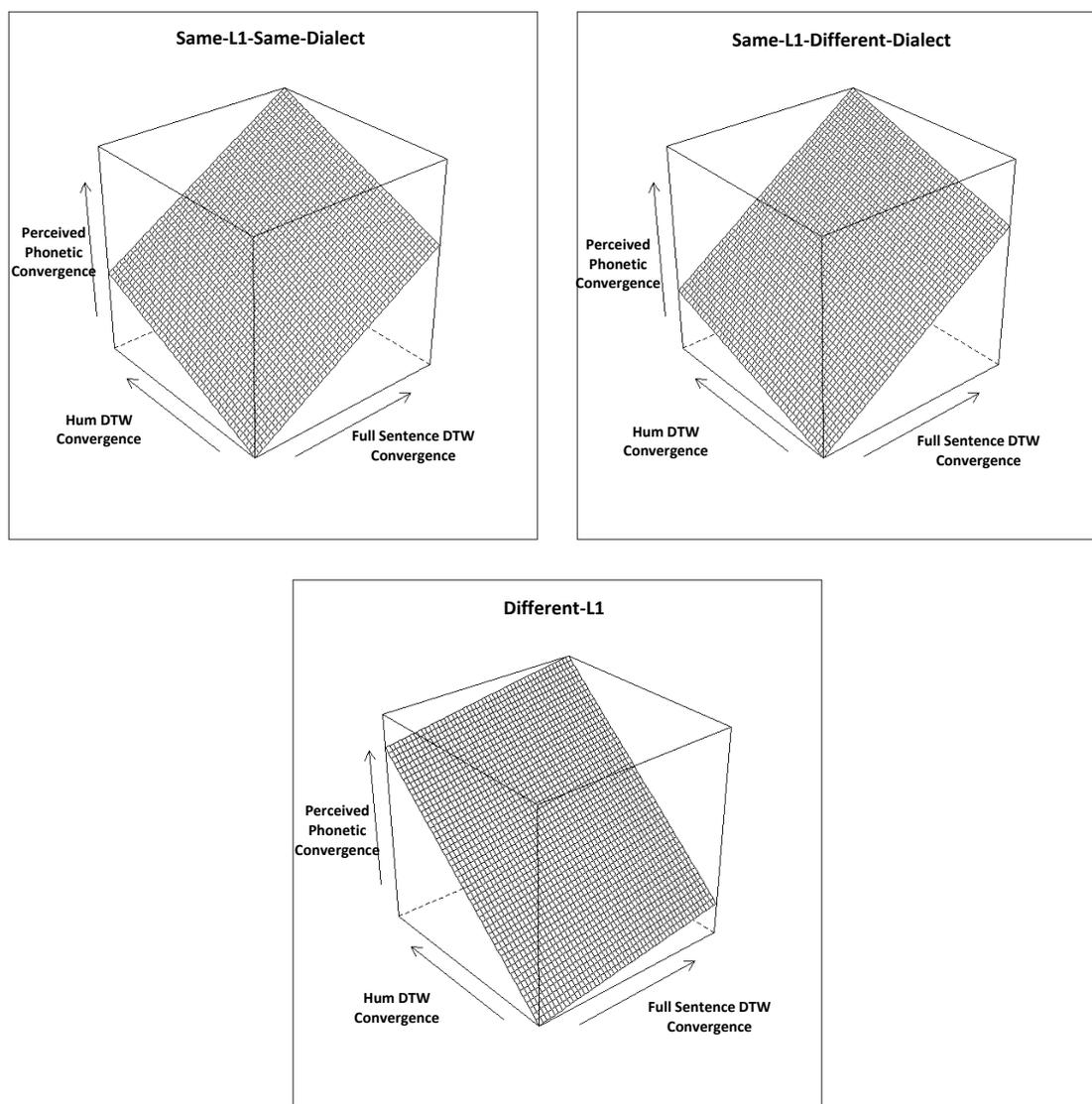


Figure 23. Schematic plots of perceived phonetic convergence predicted by hum DTW convergence and full sentence DTW convergence in the *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1* conditions

