Phonetic convergence in spontaneous conversations as a function of interlocutor language distance

MIDAM KIM*, WILLIAM S. HORTON**, and ANN R. BRADLOW*

*Department of Linguistics, Northwestern University
**Department of Psychology, Northwestern University

Abstract

This study explores phonetic convergence during conversations between pairs of talkers with varying language distance. Specifically, we examined conversations within two native English talkers and within two native Korean talkers who had either the same or different regional dialects, and between native and nonnative talkers of English. To measure phonetic convergence, an independent group of listeners judged the similarity of utterance samples from each talker through an XAB perception test, in which X was a sample of one talker's speech and A and B were samples from the other talker at either early or late portions of the conversation. The results showed greater convergence for same-dialect pairs than for either the different-dialect pairs or the different-L1 pairs. These results generally support the hypothesis that there is a relationship between phonetic convergence and interlocutor language distance. We interpret this pattern as suggesting that phonetic convergence between talker pairs that vary in the degree of their initial language alignment may be dynamically mediated by two parallel mechanisms: the need for intelligibility and the extra demands of nonnative speech production and perception.

1. Introduction

Because conversational interactions are joint activities (Clark 1996: 3), individual interlocutors often adjust their speech and language production and perception patterns depending on the particular talker-listener combination. This phenomenon of interlocutor adjustment has been studied at various levels of linguistic structure, and under various names including "coordination" (Clark 1996), "accommodation" (e.g., Giles, Coupland, and Coupland 1991; Shepard, Giles, and Le Poire 2001; Namy, Nygaard, and Sauerteig 2002; Babel 2009, 2010), "alignment" (e.g., Pickering and Garrod 2004, 2006; Kraljic, Brennan, and Samuel 2008), "audience design" (e.g., Clark and Murphy 1982), and "convergence" (e.g., Pardo 2006). In the present study, we examine adjustment at the phonetic level within spontaneous

JLP 2 (2011), 125–156
DOI 10.1515/LABPHON.2011.004
1836-6346/11/01–125

conversations between interlocutors from different native language or dialect backgrounds, with the broad goal of demonstrating bi-directional talker-listener adjustment as a possible mechanism of long-term, contact-induced language change.

A large body of work has documented talker-to-listener speech production adjustment under various conditions. For example, infant-directed speech (e.g., Cooper and Aslin 1990; Wassink, Wright, and Franklin 2007), clear speech for the benefit of listeners under adverse conditions due to a hearing loss or the presence of environmental noise (e.g., Picheny, Durlach, and Braida 1985; Payton, Uchanski, and Braida 1994; Uchanski 2005; Smiljanic and Bradlow 2009), and foreigner-directed speech (e.g., Ferguson 1971; Smith 2007; Uther, Knoll, and Burnham 2007) all represent cases of talker-initiated speech production adjustments to the communicative setting. These cases are well understood within a framework such as the Hypo- and Hyper-Theory of Speech Production (H&H Theory; Lindblom 1990), which provides for a natural account of talker variability in response to the instance-specific, competing constraints of perceptual salience and economy of effort. Similarly, listener-to-talker speech perception adjustment has been well-documented and accounted for within theories of speech perception that allow for highly flexible speech processing mechanisms (e.g., for a review see Samuel and Kraljic 2009). This body of work has demonstrated listener adaptation to individual talker characteristics (e.g., Nygaard, Sommers, and Pisoni 1994; Nygaard and Pisoni 1998; Norris, McQueen, and Cutler 2003; Eisner and McQueen 2005; Kraljic and Samuel 2005) as well as to systematic variation across groups of talkers (e.g., Bradlow and Bent 2008; Sidaras, Alexander, and Nygaard 2009).

Taken together, these separate lines of research on talker-to-listener and listener-to-talker adjustment have provided evidence for bi-directional phonetic adjustment. However, these studies have typically involved de-contextualized communicative situations, and thus provide an incomplete view of the interaction between talker-based production adjustments and listener-based perception adjustments that may operate in the more naturalistic situation of interactive dialogue. Similarly, studies of speech imitation (e.g., Goldinger 1998; Namy, Nygaard, and Sauerteig 2002; Goldinger and Azuma 2004; Shockley, Sabadini, and Fowler 2004; Delvaux and Soquet 2007; Nielson 2008; Babel 2009, 2010; Miller, Sanchez, and Rosenblum 2010) have provided evidence for perceptually-driven changes in speech production in the form of acoustic-phonetic adjustments by one talker following exposure to productions by another talker. But, like the separate studies of talker-to-listener and listener-to-talker adjustment, the speech imitation task typically involves a non-interactive, non-social setting with a tenuous connection to the real-world situation of spontaneous dialogues.

An important recent demonstration of phonetic convergence within conversational interactions was provided by Pardo (2006), who used a "Map task" (Anderson et al. 1991) to elicit conversations in which one talker takes on the role of direction "giver" while the other takes on the role of instruction "receiver." This study showed significant phonetic convergence within the conversations, with

some modulation of the convergence by talker gender and talker role. Specifically, males generally showed a larger degree of convergence to their partners than females, and within female pairs, only the giver converged towards the receiver. (See Babel, 2009, for an extensive and excellent review of work on phonetic convergence from psycholinguistic and social psychological perspectives over the past 4–5 decades.)

The investigation of talker-listener adjustment in the context of naturalistic conversations with multiple turns for each participant represents an important step toward understanding phonetic variation in real-world speech communication. It also provides a conceptual link to population-level, contact-induced change, such as might be in progress for English in its role as a global language. For example, where one speech sub-community may convey a certain vowel contrast with primarily spectral differences, another may instead realize the contrast with primarily temporal distinctions, or with spectral differences in a different range. These differences may stem from socio-phonetic factors relating to social and regional group membership, or from cross-language interaction within bilingual individuals. Provided that these group-based variations are systematic, speech communication across these sub-communities may then prompt individuals to adjust their perception and production categories to accommodate the various sound patterns. Repeated interactions may lead to long-term adjustments that will be evident in interactions with yet other sub-communities, thereby setting up the conditions for the propagation of a change through the broader speech community (Costa, Pickering, and Sorace 2008). Indeed, a substantial body of previous work on dialect change has demonstrated adjustment of specific acoustic phonetic parameters in response to a change in the ambient dialect such as occurs when an individual moves from one city to another within English-speaking Britain or USA (e.g., Munro, Derwing, and Flege 1999; Evans and Iverson 2007; amongst many others) or when the ambient language changes along with a move from one country to another (e.g., Sancier and Fowler 1997). With this general outlook in mind, the present study aimed to extend Pardo (2006) to the case of interactive dialogues between interlocutors who vary in the extent of their shared language experience.

In particular, we examined phonetic convergence under three conditions. In the same-L1/same-dialect condition, the interlocutors spoke the same dialect of either American English (2 pairs) or Korean (2 pairs). In the same-L1/different-dialect condition, the interlocutors spoke different dialects of either American English (2 pairs) or Korean (2 pairs). In the third condition, the interlocutors came from different native language backgrounds. In particular, one talker was a native talker of American English while the other was a native talker of either Korean (4 pairs) or Chinese (4 pairs); for these different-L1 conversations the language of the conversation was always English and the talkers differed in their status as a native or nonnative talker of the target language. Thus, in the present study, the "language distance" between the interlocutors varied from "close" (the same-L1/same-dialect condition), to "intermediate" (the same-L1/different-dialect condition), to "far"

(the different-L1 condition) with two possible factors determining language distance between interlocutors, namely L1 sharing and dialect sharing.

One possible outcome is that phonetic convergence within conversations would vary in relation to initial interlocutor language distance, such that greater convergence would be observed for pairs with relatively well-matched language backgrounds. This outcome would be consistent with the idea that phonetic convergence is limited to parameters and categories that are already well-established within the talkers' linguistic sound systems (Babel 2009). In a laboratory-based English speech shadowing task, Babel (2009) found that test talkers adjusted some vowel categories in the direction of the model talker's production of these categories, while leaving other vowel categories unchanged following exposure to a model talker's speech. Specifically, Babel (2009) found that the primary targets of phonetic convergence were the English low vowels /æ/ and /ɑ/, whereas the English high vowels /i/ and /u/ were left largely unchanged in the test talkers' productions. Babel's explanation for this vowel-specific convergence was that, within the English system, the low vowels /æ/ and /ɑ/ are typically subject to a higher degree of prosodically-determined variability than the high vowels /i/ and /u/ along exactly the phonetic dimension that participated in the observed phonetic convergence, namely F1/vowel height (which varies across prosodically accented and unaccented environments). Thus, in Babel's study, phonetic convergence seemed to operate within the existing phonetic repertoire of the test talkers. Under this view, interlocutors with relatively well-matched linguistic sound systems are more likely to exhibit phonetic convergence in spontaneous conversations than interlocutors with highly disparate production patterns because the variability to which the relatively well-matched talkers are exposed is more likely to be within their existing phonetic repertoires. That is, for talkers with a relatively large language distance between them, even though the distance between their vowels would seem to provide plenty of "room" for accommodation, their vowels in a given lexical item may be different enough to be outside of each other's typical vowel repertoire thereby effectively blocking phonetic convergence.

