ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 2aSC: Linking Perception and Production (Poster Session)

2aSC3. The role of voice similarity in accommodation
Sophie A. Walters*, Molly E. Babel and Grant McGuire

*Corresponding author's address: Linguistics, University of British Columbia, 2613 West Mall, Vancouver, British Columbia V6T1Z4, sophia.alex.walters@gmail.com

Studies of accommodation show that some talkers are perceived as accommodating more than others. One possibility is that the similarity of the shadower's voice to a model talker's can account, in part, for the amount of perceived accommodation. To determine this, we conducted an auditory naming task having eight model talker voices previously rated for attractiveness and prototypicality, such that the Most Attractive and Least Attractive and Most Typical and Least Typical voices for each gender were used as models. Twenty participants completed an auditory naming task with these eight voices. A separate group of twenty listeners rated the similarity of model tokens and shadower's baseline productions using a visual analogue scale. The results of this task were compared to the perceived accommodation results from a separate AXB rating task. Overall, female voices that were more different from the models showed more accommodation. This effect was not found for males, who generally showed less accommodation overall. These findings suggest that talkers either accommodate more when their voice is more distinct from the model talker's voice, or perhaps more likely, that such changes are more perceptible to listeners. Further explorations of the data are underway to tease apart these possibilities.
INTRODUCTION

As people acquire language, they adopt the speech sounds that comprise the language and dialect they are immersed in. Although some imitate more than others, listeners perceive acoustic variability and are able recognize patterns that allow them to shape their own speech while simultaneously learning from other speakers. Albeit a highly variable process, imitation appears to play an important role in shaping spoken language. Phonetic imitation is defined as the unconscious and spontaneous adoption of speech characteristics of another talker. This phenomenon has been observed in both conversational speech as well as shadowing tasks in laboratories. Past research suggests that exposure to speech stimuli causes a listener to exhibit properties of the stimuli in their own productions (Babel, 2010; 2012; Goldinger, 1997, 1998; Honorof et al. 2011; Namy et al., 2002; Nielsen, 2011; Shockley et al., 2004). While the general pattern may be for imitation, there are considerable individual differences in imitation, some of which has been accounted for by social preferences and dialect background (Babel, 2010; 2012).

Kim et al. (2011) examined phonetic imitation within spontaneous conversations between interlocutors from distinct language and dialect backgrounds in an attempt to demonstrate bi-directional talker-listener accommodation as a potential device for long-term sound change. Crucially, this study inspected phonetic convergence under three conditions of target language distance: one within dialect/language condition, one cross-dialect condition, and one cross first-language condition (with speakers of Korean and English). These dyads completed a spot-the-difference task, and sentences from these conversations were used in an XAB perceptual judgment task to assess accommodation. Listeners (n=161) in Kim et al.’s study were instructed to choose which token, A or B (composed of early- and late-conversation recordings), sounded more like target X taken from recorded conversations. However, contrary to other AXB tasks used to assess imitation (Goldinger, 1998; Pardo, 2006) listeners were told that talkers (in the A and B productions) were attempting to impersonate the model’s productions (X).

Kim and colleagues found more imitation in the within dialect/language condition, suggesting that the closer two speakers are in terms of their phonetic repertoire, the more likely they are to converge or accommodate. Kim and colleagues, however, did not measure listener judgments of perceptual space or perceived similarity of the voices used in their experiment outside of the context of “impersonation”. The set of instructions offered by Kim and colleagues is potentially problematic in that when listeners are comparing A and B tokens to X tokens extracted from conversations, late-conversation tokens which do not sound similar enough to the model are then categorized as a instance of failed imitation, despite an increase in similarity.

The current study offers a more constrained similarity environment with recorded productions from a previous phonetic imitation study, described briefly below, in which all stimuli are derived from subjects who speak the same dialect/language. Furthermore, the current study aimed to compare general perceptual similarity of talkers with previous judgments of imitation in order to begin to map the relationship between phonetic distance and imitation by assessing how similarity between voices affects listeners’ judgments of similarity between shadowers and models.