As seen in the first plot in Figure 23, in the *same-L1-same-dialect* condition, the results showed significantly higher likelihood of posttest sample selection in the XAB perception test when the degrees of DTW convergence on full sentences ($\hat{\beta} = 6.99, p < 0.0001$) and hums ($\hat{\beta} = 3.49, p < 0.05$) were higher. These two tendencies were not significantly different in the *same-L1-different-dialect* condition (see the second plot in Figure 22). However, in the *different-L1*

condition, the influence of full sentence DTW convergence on perceived phonetic convergence was significantly smaller ($\hat{\beta} = -4.29, p < 0.05$), while the influence of hum DTW convergence was significantly larger ($\hat{\beta} = 5.89, p < 0.05$), than in the *same-L1-same-dialect* condition (see the third plot in Figure 22). Importantly, these significant interactions did not change the positive contribution of the DTW convergence patterns to perceived convergence patterns.

These results suggest that phonetic accommodation judgments through both of the full and hum DTW analyses on sentences showed significantly positive relation to the human perceptual judgment on phonetic accommodation of sentences in all of the group level linguistic talker distances. In other words, although the accommodation patterns differ between the perceived accommodation judgment and the DTW accommodation judgments at the group level linguistic distances, in each of the group level linguistic distances, the DTW results were positively proportional to the perceived accommodation results.

Overall, the results on phonetic accommodation on sentences show the following tendencies:

1. Perceptual assessment results show that participants converged to all of the model talkers, regardless of dialect mismatch and L1 mismatch.
2. The DTW analyses suggest participants converged to a nonnative model talker for both of full sentences and hums, diverged from a native model talker with the same dialect background, and maintained their speech after hearing a native model talker with a different dialectal background.
3. The DTW results positively contributed to prediction of the human perceptual assessment of phonetic accommodation.

6. GENERAL DISCUSSION

6.1. Summary and discussion

The current study investigated native English talkers' phonetic accommodation towards native or nonnative model talkers in an auditory exposure situation with various measurements.

Specifically, we asked whether phonetic change could occur after auditory exposure to model talkers with different linguistic distances at two different levels: group level and item level. At the group level, we varied the linguistic distance between participants and their model talkers with dialect match/mismatch and L1 match/mismatch. We also asked whether a psychological factor, participants' implicit attitudes towards foreigners, would constrain phonetic accommodation patterns in the L1 mismatch condition. Moreover, we asked whether these phonetic accommodation patterns could be generalized from exposed items to unexposed items.

In a phonetic accommodation experiment, following the procedure of Goldinger and Azuma (2004), native English talkers read linguistic items before and after an auditory exposure phase. An important innovation of the current study regarding the materials is that we included items at three different linguistic levels, namely, monosyllabic words, disyllabic words, and sentences. This provided us with ample opportunities to look at convergence patterns as reflected by various acoustic measurements at different linguistic levels. During the exposure phase, participants in experimental groups heard the stimuli and did an item-identifying task with 9 repetitions per item. Another important difference of this study from previous studies on phonetic accommodation (e.g. Babel, 2009, 2010; Pardo et al., 2012; Pardo et al., 2010) is that we added a control group where participants were exposed to the same linguistic materials for the same item-identifying task with the same number of repetitions, but visually, not auditorily. By comparing results from the control group with results from the experimental groups, we

could observe the effect of auditory exposure excluding any item repetition effect. Moreover, participants heard or viewed only a half of the materials during the auditory or visual exposure phase, while they read all items in the pretest and posttest production phases. Comparison of productions of heard items and unheard items during the exposure phase enabled us to see whether the exposure effect generalized from old items to new items. Additionally, all participants performed an implicit association task which indicated their implicit attitudes towards foreigners. Finally, participants were asked about their impressions of their model talker's native status and dialects or origins.

Recordings of monosyllabic words were analyzed for VOT, vowel duration, F0-max, F1, and F2, and recordings of disyllabic words were analyzed for the ratios of vowel duration, F0-max, and amplitude between the first syllable and the second syllable. Sentence recordings were analyzed with DTW on the full sentences and on hums that reflected the intonation contours of the sentences. Finally, part of the sentence dataset was judged by human listeners in an XAB perception test of the phonetic accommodation patterns.