An alternative possibility is that we will observe greater phonetic convergence between interlocutors with greater language distance simply because there is more room for adjustment. This possibility would require that production targets are highly flexible and that the process of phonetic convergence is relatively unconstrained by the existing phonetic and phonological systems. This type of production flexibility could be consistent with the high degree of perceptual flexibility that has now been well established in studies of perceptual adaptation to speech variability, including adaptation to dialect variation and to foreign-accented speech. However, this kind of relatively unconstrained speech production adaptation could be difficult to reconcile with well-known limits on ultimate levels of second language speech production proficiency by late learners of a foreign language (Flege 1999; Birdsong 2004). We therefore aimed to test the hypothesis that phonetic convergence in spontaneous conversations is facilitated by relatively

well-matched language backgrounds (close "language distance") between the interlocutors.

It is important to note that both of the possible outcomes discussed above are based purely on phonetic/linguistic factors and ignore social factors having to do with the interlocutors' attitudes to talkers with language backgrounds that may differ from their own (Giles and Ogay 2007). This type of psycho-social influence has been addressed in previous work on phonetic convergence. For example, Babel (2009) found that the degree of phonetic convergence by a test talker to a model talker was influenced by the test talker's rating of the model talker's "attractiveness" and of the test talker's implicit attitude towards the model talker's race, which was assessed on the basis of an Implicit Association Task (Greenwald, McGhee, and Schwartz 1998). Babel (2009) observed greater phonetic convergence for test talkers with positive biases towards the model talker's race. Additionally, for female test talkers, high attractiveness ratings for the model talker correlated with greater phonetic convergence. This finding (amongst others that have focused on social factors in phonetic convergence and imitation such as Pardo, 2006, and Namy, Nygaard, and Sauerteig, 2002) is taken as evidence against a view of speech alignment as an automatic process (Pickering and Garrod 2004, 2006). Together with the finding of phonetic selectivity of phonetic convergence, this previous work strongly suggests that the processes of phonetic convergence are mediated by both social and linguistic biases (Babel 2009). In the present study, we focus on phonetic variables such as the talkers' native status and dialects, yet we acknowledge that psycho-social factors may also be operational.

2. Methods

2.1. *The diapix conversation elicitation task*

In general we adopted the methodology presented in Pardo (2006). However, we differed from Pardo (2006) with respect to the conversation elicitation technique. Pardo (2006) used the "Map task" (Anderson et al. 1991) in which two participants are each given a copy of a hand-drawn map with easily identified landmarks. In the Map task, the two participants cannot see each other's version of the map, and one talker, the "giver," has a route marked out on the map and must guide the other talker, the "receiver," through the set of landmarks to arrive at some destination via the same route. In the present study, we elicited conversations between the interlocutors using a new task, the "diapix" task (described further below and in detail in Van Engen et al. 2010), in which no giver or receiver role differences are imposed on the talkers. This task encourages balanced talker roles across the interlocutors as a means of encouraging bi-directional phonetic convergence.

In the diapix task, each talker is given one of two pictures, scenes A and B, which are identical except for 10 differences. The talkers are seated such that they cannot see each other's picture and are instructed to work together to find the 10

differences. The differences are created such that three involve elements that are present in scene A but absent from scene B, three involve elements that are present in scene B but absent from scene A, and four involve elements that differ across scenes A and B in terms of some detail such as color or shape. For the present study, we used diapix recordings collected as part of the Wildcat Corpus. For a complete description of all aspects of this corpus, see Van Engen et al. (2010). Diapix recordings in this corpus were made using two different picture pairs, one for each of two target languages, English (the "shop" scene) and Korean (the "beach" scene). In constructing these scenes, every effort was made to keep them similar in terms of overall "look and feel" (the same artist developed the two scenes) and in the level of detail required to identify the crucial differences. Because several of the differences involve writing in the target languages (English for the shop scene, Korean for the beach scene), it was necessary to have separate English and Korean scenes. See Appendix A for black-and-white renditions of the original color pictures and Appendix B for a list of the differences across each picture pair to be found by the interlocutors.

Wildcat Corpus diapix recordings involved two participants who were seated at desks facing opposite walls in a soundproof room. Each participant was given one of the two scenes within a diapix picture pair (the shop scene for English conversations, the beach scene for Korean conversations), which were printed on letter-sized sheets of paper. The participants were then asked to find 10 differences between the two pictures by talking out loud and working together as fast and efficiently as possible. The participants were not allowed to see each other or the other's picture. Each participant wore an AKG C420 headset microphone and their conversation was stereo recorded to separate channels for each participant using a Marantz PMD 670 flash recorder. Across the entire Wildcat Corpus, which includes diapix recordings from 38 pairs of talkers, almost all pairs managed to find all 10 differences (no pair missed more than 3 differences).

Participants in the diapix task of the Wildcat Corpus (both English and Korean) were recruited by word of mouth and through advertisements posted on the Northwestern University campus. Most participants were graduate or undergraduate students, or post-doctorate researchers at Northwestern University; a small number were partners of graduate students at Northwestern University. The native language backgrounds of the participants varied with the majority being English (n = 24), Chinese (n = 20), or Korean (n = 20). For a complete list of all participants' native language backgrounds, see Van Engen and colleagues (2010). Participant ages ranged from 18 to 34 years with the average of 25.8 years. In all language pairs, the female to male gender ratio was 1:1, that is, there were no mixed gender pairs. Each talker participated in only one conversation. All participants received payment for their participation. None of the participants reported a speech or hearing impairment at the time of testing.

For the present study, a total of 16 diapix conversations were selected from the complete set of 38 diapix conversations in the Wildcat Corpus. For the same-L1

conversations, four English conversations (out of a total of eight such conversations in the Wildcat Corpus) were selected, and four Korean conversations were selected (representing all such conversations in the Wildcat Corpus).Within each language group (English and Korean), two of the four conversations were between males and two were between females. Moreover, based on self-reported information about geographical regions where the talkers lived from their birth to 18 years within the USA or South Korea, two of the conversations within each language group (one between males and one between females) fell in the same-L1/same-dialect group while the other two fell into the same-L1/different-dialect group.

For the different-L1 conversations, we selected eight native+nonnative English conversations from the Wildcat Corpus. Four of these were between a native American English talker and a native Chinese talker, and four were between a native American English talker and a native Korean talker. Within each of these sets of four, two were between female talkers and two were between male talkers. As explained above, all of these different-L1 conversations were performed in English. These eight conversations represent the full set of native+nonnative diapix recordings in the Wildcat Corpus. See Table 1 for detailed characteristics of talkers in the native+native and native+nonnative conversations.

2.2. *Phonetic convergence assessment*

While we generally followed the methodology of Pardo (2006) quite closely, there were two important differences between our methods of phonetic convergence assessment and those of Pardo (2006). The participants in Pardo (2006) were recorded reading a set of target words (landmarks on the map) before and after performing the interactive map task. Then, following previous speech shadowing studies (e.g., Goldinger 1998; Namy, Nygaard, and Sauerteig 2002; Shockley, Sabadini, and Fowler 2004), Pardo (2006) assessed phonetic convergence by means of an AXB perceptual similarity test, in which an independent group of listeners were presented with three repetitions of the same target word. The first and the last repetitions (A and B) consisted of one talker's productions of the target words at pre-test and at post-test, and the middle repetition (X) consisted of the other talker's production of the same target word during the conversation. The listener's task was to decide which of A and B sounded more similar to X. Phonetic convergence was then quantified as the rate of post-test selection, which would indicate greater perceived phonetic similarity at the end of the conversation (at post-test) than before the conversation (at pre-test). In the present study, we used speech samples taken from early and late portions of the recorded conversations, rather than pre-and post-conversation recordings of read speech. Furthermore, since these recorded samples were slightly longer than single word recordings, we modified the AXB comparison task such that the model sample was presented first (i.e., XAB) as a means of easing the memory load on the participants in the perceptual judgment task.

Table 1. *Characteristics of talkers in native+native and native+nonnative conversations.*

Native+Native conversations					
Target language	Talker type		Dialect		
	Talker 1	Talker 2	Talker 1	Talker 2	Dialect comparison
English	ENF	ENF	Bloomington, MN	Glencoe, IL	Same[a]
	ENM	ENM	GA	GA	Same
	ENF	ENF	CA	NY/FL	Different
	ENM	ENM	AZ	PA	Different
Korean	KOF	KOF	Seoul	Seoul	Same
	KOM	KOM	Seoul	Seoul	Same
	KOF	KOF	Jeju	Gangwon	Different
	KOM	KOM	Seoul	Daegu	Different

Native+Nonnative conversations					
Target language	Talker type		L2 accentedness ratings		
	Native	Nonnative	Talker 1	Talker 2	Interpretation
English	ENF	CHF	NA[b]	−0.58	Light
	ENM	CHM		−0.53	
	ENM	KOM		0.29	
	ENF	KOF		0.31	
	ENM	KOM		0.34	
	ENF	CHF		0.35	
	ENF	KOF		0.39	
	ENM	CHM		0.5	Heavy

Note. ENF = female native English, ENM = male native English, KOF = female native Korean, KOM = male native Korean, MN = Minnesota, IL = Illinois, GA = Georgia, CA = California, NY = New York, FL = Florida, AZ = Arizona, PA = Pennsylvania.

a. Bloomington, MN and Glencoe, IL are classified as regions with the same dialect according to Labov, Ash, and Boberg (2006: 6).

b. All English talkers in native+nonnative conversations were rated as having very low accentedness (M = −1.79).