ATTRACTIVENESS, TYPICALITY, AND PHONETIC SPACE

To examine the possibility that the similarity of a shadower’s voice to a model talker’s voice can influence perceived accommodation the current study draws on stimuli taken from a previous experiment. Babel, McGuire, Nicholls and Walters (2012) sought to examine whether social preferences or cognitive novelty triggered a stronger imitative response. To this end, we presented 20 participants (10 female, 10 male) with the voices from eight model talkers who had been previously rated as the most attractive, unattractive, typical, and atypical for each gender from a set of 60 voices (30 male, 30 female) in an auditory naming paradigm. Following previous work (Goldinger, 1998; Pardo, 2006), an AXB similarity rating task was used to quantify phonetic imitation. In such a task a shadower’s baseline and shadowed production for a word are compared to the model talker whose voice elicited the shadowed production: for example, book_{baseline} – book_{model} – book_{shadowed}. If listeners choose the shadowed token as more similar to the model, it suggests that a shadower’s voice became more like the model’s as a result of auditory exposure to the model (= imitation). If listeners choose the baseline token as more similar sounding to the model, then it suggests a shadower diverged from the model after auditory exposure to that model. One hundred and fifty-nine listeners completed the AXB task, and the results are summarized in Figure 1. The horizontal axis of this figure reports the proportion of shadowed tokens (as opposed to baseline tokens) judged as more similar to the model. Values above 0.5 indicate imitation, and those below 0.5 would indicate phonetic divergence. Values hovering around 0.5 indicate that listeners were at chance in terms of assessing whether baseline or shadowed tokens sounded
more similar to the model. As can be gathered from this figure, listeners’ judgments suggest that female participants generally imitated more than male participants, but that this is subject to variation in voice type.

**FIGURE 1.** Model by Shadower Gender interaction. Pink bars represent the female shadowers and blue bars represent the male shadowers.

We compared listeners’ judgments of imitation from this previous study to the similarity judgments collected in the study described below. To reiterate, the goal of the current study was to examine how the perceptual distance between shadowers’ baseline productions and the tokens from the model talkers may contribute to listeners’ judgments of imitation. Contrary to Kim et al. (2011) and following discussions in Babel (2009, 2012), we predicted that voices which are more different will be judged as imitating more.

**SIMILARITY RATING TASK**

In this experiment we examined how perceived phonetic similarity affects listeners’ judgments of imitation. The similarity rating task was designed to allow listeners to scale their perceptual space of the tokens to which they were exposed using a digital space displayed on a computer screen in front of them. The mapping of similarity directly into a constrained space allows the listeners the freedom to express how similar or different two tokens sound relative to one another while also providing a continuum of perceptual judgments to compare to perceived imitation judgments from Babel et al. (2012).

**Methods**

**Participants**

Twenty-four participants (15 = female, 9 = male) from the University of British Columbia community completed the task. All participants were self-identified native speakers of English with no reported speech, language, or hearing disorders. They were compensated with $10CAD.
Stimuli

Stimuli consisted of the baseline productions from 20 shadowers (10 females, 10 males) who completed an auditory naming task (Babel, McGuire, Nicholls, & Walters, 2012). The model talkers were 8 voices (4 male, 4 female) who had been previously rated as the Most and Least Attractive and Most and Least Typical voices for each gender from a set of 60 voices (30 male, 30 female). The words were 15 low frequency monosyllabic words taken from the CELEX database (Baayen et al., 1993) which had /i u/ as syllable nuclei (5 words per vowel). To be clear, the shadowers’ tokens used in this study are the baseline tokens which were recorded prior to exposure to any of the model talkers.

Procedure

Listeners were seated at computer workstations equipped with a mouse and a AKG K240 headphones. In each trial participants were presented with a shadower’s baseline production and a model talker token of the same lexical item. The ISI between tokens was 500 ms. Each listener was presented with 625 randomly selected trials out of the total possibility of 2400 trials (15 words x 8 models x 20 shadowers).

Similarity was measured on a pixilated visual analogue scale (Massaro and Cohen, 1983). Listeners were instructed to judge two productions of the same word with respect to one another and record their judgments using the VAS line. On each trial listeners were to use the mouse and click along the line, as shown in Figure 2, indicating how similar the shadower and model voices were. The task lasted on average 45 minutes and participants were instructed to take breaks in between blocks as needed. Post-task procedures included a language background questionnaire and compensation transactions.

Figure 2. The Visual Analogue Scale (VAS) participants used to rate the similarity between shadowers’ baseline tokens and model talkers’ tokens. The center of the line is approximately 800, which corresponds to the vertical black lines in the subsequent figures.