Table 14. Overall phonetic accommodation patterns

Linguistic level	Measurement	Group level linguistic distance		
		Same-dialect	Same-L1	Different-L1
Monosyllabic words	VOT	C	M	C
	Vowel Duration	C	M	C
	F0-max	C	C	C
	F1	C	C	C
	F2	C	C	C
Disyllabic words	V2/V1 Duration Ratio	C	C	C
	V2/V1 F0 Ratio	C	C	C
	V2/V1 Amplitude Ratio	C	C	C
Sentences	DTW full sentence	D	M	C
	DTW hum	D	M	C
	XAB perception	C	C	C

Note. C = convergence, M = maintenance, D = divergence.

Table 14 summarizes the overall phonetic accommodation patterns in the total dataset with all of the analyses. Overall, participants converged to their model talkers regardless of their group level linguistic distances for most of the measurements, with the exceptions of convergence inhibited by dialect mismatch for VOT and vowel duration and by L1 match for full sentence DTW and hum DTW. Moreover, these convergence patterns generalized to new items in most of the cases. Finally, although the two holistic measurements on phonetic accommodation, namely, the DTW technique and XAB perception test, indicated different accommodation patterns across the three group level linguistic distances as seen in Table 10, the linear mixed effects regression analysis revealed that the DTW results positively contributed in predicting the perceived phonetic convergence patterns in each of the group level linguistic distances.

A general tendency from the overall results is that, in the passive auditory exposure setting without any social or experimental forcing factors working, participants converged towards their native or nonnative model talkers along a variety of acoustic parameters and in terms of perceived similarity. Importantly, L1 mismatch and dialect mismatch did not interfere with perceived phonetic convergence. Moreover, participants' negative attitudes towards foreigners did not affect the degree of phonetic convergence towards a nonnative model talker.

These results are in the opposite direction from some of the previous studies on phonetic accommodation. As for the linguistic distance factor, while Kim et al. (2011) found that native talkers diverged from a native partner with a different dialectal background and barely converged to a nonnative partner in a conversation, the current study found that native talkers converged to both a native model with a different dialect and a high proficiency nonnative model talker.

Moreover, adjusted phonetic changes in the datasets of monosyllabic and disyllabic words were

significantly influenced by their preexisting acoustic distances from the model talkers. For all acoustic measurements on monosyllabic and disyllabic words, the general tendency was that the preexisting acoustic distance between items spoken by the participants and the items spoken by their model talkers positively predicted the degree of phonetic convergence by the participants in the posttest in any directions of change. That is, the more acoustically distant an item was from the model talker item in any directions, the larger the degree of phonetic convergence by the participant was.

This seems quite surprising and suggests an opposite tendency from the finding of Kim et al. (2011) that interlocutor sociolinguistic distances, namely, native status match and dialect match between interlocutors, are negatively correlated with the degree of their phonetic convergence. We can think of three important points regarding this contradiction between two studies. First, the two experiments were based on different tasks, a task-oriented conversation between interlocutors for Kim et al. (2011) and passive auditory exposure to a model talker for the current study. Therefore, the sociolinguistic distances that were effective for the interactive task in Kim et al. (2011) might not have been at work in the current study.

Second, the group level sociolinguistic distances between model talkers and their interlocutors in the current study are much closer than those in Kim et al. (2011), especially for the participants of the nonnative model talkers. This is because the nonnative model talkers in the current study were selected based on their high English proficiency, while the nonnative interlocutors in Kim et al. (2011) varied in their English proficiencies. Under these conditions (asocial experiment setting and much closer group level talker distances than in Kim et al. (2011)), participants might have been able to converge towards their model talkers regardless of L1 mismatch and dialect mismatch.

Third, within these group level linguistic distances that are close enough for phonetic convergence to occur, a certain phonetic distance between participants and their model talkers might have been needed at the item level for the participants to show any significant phonetic changes. This might have led to the tendency that the closer the acoustic talker distance was, the more room the participant had to move in the phonetic space, and the larger the degree of convergence was.

Regarding the implicit attitudes factor, while Babel (2009) and Babel (2010) found that participants with positive attitudes towards a black model talker exhibited a larger degree of phonetic convergence than others, we found that participants' attitudes towards foreigners did not directly impact their degree of convergence towards a nonnative model talker. This might be because there was no "social" setting for the phonetic accommodation experiment in the current study, while there were social settings where photos of the model talker's face were visually displayed with their voices in Babel (2009) and Babel (2010). Rather, participants' implicit attitudes towards foreigners influenced the relationship between preexisting acoustic distances between participants and their model talkers and participants' degree of phonetic convergence. The pattern of the influence varied depending on the acoustic measurement types.