To obtain early and late samples from each talker's diapix recording for presentation in the XAB perceptual judgment task, three utterance snippets were extracted from the first third and the last third of each talker's portion of the conversation, using Adobe Audition 1.5. Speech samples were chosen according to the following three criteria: (1) samples were 1 to 1.5 seconds in duration, (2) samples consisted of one intonational phrase or occurred at the end of an intonational phrase, and (3) samples were fluently produced without any evident hesitations, disfluencies, background noise, or back channeling from the other talker. From each talker's recording, the first (within the first third of the entire conversation duration) and last (within the last third of the entire conversation duration) three

speech samples that fit these criteria were selected. Because it is not possible to obtain identical utterances from the early and late parts of a diapix conversation (people usually did not go back to an item they had already discussed), the extracted speech samples were all different in context (i.e., contained different words).[1] In total, 192 speech samples (3 samples × 2 time points × 2 talkers × 16 conversations) were extracted for presentation to an independent group of participants in XAB perceptual similarity tests. All speech samples were normalized to have the same overall RMS value (1.0 Pa). Transcripts of the complete set of the extracted utterances are provided in Appendix C.

A group of 121 native English talkers participated in the XAB perceptual similarity test on English speech samples. All of these participants were undergraduate students at Northwestern University. They ranged in age from 19 to 26 years, with an average of 20.6 years. All received course credit for their participation; none reported a hearing impairment at the time of testing. These native English listeners were randomly assigned to one of six XAB similarity perception tests: (1) 22 (14 female and 8 male) for the 2 female, same-L1 (native+native) English conversations; (2) 19 (13 female and 6 male) for the 2 male, same-L1 (native+native) English conversations; (3) 19 (14 female and 5 male) for the 2 female, different-language (native+nonnative) conversations with a Chinese nonnative talker; (4) 21 (16 female and 5 male) for the 2 male, different-language (native+nonnative) conversations with a Chinese nonnative talker; (5) 20 (12 female and 8 male) for the 2 female, different-language (native+nonnative) conversations with a Korean nonnative talker; and (6) 20 (10 female and 10 male) for the 2 male, different-language (native+nonnative) conversations with a Korean nonnative talker.

A group of 40 native Korean talkers also participated in the XAB perceptual similarity test. All of these participants were undergraduate or graduate students at Northwestern University. They ranged in age from 20 to 40 years, with an average of 26.7 years. All received monetary compensation for their participation; none reported a hearing impairment at the time of testing. These native Korean listeners were randomly assigned to one of two XAB similarity perception tests: (1) 20 (9 female and 11 male) for the 2 female, same-L1 (native+native) Korean conversations; (2) 20 (13 female and 7 male) for the 2 male, same-L1 (native+native) Korean conversations.

Participants in the XAB test were seated in a soundproof booth. In each trial, the participant heard a triplet of speech samples (XAB) played over headphones. The stimuli were presented via a computer running Millisecond Inquisit 2.0. Listeners were instructed to imagine that the second talker (in samples A and B) was attempting to impersonate the model talker (in sample X). The task was to select A or B as a response to the question, "Which is more similar to the MODEL, A or B?" (see Figure 1 for an illustration of triplets). We adopted this impersonation scenario because the utterances to compare (X, A, and B) were all different in content and thus hard to directly compare in terms of their phonetic detail. Since

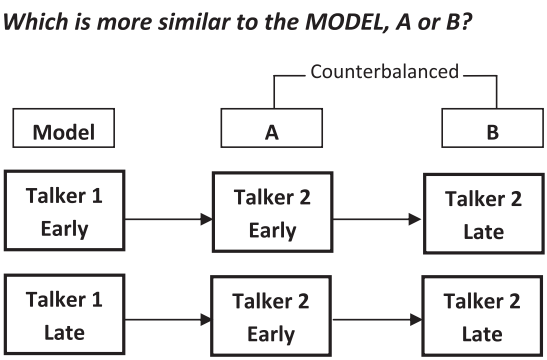

Figure 1. *An illustration of the triplets in XAB perception tests.*

accurate impersonation captures broad phonetic characteristics of a talker, it can be judged on the basis of utterances with different lexical content.

In each trial, three letter boxes for the triplet (X, A, and B) were displayed graphically on the monitor, X at the top, A at the left bottom, and B at the right bottom. Listeners' responses were entered by clicking on "A" or "B" on the monitor with a mouse. As shown in Figure 1, a triplet of speech samples consisted of a sample from one talker's early or late contribution to the conversation, and the partner's early contribution and late contribution. The inter-sample interval was 100 milliseconds. Following this scheme, all possible combinations of 12 samples from each conversation (3 from early and 3 from late parts of a conversation × 2 talkers) were created. The order of the A and B utterances was counterbalanced.

Within each XAB test condition there were 4 blocks, 2 containing speech samples from one conversation and 2 from the other conversation of the same type in terms of language pairs and talkers' gender. In each block, the model talker (X) varied from trial to trial. In this way, each listener was presented with 432 trials in total (2 conversations × 2 talkers per conversation × 2 time points per talker (i.e., early or late) × 2 orders of A and B presentation per triplet × 27 (3 × 3 × 3) possible XAB sample combinations). The trials in each block were given in random order with an 800 ms interval between trials. Participants were allowed to rest between blocks. It took approximately 1–1.5 hours for a participant to finish the task. The experimenter (MK) was careful to use only the target language of the experimental condition when interacting with participants.

2.3. *Nonnative talker accent rating*

As part of the original Wildcat Corpus compilation, a separate accent rating test was conducted as a means of assessing English language proficiency[2] for the nonnative talkers. Included in this test were speech samples from each of the nonnative talkers in the full corpus whose native language was either Korean or Chinese ($n = 34$, including the 4 Korean and 4 Chinese talkers selected for the present

study), as well as speech samples from the 8 native English talkers from the native+nonnative diapix conversations included in the present study. The native English talkers' samples were included as native-accent "anchors." A total of 378 speech samples were selected for the accent rating test: 3 samples × 3 times (the first 1/3, the middle 1/3, and the last 1/3 parts in a conversation) × 42 talkers = 378 samples. The duration of the samples was 1.5 to 2 seconds. The criteria for choosing a speech sample from a conversation were the same as for the XAB similarity tests except for the stimulus length, which, as described above, was 1–1.5 seconds for the XAB samples.

An independent group of 15 native English listeners (8 female and 7 male) participated in the accent rating test. All listeners were undergraduate students at Northwestern University and received course credit for their participation. They ranged in age from 19 to 22 years, with an average of 20 years. None of the participants reported any hearing impairment at the time of testing. None of these listeners had taken part in either the diapix recording for the Wildcat Corpus or the XAB perceptual similarity task in the present study.

The 378 speech samples were divided into 3 blocks, so that each block consisted of 126 trials (one sample per trial). The speech samples were presented over headphones in random order via a computer running Millisecond Inquisit 2.0 to the listeners in a sound proof booth. Listeners rated the accent of each speech sample on a scale of 1 (native-accented) to 9 (heavily foreign-accented) in response to the question: "How foreign is this talker's accent?" Listeners were allowed to take a rest between blocks. It took about 40 minutes for a listener to complete all 3 blocks. For the purpose of the current study, we then identified the average accentedness ratings for the subsample of nonnative talkers ($n = 8$) whose diapix conversations we examined here. These ratings were *z*-transformed to adjust for variation in use of the 9-point scale. Each talker's final average accentedness score was calculated based on the *z*-transformed accent ratings. While they cannot serve as a continuous factor in our regression models because of the small sample size ($n = 8$), these accentedness scores will allow us to gain some initial, tentative insight into the role played by phonetic proficiency in predicting phonetic convergence in native+nonnative conversations, and they confirmed that all of the nonnative talkers were easily identified by native listeners as foreign-accented English talkers (see Table 1).

3. Results

The data in this study were submitted to a set of generalized linear mixed effects regression analyses with the logit link function and binomial variance (Baayen 2010; Bates and Maechler 2010) to test whether phonetic convergence in spontaneous conversations is facilitated by relatively well matched language backgrounds (close "language distance") between the interlocutors. We included all three inter-

talker language distance conditions: the close (the same-L1/same-dialect condition), the intermediate (the same-L1/different-dialect condition), and the far (the different-L1 condition). Phonetic convergence was assessed on the basis of the likelihood of "late" response selection in the XAB perception tests. In all analyses, the model speech samples (X) were always from late parts (i.e., last third) of the diapix conversations.[3]

The dependent variable was the listener's response (early or late) in the XAB perception tests with snippets from the 4 same-L1/same-dialect conversations (native+native, 2 English and 2 Korean), 4 same-L1/different-dialect conversations (native+native, 2 English and 2 Korean), and 8 different-L1 conversations (native+nonnative, 4 with a Korean nonnative talker and 4 with a Chinese nonnative talker). The fixed effect factor was interlocutor language distance (same-L1/same-dialect vs. same-L1/different-dialect vs. different-L1). We also included three random effect factors, listener (in the XAB perception tests), talker (in the diapix conversations), and pair (in the diapix conversations). Pair gender (female or male), talker age, and conversation duration were each examined as possible control variables, but none improved the fit of the model, so no control variables were included in the final model. To make full comparisons among the three conditions of the fixed effect factor, the same generalized linear mixed effects regression analysis was carried out twice with different reference levels: one with the same-L1/same-dialect condition and the other with the same-L1/different-dialect condition as the reference level. The significance level was adjusted from 0.05 to 0.025 by Bonferroni correction, because two analyses were conducted on the same dataset for one set of results.