Results and Analysis

Figure 3 presents the averaged VAS similarity ratings for each shadowers’ voice compared to each of the model talkers. The similarity ratings for the 10 female shadowers’ voices are plotted in pink, and those for the 10 male shadowers’ voices are shown in blue. Higher VAS ratings indicate more similar-sounding voices. These density plots demonstrate several points. First, shadowers’ voices did indeed differ in terms of perceived similarity, as evidenced by the multi-modal distributions. Second, these differences in perceived similarity varied across comparisons with the different model talkers. Finally, there were gender differences based around the gender of the shadower voice; female shadowers’ voices are rated as more similar to the female model talkers, and male shadowers’ voices are rated as more similar to the male model talkers. Moreover, female’s voices are generally rated as more similar to the models with generally higher VAS ratings, although the degree to which this is true varies by model.
The relationship between shadowers’ averaged VAS similarity ratings and their averaged proportion of shadowed tokens judged more similar to the model (taken from the previous AXB task) were analyzed using Pearson’s correlation. There was a significant correlation across the entire data set \( t(157) = -3.52, p < 0.001, r = -0.27 \), indicating that voices that are rated as more different are also judged by listeners as having imitated more. Further exploration of the data revealed that this pattern was robust for female voices \( t(78) = -4.49, p < 0.001, r = -0.45 \), but non-existent for male voices \( t(77) = -0.67, p = ns, r = -0.08 \). This is plotted in Figure 4.
FIGURE 4. VAS similarity ratings by imitation judgments for each shadower. Females shadowers are in red, and males are in blue. The horizontal axis indicates VAS similarity ratings for each shadower compared to each model; the vertical line at 800 indicates the approximate midpoint of the VAS scale. The vertical axis indicates the proportion of shadowed tokens judged more similar to models’ productions compared to the baselines; the horizontal line at 0.5 marks chance performance for listeners’ judgments of imitation. Values above 0.5 indicate shadowed tokens were more likely to be judged as more similar to the model, and those below 0.5 indicate baseline tokens sounded more similar to the model – this behavior would be indicative of divergence.

Figure 5 separates the data for each model talker. While we refrain from statistical analysis due to the small sample sizes when separating the data by shadower gender and model, the plots indicate that the pattern for female shadowers is generally robust across model talkers, with the exception of the Least Typical Male model. This voice elicited strong amounts of imitation for female voices, and perceived similarity of the voice plays no role in listeners’ judgments of imitation. Inspection of Figure 5 with the male shadowers in mind reveals more variable patterns: with some models there seems to be a negative relationship such that more dissimilar voices are judged to have imitated more (Least Attractive Male, Least Typical Female), with others there is a tendency for more dissimilar voices to being judged as having diverged (Most Typical Female and Most Typical Male), while with others the amount of perceived imitation has no bearing on perceived voice similarity.
FIGURE 5. VAS similarity ratings by imitation judgments for each shadower separated by model talker. Females shadowers are in red, and males are in blue. The horizontal axis indicates VAS similarity ratings for each shadower compared to each model; the vertical line at 800 indicates the approximate midpoint of the VAS scale. The vertical axis indicates the proportion of shadowed tokens judged more similar to models’ productions compared to the baselines; the horizontal line at 0.5 marks chance performance for listeners’ judgments of imitation. Values above 0.5 indicate shadowed tokens were more likely to be judged as more similar to the model, and those below 0.5 indicate baseline tokens sounded more similar to the model – this behavior would be indicative of divergence.

DISCUSSION AND CONCLUSION

The results of this study suggest that, contrary to the findings of Kim et al. (2011), perceptual similarity is negatively correlated with perceived imitation. That is, the more dissimilar a shadower was to a model, the more likely she was judged as having imitated. While this pattern is recognized as statistically significant across the data set, further investigation revealed that the pattern was robust among female talkers, but not among males. The negative correlation between perceptual similarity and perceptual imitation reinforces the possibility of greater phonetic convergence between talkers with greater phonetic distance between them. Such a pattern is indicative of the established perceptual flexibility in relation to perceptual speech adaptability and its contribution to dialect variation and sound change. Given the finding from Babel et al. (2012) that women exhibit more spontaneous phonetic imitation than males and the evidence from the current study which suggests that females with more dissimilar voices in comparison to the model talker imitate more, we are led to consider the importance of the phonetic perceptual flexibility and production of women in phonetic innovation and variation.

An important caveat is necessary. The results of this study are inconclusive as to whether increase phonetic distance truly gives way to increased phonetic imitation or whether imitation involving more dissimilar voices is easier for listeners to assess. Future research will attempt to tease apart these two possibilities.
ACKNOWLEDGMENTS

Thanks to Graham Haber and Jamie Russell for lab assistance on this project. Parts of this work were funded by SSHRC grant #F10-04808 awarded to Molly Babel.

REFERENCES