The additional points we found from the results on the sentence dataset was the relation between two holistic judgments, one with human perception and the other with the DTW analyses. The phonetic convergence patterns of the sentences perceived by the XAB listeners did not completely match with those from the DTW analyses over the group level linguistic distances, *same-L1-same-dialect*, *same-L1-different-dialect*, and *different-L1*. This result is in line with some of previous findings. Pardo et al. (2010) found that phonetic accommodation patterns with articulation rate and vowel spectra did not converge with perceived accommodation

patterns. Babel and Bulatov (2011) found that F0 imitation patterns judged by acoustic analyses and human perception were not significantly correlated. Therefore, it seems that, whether single or multiple, segmental or holistic, acoustic judgments do not seem to converge with human perception on phonetic accommodation at the experimental setting level.

However, a linear regression between perceptual judgment and the DTW analyses on the sentence dataset in the current study revealed that the DTW on full sentences and hums were actually in positive relation with the perceptual judgment in each of the group level linguistic distances. Thus we can assume that, although the perceived accommodation patterns and the DTW accommodation patterns suggest different accommodation patterns across the group level linguistic distances, the perceived accommodation patterns could be partially predicted with the DTW distance changes in all the group level linguistic distances. Pardo (2010) also found similar results. In her post-analysis on the results of perceived phonetic accommodation patterns through task-oriented conversations between same-gender native English talkers in Pardo (2006), the variability of perceived accommodation was significantly accounted for by the variability of average F0 and duration of utterances.

The results from the adjusted phonetic changes revealed that participants generalized their phonetic accommodation patterns from exposed to unexposed items. This was noticeable in the present study for the acoustic measurements of the monosyllabic words and disyllabic words. Previously, Nielsen (2011) found that participants generalized their convergence to an extended VOT of words that started with /p/ to VOT of words that started with /k/. In the current study, we found evidence for accommodation generalization from more diverse measures and levels. Therefore, we can assume that phonetic accommodation might not stop as a transient change for

the moment, but can be applied to new items in the talkers' linguistic system and lead to long-term language learning.

In conclusion, in an auditory exposure experiment, we found robust evidence for phonetic convergence by native English talkers towards both native and nonnative model talkers based on acoustic analyses of English monosyllabic and disyllabic words, and human perception of sentences. Although DTW analyses of sentences revealed different patterns from human perception of phonetic convergence across different group level linguistic distances, the DTW results were positively related to perceived phonetic convergence on an item-by-item level. While phonetic convergence generally was not inhibited by dialect mismatch, L1 mismatch, and participants' negative attitudes, the absolute model talker-participant acoustic distance for each item before exposure was positively correlated with the degree of their phonetic convergence: the more acoustically distant a linguistic item was from the model talker's, the larger the degree of phonetic convergence was for that item. Lastly, the observed phonetic convergence was robust in that it generalized from exposed items to novel (unexposed) items at all levels of acoustic measurements.

6.2. Implications, limitations, and future studies

As introduced in Chapter 3, the dependent measure for phonetic convergence in the current study, adjusted phonetic changes, covers the case of overshooting convergence, not only phonetic convergence with shortened absolute Euclidean distance from the model talker value, which was the dependent measure for some previous studies (Babel, 2009, 2010; Pardo et al., 2012; Pardo et al., 2010). Therefore, this new dependent measure might actually have increased the possibility of observing phonetic convergence, whereas the previous studies failed to find reliable phonetic

convergence patterns on all acoustic measurements they attempted. This turned out to be true, as we found that phonetic convergence is a robust phenomenon in terms of many acoustic correlates of monosyllabic and disyllabic words and DTW features of sentences and their intonation patterns. As there is no crucial reason to exclude overshooting phonetic convergence from phonetic convergence accounts, adjusted phonetic changes might be conceptually better and be more productive than changes in the absolute Euclidean distance to model talkers.

Moreover, unlike in the previous studies (e.g. Babel, 2009, 2010; Goldinger & Azuma, 2004; Pardo et al., 2012; Pardo et al., 2010), we introduced the control condition where participants were exposed to words and sentences visually, but not auditorily. We subtracted the averaged posttest-pretest changes that occurred in this condition from the posttest-pretest changes in the experimental conditions for the calculation of the final phonetic convergence dependent measure, adjusted phonetic change. This process enabled us to exclude the potential linguistic repetition effect from the phonetic accommodation effect. As it is not known how much repetition effects contribute to phonetic accommodation in experiment settings, this experiment and analysis design for phonetic accommodation research allows us to take a closer look at phonetic accommodation effects without the interference of repetition effects.