Results of these analyses showed a significantly higher likelihood of late responses for same-L1/same-dialect pairs (i.e., the "close" condition in language distance) than either for same-L1/different-dialect pairs (i.e., the "intermediate" condition in language distance) ($\hat{\beta} = -0.58$, $SE = 0.14$, $z = -4.05$, $p < 0.001$) or for different-L1 pairs (i.e., the "far" condition in language distance) ($\hat{\beta} = -0.5$, $SE = 0.12$, $z = -3.91$, $p < 0.001$). This indicates that phonetic convergence may be facilitated when the interlocutors' language distance is relatively close due to a shared native language background and a shared dialect background. Additionally, the results showed that the same-L1/different-dialect condition did not differ from the different-L1 condition in terms of the likelihood of late response ($\hat{\beta} = 0.07$, $SE = 0.12$, $z = 0.61$, $p = 0.53$). Thus, we can see that sharing the same L1 but not dialects between interlocutors (the "intermediate" condition) does not facilitate phonetic convergence compared to interlocutors' having different L1s. See Figure 2 for the overall phonetic convergence patterns of the three language distance conditions and Table 2 for each talker's convergence pattern and averaged percentage of XAB listeners' late sample selections (i.e., convergence likelihood).

A potential confounding factor in this analysis is the native versus nonnative status of the talkers. That is, nonnative talkers in the different-L1 condition may in general show a lower (or higher) tendency towards phonetic convergence than na-

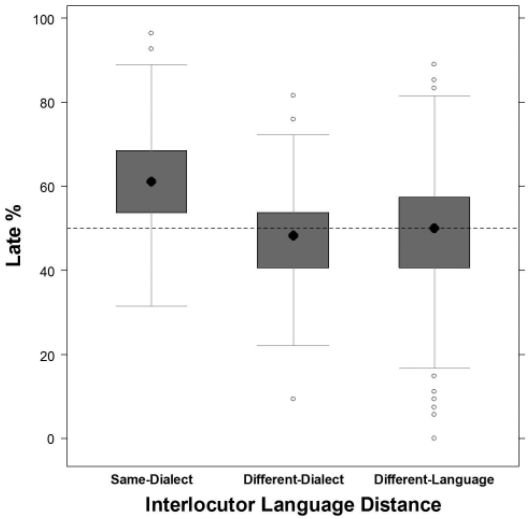

Figure 2. *Boxplot of perceived phonetic convergence as a function of interlocutor language distance.*[4] *Same-Dialect represents the same-L1|same-dialect condition, Different-Dialect, the same-L1|different-dialect condition, and Different-Language, the different-L1 condition. Late % represents the rate of XAB listeners' selection of the late sample as a better match to the model speech than the early sample.*

tive talkers of the target language. To control for this potential confound, we ran an additional set of analyses excluding two types of data: XAB responses on Korean speech samples by native Korean talkers (from Korean native+native conversations) and English speech samples spoken by nonnative English talkers (from English native+nonnative conversations). Then we compared phonetic convergence (in terms of likelihood of late sample selection) only by native English talkers conversing with either a native English-speaking interlocutor or a nonnative English-speaking interlocutor. The dependent variable was XAB response (early or late). The partner's target language status (native or nonnative) was entered as a fixed effect factor. Listeners in the XAB perception tests, and talkers and pairs in the diapix conversations were entered as random effect factors. Pair gender, talker age, and conversation duration were considered as possible control variables, but none increased the fit of the model. Therefore, no control variable was included in the final model. The same analysis was conducted twice with different reference levels for full comparisons among the three language distance conditions. Thus the significance level was adjusted from 0.05 to 0.025 by Bonferroni correction.

Results from this analysis confirmed the language distance effect found in the general model: Native English talkers paired with fellow native English talkers from the same dialect background were significantly more likely to be judged as having converged to their conversation partners than native English talkers who

Table 2. *Convergence patterns of the talkers in native+native and native+nonnative conversations.*

Native+Native conversations								
Target language	Convergence patterns[a]					Dialect comparison	Averaged late %[b]	
	Talker 1				Talker 2		Talker 1	Talker 2
English	ENF		↔		ENF	Same	59.9**	62.6**
	ENM		↔		ENM	Same	65.6**	61.4**
	ENF		←		ENF	Different	50.9	55.6*
	ENM				ENM	Different	47.5	50.1
Korean	KOF		↔		KOF	Same	61.3**	57.6*
	KOM		↔		KOM	Same	59.5**	65.2**
	←	KOF			KOF	→ Different	44.2*	42.2**
	←	KOM			KOM	→ Different	46.4*	40.4**

Native+Nonnative conversations								
Target language	Convergence patterns[a]					L2 accentedness	Averaged late %[b]	
	Native				Nonnative		Native	Nonnative
English	ENF				CHF	→ Light	47.1	23.8*
	ENM		→		CHM		57.4**	49.6
	ENM		←		KOM		47	55.3**
	ENF		←		KOF		50.6	64.1**
	ENM		←		KOM		49.5	59.3**
	ENF				CHF		51	52.9
	←	ENF			KOF		42.5*	47.8
	←	ENM			CHM	→ Heavy	43.6*	40.6**

Note. ENF = female native English, ENM = male native English, KOF = female native Korean, KOM = male native Korean, CHF = female native Chinese, CHM = male native Chinese.

a. Arrows indicate direction of any significant phonetic adjustment (i.e., ↔: symmetric convergence, ←: asymmetric convergence, when originating from one speaker to another, and divergence, when originating from one speaker to the outside).

b. Late % is the percentage of XAB listeners' selection of late speech samples for a better match to model samples. ** and * indicate Late % different from 50% within 99.5% and 95% confidence intervals, respectively.

were paired with either fellow native English talkers from different dialect backgrounds ($\hat{\beta} = -0.47$, $SE = 0.11$, $z = -4.05$, $p < 0.001$) or nonnative English talkers ($\hat{\beta} = -0.57$, $SE = 0.10$, $z = -5.25$, $p < 0.001$). Again, the same-L1/different-dialect condition and different-L1 condition did not differ in their effect on phonetic convergence ($\hat{\beta} = -0.1$, $SE = 0.1$, $z = -0.93$, $p = 0.34$).

Another potential confounding factor is the target language of conversations. It might be that conversations in one language lead to more phonetic convergence in general than conversations in another language. To investigate this possibility, we compared phonetic convergence patterns of the two types of native+native (same-

L1) conversations, namely, English conversations between two native English talkers and Korean conversations between two native Korean talkers. The dependent measure was listener response (early or late) from the XAB perception tests with speech samples from the 8 native+native (English or Korean) conversations. The fixed effect factors were dialect match (same or different) and target language (English or Korean). The dialect-by-language interaction was also included in the model. The random effect factors were listeners in the XAB perception tests, and talkers and pairs in the diapix conversations. Pair gender (female or male), conversation duration, and talker age were each examined to see if they improved the fit of the model as possible control variables. None of these control variables significantly improved the fit of the model, therefore no control variable was included in the final model. The condition number for multicollinearity of the fixed effect factors, namely, dialect match and target language was under 10 (value = 9.04).

The regression results indicated that closer language distance, namely, the same-dialect condition showed a significantly higher likelihood of phonetic convergence than the different-dialect condition ($\hat{\beta} = 0.46, SE = 0.07, z = 5.93, p < 0.001$). Additionally, English as the target language significantly increased the probability of late sample selection relative to Korean as the target language ($\hat{\beta} = -0.31$, $SE = 0.08$, $z = -3.6$, $p < 0.001$). Finally, the interaction of dialect match and target language was also significant ($\hat{\beta} = 0.23$, $SE = 0.11$, $z = 2.07$, $p < 0.05$), with a smaller dialect effect for English diapix conversations than for Korean diapix conversations.

In summary, we observed a facilitation of phonetic convergence in cases where the interlocutors had a relatively close language distance. Sharing the same language and the same dialect with a partner in a conversation facilitates phonetic convergence more than sharing the same language but not dialect or having different languages. However, the effects of interlocutors sharing the same language but not dialect and of interlocutors having different languages were not significantly different. In other words, sharing the same language and dialect was the only condition amongst the three language distance conditions where phonetic convergence was likely to occur.

At first glance, the pattern of greater convergence for same-dialect pairs than for different-dialect pairs seems to contradict the finding of Pardo's (2006) map task conversations, in which apparently different-dialect pairs (based on biographical information) showed convergence early in the conversation. Although dialectal variation was incidental to the main purpose of that study and was not an object of systematic analysis, Pardo did assess some degree of dialect variation in terms of the production of words that were selected to be diagnostic of marked regional differences in American English (i.e., the vowel contrasts in the following lexical sets: *marry-Mary-merry, cot-caught*, and *pen-pin*). This assessment indicated that all talkers distinguished these words in their speech. Thus, it is possible that the dialect differences between talkers in Pardo's study may have been less extensive than the dialect differences in the different-dialect pairs of the present study.