In sum, we made two innovations for the new dependent measure of phonetic accommodation, adjusted phonetic change. On one hand, this new measure enlarged the opportunity to observe phonetic convergence, compared to the absolute Euclidian-distance measure used in previous studies. This was done by including overshooting convergence cases, not only the decrease of absolute distance to model values, into the area of phonetic convergence. On the other hand, we excluded potential repetition effects from the final phonetic accommodation measure by introducing the control condition. This enabled a more conservative

and rigorous investigation of phonetic accommodation. Taken together, the adjusted phonetic change formula can provide a better tool than previous measures in investigating phonetic accommodation along various acoustic-phonetic dimensions.

Although acoustic correlates might positively contribute to human perception of phonetic convergence, it is not that they all work in the same way for the same comparison. For example, in our data, the DTW analyses showed different accommodation results from the perceived phonetic convergence patterns across the three group level linguistic distances, although they were positively proportional to perceived phonetic convergence. Therefore, a very important message of the current study is that phonetic accommodation analyses that depend on a single or a few acoustic cues might draw very different results from analyses that used different single or multiple acoustic cues. If we do not consider the reason of these different results, it might be very hard to draw conclusions regarding phonetic accommodation. That is, it may appear that sometimes people converge to each other, other times they do not. In fact, while talkers do change their speech styles with various phonetic and acoustic features, the way individual talkers utilize individual phonetic and acoustic features might vary. Thus it might be hard to capture these exposure-induced modifications with a certain set of single cues, or even with a mechanically holistic judgment, the DTW technique. Nevertheless, ordinary listeners can judge the talkers' accommodation patterns by integrating all available phonetic and acoustic features. Therefore, while every study cannot and does not have to perform many acoustic measurements at all linguistic levels or conduct human perceptual judgment tests, it is clear that phonetic accommodation is a multifaceted phenomenon and we need to interpret results of individual studies carefully.

Theoretically, the current research sheds light on the speech perception-speech production link. Under a passive auditory exposure setting and with no instruction for imitation, participants in the current study changed their speech production towards the model talkers only by listening to their read speech. This suggests that speech perception can lead to speech production change quite automatically and without any socially forcing factors. Critically, this process was not constrained by group level linguistic and psychological factors, such as dialect mismatch, L1 mismatch, and participants' negative attitudes towards foreigners. However, the degree of phonetic convergence was constrained by subtle acoustic distances between model talkers and participants on individual words. Therefore, we observed that phonetic convergence can occur in a fast and asocial manner, but still conditioned by fine-grained preexisting linguistic talker distances.

This supports the idea of conditional automaticity on the perception-behavior link (Bargh, 1989; Pickering & Garrod, 2004). That is, while automaticity can be traditionally defined as a process that occurs without awareness, is effortless, unintentional, autonomous, and uncontrollable, these features might not necessarily occur at the same time. Rather, any "automatic" processing might come with a certain set of such features, thus conditioned by the other missing features. Specifically, the automaticity of convergence is assumed to occur at the post-conscious level, where the stimulus has to be under the talker's awareness. In other words, the talker has to listen to the stimulus. It is considered that, through this awareness, many linguistic, psychological, and social factors influence the relatively automatic phonetic accommodation process.

Another theoretical implication of the current study is that this line of research can give us a hint as to a mechanism of language change, which may be caused in part by the spread of

individual production changes that follow perception of different patterns of speech input. The experiment was a controlled, lab-based investigation of this sort of real-world language change. In particular, the current study can provide some insight into language change induced by contact with different dialects and L2 talkers. The results of the current study suggest that interlocutor dialectal differences or the native status of interlocutors might not necessarily affect the degree of phonetic convergence by the interlocutors. That is, one can imitate talkers from different dialectal backgrounds, and can even imitate nonnative talkers quite automatically. The more important constraint for the degree of phonetic convergence was suggested to be the interlocutors' idiosyncratic production distances from each other.