Therefore, the findings of Pardo (2006) and the current study might not actually be contradictory to each other in terms of the dialect effect. Also, in the present study we find a smaller dialect-based effect for English than Korean, which may mirror the magnitude of the dialect differences across the Korean and English different-dialect pairs. Extensive dialectal analyses were beyond the scope of the present study, but we should note that the different Korean dialects in question here are vastly different, possibly more so than the particular English dialect differences included here.

Additionally, it is important to acknowledge that target language was not the only difference between native+native English conversations and native+native Korean conversations. There were also other possible non-linguistic, cultural differences that may have played a role in this interaction, and a major difference between the two types of native+native conversations in this study was the different set of diapix scenes. Specifically, for the English conversations we used the shop scene whereas for the Korean conversations we used the beach scene. While we tried to make the scenes similar in their general style and overall difficulty (as noted above in the Methods section), there were numerous differences across the scenes that may have contributed to the dialect-by-language interaction. Nevertheless, it is noteworthy that the same overall pattern of greater convergence for same-dialect than for different-dialect pairs was observed in both languages.

Finally, while there was insufficient power in the dataset from the 8 different-L1 diapix conversations to fully test whether accentedness of the nonnative talker would predict convergence likelihood, we can note here a hint of a pattern that would need to be confirmed with additional data. Specifically, from Table 2 it appears that convergence of the nonnative talker to the native talker occurred in the conversations involving nonnative talkers with accentedness ratings in the middle of the accentedness scale. In contrast, the nonnative talkers at the extremes of the accentedness range were more likely to demonstrate divergence from the native talker. However, this informal observation is based on a very limited dataset and awaits future studies for verification.

4. Discussion

A central issue of concern in work on accommodation and other forms of convergence within conversations is the automaticity versus non-automaticity of interlocutor alignment. As proposed in Pickering and Garrod's (2004, 2006) interactive automatic alignment account of dialogue processing, nonconscious, automatic priming of linguistic representations is the engine that drives convergence across interlocutors. Because language comprehension and production are assumed to rely on the same underlying representations, talkers and listeners can become automatically "aligned" over the course of a conversation, in a process known as "input-output coordination." Importantly, such interactive alignment can occur

across all levels of language use (e.g., Garrod and Anderson 1987; Branigan, Pickering, and Cleland 2000; Cleland and Pickering 2003). On this view, the phonetic representations activated as part of comprehension are then available to automatically shape subsequent language production, and talkers and listeners should become more similar (or aligned) later in conversations than they were at the beginning. In the present study, however, talkers in spontaneous conversations showed not only alignment (convergence), but also a lack of alignment (maintenance and divergence), depending on the language and dialect match between interlocutors, providing evidence against a straightforward interactive alignment model. Instead, in accordance with Babel's (2009) suggestion of phonetic convergence as constrained by the existing phonetic repertoire, our results generally support the hypothesis that closer interlocutor language distance facilitates phonetic convergence between talkers in conversations.

The simplest version of the interactive alignment account, in which alignment is seen as a one-way process (i.e., toward convergence) driven solely by automatic priming processes, cannot easily explain the variant phonetic convergence patterns observed in the present study, as well as those found in similar studies (e.g., Pardo 2006; Babel 2009; Lewandowski and Dogil 2010; Pardo, Jay, and Krauss 2010; Nielsen 2011). Instead, a lack of alignment would result either from effortful repair of initially misaligned representations or from explicit and effortful interlocutor decisions not to align to each other (Pickering and Garrod 2004, 2006). For example, within Communication Accommodation Theory (CAT; Giles, Coupland, and Coupland 1991; Shepard, Giles, and Le Poire 2001), linguistic adjustment between talkers in a conversation is a strategic behavior with which "interlocutors achieve a desired social distance between self and interacting partners" (Shepard et al. 2001). Thus, in this view, convergence, divergence, and maintenance are all available to talkers in conversational interactions depending on the "desired social distance." In the diapix conversations of the present study, it seems unlikely that misalignment repair mechanisms would explain the variation in phonetic convergence patterns, because all diapix conversation pairs succeeded in finding almost all of the differences between the pictures. This suggests that the situation models of the interlocutors were likely to have been successfully aligned by the end of the conversations, despite the different patterns of phonetic adjustment (convergence, maintenance, and divergence). Thus in our study, as in others (e.g., Bourhis and Giles 1977; Bilous and Krauss 1988), it appears that misalignment at the phonetic level can occur without a persistent failure of comprehension between talkers.

It is plausible that the native talkers in the different-L1 (native+nonnative) talker pairs adopted a speech production strategy that could be characterized as a "decision" against phonetic convergence. Specifically, if they adopted a clear speaking style or "foreigner talk" (Ferguson 1971; Clyne 1981; Perdue 1984; Evans 1987) as a means of assisting their nonnative partners in the diapix task, then their productions may have diverged phonetically from their partner's productions over

the course of the diapix conversations. Furthermore, the nonnative status of one's partner is likely to be particularly salient in the case of a very heavily-accented interlocutor. Thus, in the case of the native talkers in the native+nonnative talker conversations, the variation in phonetic convergence from negative (divergence) to zero (maintenance) to positive (just one case) could be explained by the notion of a strategic "decision" to achieve a desirable social distance between self and interacting partners (e.g., Gallois et al. 1988; Shepard et al. 2001). However, in a study with highly controlled materials such as the current study, such a decision could have been based on linguistic rather than social-psychological considerations. That is, the adoption of a nonnative-oriented speech mode in the diapix conversations would serve the purpose of intelligibility enhancement, which would be linked to target language proficiency of the nonnative partner, rather than the explicit manipulation of social distance. In this regard, it is potentially relevant that all of the nonnative talkers in the present study (regardless of proficiency level) were very clearly foreign-accented (accentedness z score ≥ -0.58). Therefore, the phonetic adjustment patterns on the part of the native talkers in the native+nonnative diapix conversations were less likely to be socially (desire for greater social distance) than linguistically (attempt to enhance intelligibility) motivated, since the social considerations would presumably have been similar across all nonnative partners. Nevertheless, it remains for future work to sort out the contributions of the nonnative talker's proficiency in the target language and the native talker's implicit attitudes towards a partner from a different native language background.

This strategic explanation is less plausible for the nonnative talkers than for the native talkers in the native+nonnative conversations. In the context of the diapix task, which is essentially a cooperative problem solving game, it seems unlikely that the nonnative talkers would have explicitly decided not to align to their native partners or have adopted either a socially- or linguistically-motivated strategy of increased phonetic distance. It may seem that the strong desire of nonnative talkers to express their linguistic and cultural identity through their English speech production might function as a motivation toward social distancing and away from convergence (Zuengler 1982). However, the desire to achieve a high level of English intelligibility (i.e., to be understood) would likely have been a strong motivation for these nonnative talkers to align to their native partners as a means of increasing overall English intelligibility.

What factors might then have interfered with nonnative talkers' phonetic convergence? On the particular case of dialogues involving native and nonnative talkers, Costa and colleagues (2008) suggest that an imbalance in target language proficiency across the native and nonnative talkers may inhibit the automatic alignment mechanism proposed by Pickering and Garrod (2004, 2006). In particular, Costa and colleagues (2008) suggest the high attentional demand and processing load involved in nonnative speech production may interfere with the overall automaticity of the conversation and inhibit inter-talker alignment. Thus, while the nonnative talkers may have had a strong intelligibility-based motivation to align to their native inter-

locutors, the extra demands of second language production may have interfered with any alignment process. Presumably, this interference would have been greater for nonnative talkers with relatively low target language (i.e., English) proficiency.

Just as the high attentional demands and processing load involved in second language production may have acted as an inhibitory influence for the nonnative talkers, it is also possible that the high attentional demands and processing load involved in the perception of nonnative speech (i.e., foreign-accented speech) on the part of the native talkers may have had an inhibitory effect with respect to the processes of phonetic convergence by the native talkers. This possibility was supported by the data showing greater phonetic convergence in the native+native than in the native+nonnative diapix conversations. Moreover, in the diapix conversations in the same-L1/different-dialect condition, the relatively high attentional demands and processing load involved in cross-dialect communication may have had a similar inhibitory effect with respect to the processes of phonetic convergence. Thus, on this view, the general language-distance effect on phonetic convergence observed in this study may be related to an inhibitory effect of relatively high processing load due to both nonnative speech production and foreign-accented speech perception.

Overall, in these diapix conversations, we speculate that the observed language-distance-linked phonetic convergence patterns can be accounted for by two parallel mechanisms: the need for intelligibility and the extra demands of nonnative speech production and perception. The first mechanism, the need for intelligibility, may inhibit phonetic convergence due to the adoption of an intelligibility-enhancing speech style ("clear" or "foreigner" speech). This would occur particularly in situations where the interlocutors have a relatively far language distance due to different native languages or dialects – i.e., in situations where speech intelligibility is most likely to be compromised. The large literature on clear speech (for a review see Smiljanic and Bradlow 2009) has identified a wide range of acoustic-phonetic features of clear versus conversational/plain speech. Thus, this account of the observed phonetic convergence patterns could be empirically verified by an acoustic analysis of the speech in our diapix conversation recordings. For example, we predict that markers of clear speech (slower speaking rate, more pauses, reduced final consonant lenition, etc.) would be most evident in later portions of native talker speech in diapix conversations with low proficiency nonnative interlocutors relative to early portions of the same conversations, and relative to later portions of conversations with high proficiency nonnative interlocutors. An empirical prediction of the second proposed mechanism is that phonetic convergence within interlocutors with an initially close language distance should be inhibited by other sources of an increase in high attentional demands and processing load. For example, the addition of background noise or the simultaneous performance of some other cognitive resource demanding task (e.g., operating a motor vehicle or another mechanical device) should also have an inhibitory effect on phonetic convergence even when language distance is relatively close.

While our data suggest a link between phonetic convergence and language distance, there are several other uncontrolled factors in the diapix conversations that may have contributed to the pattern of findings. First, a potentially relevant factor in phonetic convergence patterns is the talker roles adopted in the conversations. As discussed previously, Pardo (2006) found variation in phonetic convergence patterns depending on the talker role as the giver or receiver of information. In the Wildcat Corpus diapix conversations, there were no specific roles assigned to the talkers, and the diapix task was designed to avoid the adoption of "leader" and "follower" roles. It is possible that the talkers could have spontaneously chosen different roles. However, in other analyses of the full set of diapix conversations in the Wildcat Corpus (including those in the present study), we have observed considerable role variation both within and across the diapix conversation pairs (Van Engen et al. 2010).

A second factor that may have influenced our findings is that we assessed convergence on the basis of early and late samples extracted from within the diapix conversations themselves. This is in contrast to Pardo (2006) in which convergence was assessed on the basis of comparison between identical items (i.e., the same word) produced at pre-test, within the conversation (at task), and at post-test. Note that the pre- and post-test productions were read speech, whereas the within-conversation item was spontaneous speech. With this design, Pardo (2006) was able to determine that the talkers converged relatively early in the conversation (pre vs. task comparison) and that the convergence increased by the end and even beyond the end of the conversation (task vs. post, pre vs. post comparisons). In the present study, we always compared spontaneous, within-conversation items with other spontaneous, within-conversation items (from early and late portions of the conversations). While this design does not allow us to assess the pace, persistence, and/or attenuation of convergence over time, it has the important advantage of avoiding the confound of read versus spontaneous speech that is inherent in comparisons with pre- and post-test items.

A limitation of our data set is that we have presented only listeners' perceptual judgments in the XAB perception tests as our index of phonetic convergence. The set of possible acoustic features that underlie the observed patterns of phonetic convergence is very extensive, and unfortunately we have not yet traced the perceptual patterns to specific acoustic-phonetic features in the diapix recordings. A significant challenge in this regard is the fact that our XAB speech samples were all different utterances, making direct acoustic comparison at the lexical or sub-lexical levels impossible. Nevertheless, we attempted some acoustic measurements in terms of features such as speaking rate and pitch range. However, these analyses did not yield interpretable findings. One possible approach to this issue involves developing a different set of diapix scenes in which numerous repetitions of key phrases are effectively elicited at various points in the conversation (for progress on this front see Baker and Hazan 2009). As described in the Introduction, studies that have identified acoustic features of convergence have involved

shadowing or speech imitation, which may highlight fine grained phonetic adaptations that are less apparent in spontaneous dialogues. Nevertheless, the XAB perceptual judgment tasks in both the present study and Pardo (2006) indicate that some interlocutor-oriented adjustment is operating in spontaneous dialogues and ultimately we should be able to identify the acoustic parameters that give rise to the perception of convergence. We attempted to eliminate (or at least, minimized) all obvious non-phonetic bases for similarity judgment in the present study by selecting samples that were similar in acoustic length, fluently produced throughout their duration, and that contained no evident hesitations, disfluencies, background noise, or back channeling from the other talker. Thus, it is unlikely that the listeners had clear access to non-phonetic dimensions (such as broader linguistic, social, or esoteric information) on which to base their XAB judgments. The challenge of identifying the critical phonetic features remains open.

While the present data suggest that the language experiences of interlocutors play an important role in determining patterns of phonetic convergence, future work would ideally examine a far larger set of conversations where a wider range of dialect and native-language pairings can be explored. With more spontaneous conversation recordings available it may be possible to shift the balance of data from many listener judgments of relatively few speech samples to many direct measurements of speech adjustments in terms of precise phonetic, phonological, and other linguistic structural features in a large sample of conversations. As tools for capturing and analyzing very large speech corpora become more widely available (e.g., from radio, television, and other broadcast media), this approach may become quite feasible and will likely lead to breakthroughs in our understanding of the conditions that lead to phonetic change on individual and population levels.

5. Conclusion

We have obtained evidence in support of a relationship between interlocutor language distance and phonetic convergence in conversations within pairs of native English talkers, within pairs of native Korean talkers, and in conversations between a native English talker and a nonnative talker of English. Specifically, we found that within pairs of native talkers of the target language (either English or Korean), a match in regional dialect facilitated phonetic convergence. This stands in contrast to a lack of phonetic convergence between pairs of talkers who did not share a regional dialect or came from different native language backgrounds (with one being a native and the other a non-native talker). We interpret these results as suggesting that phonetic convergence between talker pairs that vary in the degree of their initial language alignment may be dynamically mediated by two parallel mechanisms: the need for intelligibility and the extra demands of nonnative speech production and perception.

Appendix A

Pictures for the diapix conversations

a.

b.

Figure A1. *Shop scenes for English conversations.*

a.

b.

Figure A2. *Beach scenes for Korean conversations.*

Appendix B

Lists of the differences between the two versions of the English and Korean pictures.

Table B1. *Differences in the English pictures (shop scenes).*

Version A	Version B
Rightmost sign shows a cat and bowl	Rightmost sign shows a sheep on grass
Middle shop left sign mentions pork chop	Middle shop left sign mentions lamb chop
Middle shop right sign mentions cheese soup	Middle shop right sign mentions beef soup
The woman's shoes are red	The woman's shoes are green
No beehive in the tree	Beehive in the tree
Rightmost shop door has paw prints	Rightmost shop door has no paw prints
Rightmost shop name is "Pet Shop"	Rightmost shop name is Pete's Pet Shop
No bench on the street	Bench on the street
The boy is carrying a box	The boy is not carrying anything
Leftmost shop name is "Boss's Booze"	No name on the leftmost shop

Table B2. *Differences in the Korean pictures (beach scenes).*

Version A	Version B
Top left girl is running after her hat	Top left girl is walking with her hat on
Sitting lady has curled hair	Sitting lady has straight hair
Dog with sitting lady is a poodle	Dog with sitting lady is not a poodle
Sitting lady holding a fan in left hand	Sitting lady holding nothing in left hand
Pizza box on the table	Menu on the table
Lady by the table is wearing pants	Lady by the table is wearing a skirt
Clouds in the sky	No clouds in the sky
One wave in the sea	Two waves in the sea
There is a boy in the sea	There is no boy in the sea
Food stand menu reads '사과 자두' (/sakwa cadwu/, apple plum)	Food stand menu reads '수박' (/swupak/, watermelon)
Boys are not playing with a ball	Boys are playing with a soccer ball

Note. There were 11 differences in the Korean pictures, but the subjects were instructed to find 10 differences between the pictures.

Appendix C

Utterances used for speech samples in the XAB perception tests

Table C1. *Samples from English conversations between two native English talkers.*

Conversation	Talker	Early samples	Late samples
ENF-ENF 1	Talker 1	do you have a beehive	I thought it was the part of the honey comb but
		on the tree in the right corner	also has some bees below it
		it's like pretty much in front of it	she's walking to the left
	Talker 2	is it at the top	leaves in the trees.
		she has red shoes in mine.	does the little boy have a visor
		I don't have a park bench	does he have the mustache the man
ENF-ENF 2	Talker 1	bottom left or top left	there's like two levels of green
		mine has red high heel shoes	and groceries all caps
		do you have foot prints on the door	and there's nothing in front of the like
	Talker 2	how about I'll tell you what's in my picture	yeah just the green beach
		she's caring a blue purse	you had the two boxes
		and bees flying around it	but there's nothing around that corner
ENM-ENM 1	Talker 1	pink sign on the left says Boss's Booze	crossing a long squiggly line
		colors also pink purple pink	and a small squiggly line to her right
		the white band yeah and then the sky	slightly to the left of the angry man
	Talker 2	does yours have a martini glass on it	where is the woman in your picture
		and pink purple pink.	is there anything on the ground in your picture
		and you said the tan doors.	and you had a different sign for groceries night
ENM-ENM 2	Talker 1	okay here's what I have, alright here's my grocery store	I do little yellow uh circles you know
		I've got a pork chop special	no I've got paw prints on the door
		above that little yellow window	can you make sure can you confirm you have nine differences
	Talker 2	starting with the text on the page	ah kid's wearing a blue visor
		do you have groceries Pete's Pet Shop	green shirt blue shorts and kind of blue shoes
		one ninety five three exclamation points	five apples and uh

Note. ENF = female native English talker; ENM = male native English talker.