Practically, understanding the speech perception-production link has implications for communication in general. It has been suggested that phonetic convergence can yield better flow for communication (e.g. Pickering & Garrod, 2004; Shepard et al., 2001). Especially for global communication, the results of the current study suggest that nonnative interlocutors would not necessarily be hard to converge to. That is, native English talkers phonetically converged towards nonnative model talkers in terms of all acoustic measurements for monosyllabic and disyllabic words and DTW features on full sentences and hums, and they were perceived to have imitated nonnative model talkers' sentences. Thus, in a future study, we can ask whether native talkers' phonetic imitations towards nonnative talkers could lead to better communicative efficiency between the interlocutors in global communication.

Another practical implication of this work is for research on human-computer conversational systems. Recently, researchers on human-computer conversational systems have been studying acoustic correlates that might be involved in rapport in human-human dialogues, so that they can apply the findings to the design of better human-computer interfaces where

people can talk with a computer system in a more natural and cooperative manner (Worgan & Moore, 2011). The acoustic correlates of phonetic accommodation found in the current study are exactly what this line of research would need. They can apply the acoustic correlates related to phonetic accommodation to their human-computer dialogue system, so that the computer system can accommodate to human speech in an appropriate manner. Moreover, the diverse patterns of the acoustic correlates for phonetic accommodation to different model talkers in different conditions might have implications in designing a more personalized human-computer conversational system.

While the current study focused on finding general linguistic features of phonetic accommodation, we can also investigate language specific features that can be used for phonetic accommodation. In fact, all monosyllabic and disyllabic words picked and the new sentences made for the current study originally aimed at studying phonetic accommodation on systematic differences between English and Korean. That is, certain English linguistic features that do not match with the Korean features might be contributing to our nonnative model talkers' nonnative accent, and this might be transferred to participants who heard the nonnative model talkers in a systematic way. For example, while the English two-way stop distinction is known to use VOT mainly, the Korean three-way stop distinction is known to utilize both VOT and F0 of the vowel (e.g. Kim, 2004). Importantly, in the Korean system, the stop voicing category, "aspirated", is related to heavy aspiration and high F0, while the category, "lenis", is related to light aspiration and low F0. This L1 feature might exist in our Korean nonnative model talkers' English accent, as an additional F0 distinction on the English voiced (no or light aspiration, so low F0) vs. voiceless (heavy aspiration, so high F0) voicing distinction. While the current study lacks this line of investigation, in a future study, we can ask whether this kind of L2-induced systematic

linguistic differences can be accommodated by native participants, resulting in their adopting the nonnative way of voicing distinction with F0. Furthermore, we can investigate whether the nonnative model talkers had any systematic differences from the native model talkers in terms of vowel formants of monosyllabic words, word level stress patterns in disyllabic words, and intonation patterns in the new sentences, and whether these linguistically systematic differences were also targets of phonetic accommodation.

In the current study, we found robust evidence of phonetic accommodation in a very passive, asocial setting. Moreover, we tested dialect mismatch, L1 mismatch, and implicit attitudes towards foreigners as potential intervening factors on phonetic accommodation, and found that they influence the phonetic accommodation process in a very subtle way, not directly the degree of convergence, but various interactions with preexisting acoustic distances. Therefore, we can consider the results of the current study as a baseline for the phonetic accommodation research. That is, phonetic convergence occurs in a passive overhearing setting with no interaction with the talker, no instruction to imitate, no production practice, and no training or feedback, and it is still conditioned by linguistic and psychological factors in various ways. And now we can move on to exploring more ecologically valid intervening conditions for phonetic accommodation and filling the gaps among previous research results. For example, while nonnative talkers are known to be hard to converge to by native talkers (Kim et al., 2011), the current study found that high proficiency enables native talkers to converge to some nonnative talkers. We also know that, for women pairs, being a receiver in a task with between-talker information asymmetry facilitates the talker's phonetic convergence towards the partner (Pardo, 2006). Then, imagine a conversation between a female native talker and a female nonnative talker with low target language proficiency. Would a receiver role for the native talker

help her phonetic convergence to her low proficiency nonnative partner? In other words, would the female receiver facilitation effect override the inhibitory effect of nonnative low proficiency on phonetic convergence? These questions will help us finding various intervening factors for the speech perception-production link in the linguistic, psychological, and social layers.