Table C2

a. *Samples from Korean conversations between two native Korean talkers.*

Conversation	Talker	Early samples	Late samples
KOF-KOF 1	Talker 1	옆에는 빨간 사과가 있어요	뭐지 부채질을 하고 있어요
		그 밑에는 시원한 수박	바다 바다 위쪽에
		그 옆에 빨간 사과	바다 위쪽으로 왼쪽이요
	Talker 2	주전자 뚜껑 색깔 뭐야	어떤 여자애 뒷 모습 보여요
		아 잠깐만 시원한 수박	그 다음에 옆에 강아지 보여요
		그 아랜 차가운 아이스 티 있고	갈매기 가운데 보여요
KOF-KOF 2	Talker 1	갈매기 두 마리랑	그 골대는
		구름 두 개가 따로 이렇게 있어요	축구 하는 애들 사이에 있어요
			발자국 네 개 있어요
		그리고 저 그 어 그 파도 두 갠데	
	Talker 2	두 개가 이렇게 있고	예 오른손은 뒤로 이렇게
		풍덩 사람 빠지는 거	돼지 꼬리 긴 거
		모자 날아 가는 거 있어요	모자 날아 가는 표시 하고
KOM-KOM 1	Talker 1	오른쪽 위부터	아 가운데 큰 파도
		조그만 달팽이들 있고요	그 밑에 엠 자 하나 그죠
		반팔 티 입었어요	그게 그 왼쪽으로가
	Talker 2	그 그리고 그 밑에로는	엠 자가 쪼끄만 게 하나 있고
		너울지는 게 지금	그럼 토탈 세 개가 있는 건가요 그 쪽에
		달팽이처럼 해 가지고	엠 자 같이 생긴 게 여덟 개
KOM-KOM 2	Talker 1	파전하고 피자	아저씨 목걸이는 있나요
		그 다음에 의자가 두 개	악세사리를 하고 있는
		그 다음에 스커트는 빨간 색	왼쪽에 있는 아가씨도 없고요
	Talker 2	피자가 두 개 있네요	옷걸이는 없습니다
		하얀 색에 노란색	오른쪽에 큰 파도는
		흰 색이랑 파란 색이	피서는 부산에서

Note. KOF = female Korean nonnative English talker; KOM = male Korean nonnative English talker.

b. *English translations of samples from Korean conversations between two native Korean talkers.*

Conversation	Talker	Early samples	Late samples
KOF-KOF 1	Talker 1	there is a red apple next to it	what she's fanning herself
		a watermelon below it	the sea above the sea
		a red apple next to it	above the sea to the left
	Talker 2	what's the color of the lid of the kettle	can you see the back of a girl
			then can you see a dog next to her
		/a/ ('ah') wait a cool watermelon	do you see a seagull at the center
		you have cool iced tea below it	

Table C2 (*Continued*)

Conversation	Talker	Early samples	Late samples
KOF-KOF 2	Talker 1	with two seagulls	that goalpost
		do you have two clouds separately	is it between the kids playing soccer
		and /ce/ (‘that’) /ku/ (‘the’) /e/ (‘uh’) I have two waves	are there four footprints
	Talker 2	I have two of them	yes the right hand is backward like this
		a person is falling with a splash	a long pig tail
		do you have a hat blowing	mark the hat blowing
KOM-KOM 1	Talker 1	from the top right corner	ah the big wave at the center
		there are small snails	a letter M below it right
		is he/she wearing a t-shirt	that is, the left side is
	Talker 2	/ku/ (‘the’) and below it	there’s a small letter M, and
		the shape of the wave is	then do you have three in total there
		it looks like a snail	eight ones that looks like M
KOM-KOM 2	Talker 1	/pacen/ (‘green onion pancake’) and pizza	does he have a necklace
		next two chairs	the one who’s wearing jewelry
		next the skirt is red-colored	there is no lady on the left side
	Talker 2	there are two pieces of pizza	there is no hanger
		white and yellow	the big wave on the right side is
		white and blue	your vacation in Busan

Note. KOF = female Korean nonnative English talker; KOM = male Korean nonnative English talker.

Table C3. *Samples from English conversations between a native English talker and a Chinese learner of English.*

Conversation	Talker	Early samples	Late samples
ENF-CHF 1	ENF	is she wearing red shoes	what side is the poster on
		is it a really really light pink	for the sheep dogs
		does she have a blue bag	is the man holding anything
	CHF	it’s green	is it have leaves
		on the middle part	the house of the beef
		they don’t have a name	I don’t know how to say
ENF-CHF 2	ENF	for the names of the shops	the missing beehive
		mine is missing a name	the paw print on the door
		next to the lamb chop special	the pet shop sign
	CHF	there are actually signs	and there’s a foot prints
		yeah it’s basically red	like a dog’s foot prints
		pork chop special	oh mine only have um

Table C3 (*Continued*)

Conversation	Talker	Early samples	Late samples
ENM-CHM 1	ENM	on the lower left hand corner What other things are they wearing It doesn't have a name on mine	um we're talking about the baby with a light green background and that's it no there's no hive either
	CHM	in the downside of the paper so the color maybe differences and the hair is yellow	so you only have one animal below the bar I have nothing oh let me see
ENM-CHM 2	ENM	on the left that says Boss's Booze no mine says pork chop special you have two boxes behind the boy	does he have brown shoes on you said there's uh it's a sheep and I think we found ten
	CHM	on the left of what, pardon me do you have two pictures and I have one there's a sign that says groceries	and then in the groceries like uh the guy wearing the apron I have leaves on the trees

Note. ENF = female native English talker; ENM = male native English talker; CHF = female Chinese nonnative English talker; CHM = male Chinese nonnative English talker.

Table C4. *Samples from English conversations between a native English talker and a Korean learner of English.*

Conversation	Talker	Early samples	Late samples
ENF-KOF 1	ENF	there's a tree to the left of the tree and wearing green shoes	is your sky blue do you have nine differences with the pink sign
	KOF	what is Pete's Pet Shop footsteps I don't understand	next to cat some vegetable in the board grocery store
ENF-KOF 2	ENF	mine says Boss's Booze Pete's Pet Shop yours says it's paw prints on the door	my little boy is holding a box is there anything in the box and the martini glass
	KOF	there's no sign martini glasses maybe apples	no he's just bars sign white shirts
ENM-KOM 1	ENM	it says Boss's Booze that's on the top of the sign okay I have the same thing	okay so that's another difference do you have anything in the sky okay yeah I don't have a bench there
	KOM	drawing color is red one I can't find any blue one the window what's color of window	the sign have one lamb the color is yellow bee's house and there's some bee

Table C4 (*Continued*)

Conversation	Talker	Early samples	Late samples
ENM-KOM 2	ENM	I have bees buzzing right below the beehive	do you have two boxes next to him on the ground
		with the strap that goes over her shoulder	a blue sign with the dog the brown dog
		starting to the left side of the picture	do you have anything else in your picture that we haven't talked about
	KOM	mine wears white one	but I cannot recognize this
		mine wears red shoes	one on top of another
		that's left side middle ground	and the top one is kind of open

Note. ENF = female native English talker; ENM = male native English talker; KOF = female Korean nonnative English talker; KOM = male Korean nonnative English talker.

Acknowledgments

The authors would like to thank Kelsey Mok, Page Piccinini, and Sudha Ayala for their assistance in data collection. This work was supported by grant R01-DC005794 from NIH-NIDCD.

Correspondence e-mail address: midam-kim@northwestern.edu

Notes

1. The content of diapix conversations changes quickly enough as talkers move across the scenes that, for XAB triplets in which the X sample was taken from a late portion of the conversation, there was no greater semantic or lexical overlap between the model (X) and late test sample than between the model and early test sample. This was confirmed by a count of the number of words shared within all of our triplets (provided in Appendix C). On average, model and late samples shared 6% of their words, and model and early samples shared 7% of their words.
2. Note that L2 proficiency in the current study was measured by accentedness ratings, which is a proxy for phonetic proficiency specifically rather than for language proficiency in general.
3. Even though the XAB perception tests included trials where X was an "early" sample (i.e., from the first third of the diapix conversation) as well as trials where X was a "late" sample (i.e., from the last third of the diapix conversation), we limited our main analyses to XAB trials where the model, X, was a late sample. We chose to focus on the late X results based on the assumption that "on-line" phonetic accommodation patterns (a talker accommodating to the partner in real time at the end of the conversation, namely, accommodating to the partner's "late" samples) would be more stable than "global" phonetic accommodation (a talker changing speech style at the end of the conversation, reflecting the partner's speech from the beginning, namely, the partner's "early" samples). To test this assumption, we conducted a generalized linear mixed effects analysis. The dependent measure was all talkers' XAB responses (early vs. late), the fixed effect factor was model timing (early X vs. late X), and the random effect factors were listener, talker, and pair. The results showed that models with late samples were significantly more likely to lead to late sample

selection by the XAB perception test participants ($\hat{\beta} = 0.12$, $SE = 0.015$, $z = 8.29$, $p < 0.001$), thus providing more judgments towards phonetic convergence. Therefore, we excluded all XAB perception test responses where the models were early samples.

4. The box plot was made with the xyplot function in R (Sarkar 2008).

References

Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry Thompson, & Regina Weinert. 1991. The HCRC Map Task corpus. *Language and Speech* 34. 351–366.

Baayen, R. Harald. 2010. languageR: Data sets and functions with "Analyzing linguistic data: A practical introduction to statistics" (R package version 1.0). http://CRAN.R-project.org/package=languageR.

Babel, Molly. 2009. Phonetic and social selectivity in speech accommodation. Ph.D. dissertation, Department of Linguistics, University of California, Berkeley.

Babel, Molly. 2010. Dialect convergence and divergence in New Zealand English. *Language in Society* 39. 437–456.