REFERENCES

- Abercrombie, D. (1967). *Elements of general phonetics* (Vol. 203). Edinburgh: Edinburgh University Press.
- Alshangiti, W., & Evans, B. G. (2011). Regional accent accommodation in spontaneous speech: Evidence for long-term accent change? *Proceedings of 17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong, 17-21 August, 224-227.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*: Cambridge University Press.
- Baayen, R. H. (2011). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". R package version 1.4. <http://CRAN.R-project.org/package=languageR>
- Babel, M. E. (2009). *Phonetic and social selectivity in speech accommodation*. PhD dissertation, University of California, Berkeley.
- Babel, M. E. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, 39(04), 437-456.
- Babel, M. E. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177-189. doi: 10.1016/j.wocn.2011.09.001
- Babel, M. E., & Bulatov, D. (2011). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, 1-36.
- Baker, R. E., & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*, 52(4), 391-413.

- Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3-51): Guilford Press.
- Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42. <http://CRAN.R-project.org/package=lme4>
- Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht: Foris.
- Berndt, D., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *Proceedings*.
- Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.1.29. <http://www.praat.org/>
- Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, ethnicity and intergroup relations* (Vol. 13, pp. 119-135). London: Academic Press.
- Bradlow, A., Pisoni, D., AkahansYamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English/r/and IV: IV. Some effects of perceptual learning on speech production. *Jourml of the Acoustical Society of America*, 101(4), 2299-2223.
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977-990.
- Burt, S. M. (1998). Monolingual children in a bilingual situation: Protest, accommodation, and linguistic creativity. *Multilingua*, 17(4), 361-378.
- Chan, C. L. (2009). TriggerWave [Computer program]. Linguistics, Northwestern University. http://groups.linguistics.northwestern.edu/documentation/triggerwave_home.html

- Chan, C. L. (2010). NUALigner 2 [Computer program]. Linguistics, Northwestern University.
http://groups.linguistics.northwestern.edu/documentation/nualigner_home.html
- Cho, M.-H. (2004). Rhythm typology of Korean speech. *Cognitive Processing*, 5(4), 249-253.
doi: 10.1007/s10339-004-0023-1
- Clark, H. H. (1996). *Using language* (Vol. 4): Cambridge University Press Cambridge.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech* (Vol. 3): Harvard Univ Press.
- Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue.
Language and Cognitive Processes, 23(4), 528-556.
- Coupland, N. (1984). Accommodation at work: Some phonological data and their implications.
International Journal of the Sociology of Language, 1984(46), 49-70.
- Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, 64(2-3), 145-173.
- Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in experimental social psychology*, 33, 1-40.
- Ellis, D. P. W. (2003). Dynamic Time Warp (DTW) in Matlab.
<http://www.ee.columbia.edu/~dpwe/resources/matlab/dtw/>
- Ellis, D. P. W. (2005). PLP and RASTA (and MFCC, and inversion) in Matlab.
<http://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/>
- Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, 49(3), 396-413. doi: 10.1016/s0749-596x(03)00072-x

- Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "New" and "Old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26(5), 489-504.
- Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27, 765.
- Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1(2), 126-152.
- Gallois, C., Giles, H., Jones, E., Cargile, A. C., & Ota, H. (1995). Accommodating intercultural encounters: Elaborations and extensions. In R. Wiseman (Ed.), *Intercultural Communication Theory* (pp. 115-147). Thousand Oaks, CA: Sage.
- Garrod, S., & Pickering, M. J. (2007a). Alignment in dialogue. In M. G. Gaskell & G. Altmann (Eds.), *The Oxford handbook of psycholinguistics* (pp. 443-451). New York: Oxford University Press.
- Garrod, S., & Pickering, M. J. (2007b). Automaticity in language production in monologue and dialogue. In A. S. Meyer, L. R. Wheeldon & A. Krott (Eds.), *Automaticity and control in language processing* (pp. 1-21). New York: Psychology Press.
- Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2), 292-304.
- Giles, H. (1973). Accent mobility: A model and some data. *Anthropological linguistics*, 15(2), 87-105.
- Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation: Developments in applied sociolinguistics*: Cambridge Univ Pr.

- Goldinger, S. D. (1997). Words and voices: Perception and production in an episodic lexicon. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 33-66). New York: Academic Press.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review*, *105*(2), 251-279.
- Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic bulletin & review*, *11*(4), 716-722.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of personality and social psychology*, *85*(2), 197-216.
- Gregory, S. W., Dagan, K., & Webster, S. (1997). Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior*, *21*(1), 23-43.
- Gregory, S. W., Green, B. E., Carrothers, R. M., Dagan, K. A., & Webster, S. (2001). Verifying the primacy of voice fundamental frequency in social status accommodation. *Language and Communication*, *21*, 37-60.
- Gregory, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology*, *70*(6), 1231.

- Gregory, S. W., Webster, S., & Huang, G. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language & Communication; Language & Communication*.
- Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America*, 130, 2139.
- Honorof, D. N., Weihing, J., & Fowler, C. A. (2011). Articulatory events are imitated under rapid shadowing. *Journal of Phonetics*, 39(1), 18-38. doi: 10.1016/j.wocn.2010.10.007
- Inquisit. (2009) [Computer program]. Version 3.0.4.0 Seattle, WA: Millisecond Software.
- Jungers, M. K., & Hupp, J. M. (2009). Speech priming: Evidence for rate persistence in unscripted speech. *Language and Cognitive Processes*, 24(4), 611-624.
- Jungers, M. K., Palmer, C., & Speer, S. R. (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing*, 1(2), 21-35.
- Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *61*(5), 1337-1351.
- Kim, M. (2004). Correlation between VOT and F0 in the perception of Korean stops and affricates. *Proceedings of The 8th International Conference on Spoken Language Processing*, 49-52.
- Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, 2(1), 125-156. doi: 10.1515/labphon.2011.004

- Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*(1), 54-81.
- Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English : phonetics, phonology, and sound change : a multimedia reference tool*. Berlin ; New York: Mouton de Gruyter,.
- Lee, J. P., & Jang, T. Y. (2004). A comparative study on the production of inter-stress intervals of English speech by English native speakers and Korean speakers. *Proceedings of International Conference on Spoken Language Processing*.
- Lee, O., & Kim, J. (2005). Syllable-timing interferes with Korean learners' speech of stress-timed English. *Speech Sciences*, *12*(4), 95-112.
- Lewandowski, N. (2011). *Talent in nonnative phonetic convergence*. PhD Dissertation, Universit at Stuttgart.
- Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of Acoustical Society of America*, *32*(4), 451-454. doi: <http://dx.doi.org/10.1121/1.1908095>
- MATLAB. (2010) [Computer program]. Version 7.11.0.584 Natick, Massachusetts: The MathWorks Inc.
- Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation. *Journal of Language and Social Psychology*, *21*(4), 422-432.
- Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, *32*(5), 790.
- Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*, 132-142.

- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382-2393.
- Pardo, J. S. (2010). Expressing Oneself in Conversational Interaction. In E. Morsella (Ed.), *Expressing Oneself/Expressing One's self: Communication, cognition, language, and identity* (pp. 183-196). London: Psychology Press.
- Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*(1), 190-197. doi: 10.1016/j.wocn.2011.10.001
- Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, perception & psychophysics*, *72*(8), 2254-2264. doi: 10.3758/APP.72.8.2254
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02), 169-190.
- Pike, K. L. (1945). *The Intonation of American English*. Ann Arbor, MI: University of Michigan Press.
- Ramus, F., Nespors, M., & Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. *Cognition*, *75*(1), AD3-AD30. doi: 10.1016/s0010-0277(00)00101-3
- Shepard, C. A., Giles, H., & Le Poire, B. A. (2001). Communication accommodation theory. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 33-56). New York: Wiley.
- Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Attention, Perception, & Psychophysics*, *66*(3), 422-429.
- Tilsen, S. (2009). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics*, *37*(3), 276-296.

- Turetsky, R. J., & Ellis, D. (2003). Ground-truth transcriptions of real music from force-aligned midi syntheses. *Proceedings*.
- Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35(4), 445-472. doi: 10.1016/j.wocn.2006.12.001
- Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, 53(4), 510-540.
- The Versant™ English Test: Test Description & Validation Summary*. (2008). Palo Alto, CA: Pearson Education.
- Worgan, S. F., & Moore, R. K. (2011). Towards the detection of social dominance in dialogue. *Speech Communication*, 53, 1104-1114.
- Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08*.
- Zuengler, J. (1982). Applying accommodation theory to variable performance data in L2. *Studies in Second Language Acquisition*, 4(02), 181-192.