Baker, Rachel, & Valerie Hazan. 2009. Acoustic-phonetic characteristics of naturally-elicited clear speech in British English. *Journal of the Acoustical Society of America* 125. 2729.

Bates, Douglas, & Martin Maechler. 2010. lme4: Linear mixed-effects models using S4 classes (R package version 0.999375-35). http://CRAN.R-project.org/package=lme4.

Bilous, Frances R., & Robert M. Krauss. 1988. Dominance and accommodation in the conversational behaviors of same- or mixed-gender dyads. *Language and Communication* 8. 183–194.

Birdsong, David. 2004. Second language acquisition and ultimate attainment. In Alan Davies & Catherine Elder (eds.), *Handbook of Applied Linguistics*, 82–105. London: Blackwell.

Bourhis, Richard Y., & Howard Giles. 1977. The language of intergroup distinctiveness. In Howard Giles (ed.), *Language, Ethnicity and Intergroup Relations*, 119–135. London: Academic.

Bradlow, Ann R., & Tessa Bent. 2008. Perceptual adaptation to nonnative speech. *Cognition* 106. 707–729.

Branigan, Holly P., Martin J. Pickering, & Alexandra A. Cleland. 2000. Syntactic co-ordination in dialogue. *Cognition* 75. B13–25.

Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.

Clark, Herbert H., & Gregory L. Murphy. 1982. Audience design in meaning and reference. In Jean François Le Ny & Walter Kintsch (eds.), *Language and Comprehension* (Advances in Psychology 9), 287–299. Amsterdam: North-Holland Publishing Company.

Cleland, Alexandra A., & Martin J. Pickering. 2003. The use of lexical syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language* 49. 214–230.

Clyne, Michael G. 1981. "Second generation" foreigner talk in Australia. *International Journal of the Sociology of Language* 28. 69–80.

Cooper, Robin Panneton, & Richard N. Aslin. 1990. Preference for infant-directed speech in the first month after birth. *Child Development* 61. 1584–1595.

Costa, Albert, Martin J. Pickering, & Antonella Sorace. 2008. Alignment in second language dialogue. *Language and Cognitive Processes* 23. 528–556.

Delvaux, Véronique, & Alain Soquet. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64. 145–173.

Eisner, Frank, & James M. McQueen. 2005. The specificity of perceptual learning in speech processing. *Perception & Psychophysics* 67. 224–238.

Evans, Bronwen G., & Paul Iverson. 2007. Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America* 121. 3814–3826.

Evans, Mary. 1987. Linguistic accommodation in a bilingual family: One perspective on the language acquisition of a bilingual child being raised in a monolingual community. *Journal of Multilingual and Multicultural Development* 8. 231–235.

Ferguson, Charles A. 1971. Absence of copula and the notion of simplicity: A study of normal speech, baby talk, foreigner talk and pidgins. In Dell H. Hymes (ed.), *Pidginization and Creolization of Languages*, 141–150. New York: Cambridge University Press.

Flege, Jim E. 1999. Age of learning and second-language speech. In David Birdsong (ed.), *Second Language Acquisition and the Critical Period Hypothesis*, 101–132. Hillsdale, NJ: Lawrence Erlbaum.

Gallois, Cynthia, Arlene Franklyn-Stokes, Howard Giles, & Nikolas Coupland. 1988. Communication accommodation in intercultural encounters. In Young Yun Kim & William B. Gudykunst (eds.), *Theories in Intercultural Communication*, 157–185. Newbury Park, CA: Sage.

Garrod, Simon, & Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27. 181–218.

Giles, Howard, Nikolas Coupland, & Justine Coupland. 1991. Accommodation theory: Communication, context, and consequence. In Howard Giles, Justine Coupland, & Nikolas Coupland (eds), *Contexts of Accommodation: Developments in Applied Sociolinguistics*, 1–68. Cambridge: Cambridge University Press.

Giles, Howard, & Tania Ogay. 2007. Communication accommodation theory. In Bryan B. Whaley & Wendy Samter (eds.), *Explaining Communication: Contemporary Theories and Exemplars*, 293–310. Mahwah, NJ: Lawrence Erlbaum.

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

Goldinger, Stephen. D., & Tamiko Azuma. 2004. Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review* 11. 716–722.

Greenwald, Anthony. G., Debbie E. McGhee, & Jordan L. K. Schwartz. 1998. Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology* 74. 1464–1480.

Kraljic, Tanya, Susan E. Brennan, & Arthur G. Samuel. 2008. Accommodating variation: Dialects, idiolects, and speech processing. *Cognition* 107. 51–81.

Kraljic, Tanya, & Arthur G. Samuel. 2005. Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology* 51. 141–178.

Labov, William, Sharon Ash, & Charles Boberg. 2006. *Atlas of North American English*. New York: Mouton de Gruyter.

Lewandowski, Natalie, & Grzegorz Dogil. 2010. Identity negotiation in native-nonnative dialogs: Quantifying phonetic adaptation. In Rodulf De Cillia, Helmut Gruber, Michal Krzyzanowski, & Florian Menz (eds.), *Diskurs – Politik – Identität Discourse – Politics – Identity: Festschrift für Ruth Wodak*, 389–399. Berlin, New York: Mouton de Gruyter.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In William J. Hardcastle & Alain Marchal (eds.), *Speech Production and Speech Modeling*, 403–439. Amsterdam: Kluwer Academic.

Miller, Rachel M., Kauyumari Sanchez, & Lawrence D. Rosenblum. 2010. Alignment to visual speech information. *Attention, Perception, & Psychophysics* 72. 1614–1625.

Munro, Murray J., Tracey M. Derwing, & Jim E. Flege. 1999. Canadians in Alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics* 27. 385–403.

Namy, Laura, Lynne C. Nygaard, & Denise Sauerteig. 2002. Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology* 21. 422–432.

Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics*. doi:10.1016/j.wocn.2010.12.007.

Norris, Dennis, James M. McQueen, & Anne Cutler. 2003. Perceptual learning in speech. *Cognitive Psychology* 47. 204–238.

Nygaard, Lynne C., & David B. Pisoni. 1998. Talker-specific learning in speech perception. *Perception & Psychophysics* 60. 355–376.

Nygaard, Lynne C., Mitchell S. Sommers, & David B. Pisoni. 1994. Speech perception as a talker-contingent process. *Psychological Science* 5. 42–46.

Pardo, Jennifer S. 2006. On phonetic convergence during conversational interaction. *Journal of Acoustic Society of America* 119. 2382–2393.

Pardo, Jennifer S., Isabel Cajori Jay, & Robert M. Krauss. 2010. Conversational role influences speech imitation. *Attention, Perception, & Psychophysics* 72. 2254–2264.

Payton, Karen L., Rosalie M. Uchanski, & Louis D. Braida. 1994. Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America* 95. 1581–1592.

Perdue, Clive (ed.). 1984. *Second Language Acquisition by Adult Immigrants: A Field Manual*. Rowley, MA: Newbury House.

Picheny, Michael A., Nathaniel I. Durlach, & Louis D. Braida. 1985. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research* 28. 96–103.

Pickering, Martin. J., & Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral Brain Science* 27. 169–190.

Pickering, Martin J., & Simon Garrod. 2006. Alignment as the basis for successful communication. *Research on Language and Computation* 4. 203–228.

Samuel, Arthur G., & Tanya Kraljic. 2009. Perceptual learning for speech. *Attention, Perception, Psychophysics* 71. 1207–1218.

Sancier, Michele L., & Carol A. Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25. 421–436.

Sarkar, Deepayan. 2008. *Lattice: Multivariate Data Visualization with R.* New York: Springer.

Shepard, Carolyn A., Howard Giles, & Beth A. Le Poire. 2001. Communication accommodation theory. In W. Peter Robinson & Howard Giles (eds.), *The New Handbook of Language and Social Psychology*, 33–56. New York: Wiley.

Shockley, Kevin, Laura Sabadini, & Carol A. Fowler. 2004. Imitation in shadowing words. *Perception and Psychophysics* 66. 422–429.

Sidaras, Sabrina K., Jessica E. D. Alexander, & Lynne C. Nygaard. 2009. Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America* 125. 3306–3316.

Smiljanic, Rajka, & Ann R. Bradlow. 2009. Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Linguistics and Language Compass* 3(1). 236–264.

Smith, Caroline L. 2007. Prosodic accommodation by French speakers to a nonnative interlocutor. *Proceedings of the XVIth International Congress of Phonetic Sciences.* Saarbrucken, Germany.

Uchanski, Rosalie M. 2005. Clear speech. In David B. Pisoni & Robert Remez (eds.), *The Handbook of Speech Perception*, 207–235. Malden, MA/Oxford, UK: Blackwell.

Uther, Maria, Monja Knoll, & Denis Burnham. 2007. Do you speak E-N-G-L-I-S-H? A comparison of foreigner- and infant-directed speech. *Speech Communication* 49. 2–7.

Van Engen, Kristin J., Melissa Baese-Berk, Rachel E. Baker, Midam Kim, & Ann R. Bradlow. 2010. The Wildcat Corpus of native- and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language & Speech* 53. 510–540.

Wassink, Alicia Beckford, Richard A. Wright, & Amber D. Franklin. 2007. Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics* 35. 363–379.

Zuengler, Jane. 1982. Applying accommodation theory to variable performance data in L2. *Studies in Second Language Acquisition* 4. 181–192.