PHONETIC CONVERGENCE AFTER PERCEPTUAL EXPOSURE TO NATIVE AND NONNATIVE SPEECH: PRELIMINARY FINDINGS BASED ON FINE-GRAINED ACOUSTIC-PHONETIC MEASUREMENT

Midam Kim

Northwestern University, USA
midam-kim@northwestern.edu

ABSTRACT

This study investigates phonetic convergence by native English speakers after exposure to speech by a native or a nonnative speaker of English. Participants 1) read two word sets, 2) were exposed to one of the word sets either through auditory (experimental group 1 & 2) or visual inputs (control groups 1 & 2), and 3) read both word sets again. Preliminary results showed a tendency for participants to converge towards nonnative models but not towards native models, and to show no specificity of phonetic convergence patterns between the word sets they heard and the word sets did not hear.

Keywords: phonetic convergence, interlocutor language distance, generalization

1. INTRODUCTION

When we are exposed to speech that may deviate substantially from our own speech, we might modify our own speech accordingly. The present study is part of a larger research project which examines native English speakers' phonetic modifications after perceptual exposure to native and nonnative readings of English words and sentences. This paper focuses on exposure-induced modification in monosyllabic words.

Previous work found that native and nonnative English speakers converged towards their partners in the course of spontaneous conversation, and the degree of phonetic convergence was mediated by the interlocutors' language distance, namely, their sharing of dialects and native status [9]. That is, speakers tended to converge more towards an interlocutor whose language background was closer to their own. This is in line with findings in the literature that phonetic convergence is influenced by various linguistic and social factors such as speakers' attitude towards the model [2, 7], speakers' gender and conversational roles [11, 12]. However, in the previous work [9], because the speech samples were taken from spontaneous conversational data and all differed in content, fine-grained acoustic analyses to assess speakers' phonetic convergence patterns could not be conducted. Instead, perceptual similarity tests where the early and late samples of a conversation were compared for a better match to the partner's speech samples were used. Perceptual similarity tests are useful in that they provide holistic judgments on phonetic accommodation taking into account all acoustic-phonetic dimensions [8, 11, 12]. In contrast, parametric acoustic measurements on specific segments cannot easily capture crucial parameter combinations. However, in order to track down how perceived accommodation patterns are actually realized at the phonetic level, acoustic measurements on specific phonetic features are essential.

In the current study, the same research question as in the previous work [9] is investigated, that is, is phonetic convergence negatively correlated with interlocutor language distance? However, in the current study, instead of participating in spontaneous conversations, native English speakers heard a native or nonnative English model speaker reading words. A control group was exposed to written rather than spoken words. The participants also read the same words before and after the exposure phase. In this way, participants' pretest and posttest utterances with fine-grained acoustic measurements could be directly compared to show which was closer to their model speaker's utterances. This would allow us to investigate phonetic convergence on a rigorous acoustic basis, see also [2, 6, 10, 12]. In this paper, specifically, results are reported based on duration of the initial consonant and vowel of monosyllabic words at pretest and at posttest relative to the model's productions as the index of phonetic convergence.

An additional research question was added to this paper, that is, can speakers generalize their phonetic convergence patterns to unexposed items?

To incorporate this into the experiment, two word sets were established, and only one of the two sets was exposed to the participants during the exposure phase. Then, in their pretest and posttest phases, they read both sets.

Based on the results from the previous work [9], it was predicted that participants would converge more towards a fellow native English model speaker than towards a nonnative English model speaker. Additionally, based on the results from [10], it was predicted that participants would generalize their phonetic convergence patterns to new items.
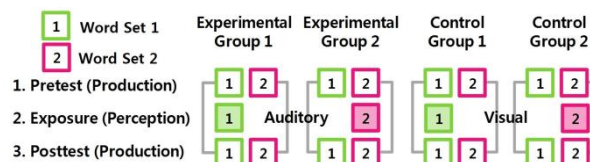
2. EXPERIMENT

2.1. Method

2.1.1. Materials

To test if phonetic accommodation effects are transferred to unexposed items by participants, two sets of English monosyllabic words (n = 63) were established, considering the following conditions: 1) The two sets differ in the place of articulation of the initial consonant. In Set 1, the words start with a bilabial stop, and in Set 2, with an alveolar stop. 2) In each word set, half of the words have voiced initial stops, and the other half, voiceless initial stops. 3) Likewise, in each word set, half of the words have voiced final consonant, and the other half of the words, voiceless final consonant. 4) The vowels, /æ, ɛ, i, ɪ, ɑ, ʌ, u, ʊ/, were controlled to be the same over the two sets. 5) Following [6]'s finding that only low frequency words were successfully imitated, the criterion for word frequency was set to be under 30 per million words in SUBTLEXus [9]. Most of the words chosen (90%) fulfilled this condition, while the highest frequency of the other words was 76 per million words. Considering all these conditions, 32 words were chosen for Set 1 (words with bilabial initial stops), and 31 words were chosen for Set 2 (words with alveolar initial stops).

Two female monolingual native American-English speakers and two female nonnative English speakers whose native languages were Korean were recorded reading all words in random order in a sound booth. The recordings were made to a computer with the sampling rate of 48000 Hz. The recordings were used as model speech stimuli in the perception phase for participants.

2.1.2. Procedure

Two experimental groups and two control groups were tested for generalization effects on unexposed items using Set 1 and Set 2 (Figure 1). All participants followed three phases: 1) pretest production, 2) auditory or visual exposure, and 3) posttest production. 1) In the pretest, participants in all conditions were recorded reading Set 1 and Set 2 out loud. 2) During the exposure phase, participants in experimental conditions heard only one of Set 1 or Set 2, read by either a native or nonnative model speaker, with 9 repetitions of each word in random order. The inter-sample interval was 100 ms. On each trial, the participants heard a word and selected the critical item written in standard English orthography on a computer display that included the target item plus seven alternatives. This item-identifying task was intended to encourage participants to focus on listening to the stimuli, but they were not given any direct task training or any feedback. Participants in the control conditions viewed orthographic representations of 9 repetitions of words taken from either Set 1 or Set 2, and did the same item-identifying task during the exposure phase, with no auditory stimulation. 3) In the posttest, all participants in the four experimental conditions read all words in Set 1 and Set 2 again. In all reading phases, the words were displayed to the participants on the computer monitor in random order, and all readings were recorded to another computer with the sampling rate of 48000 Hz. By comparing pre-to-post differences across experimental and control groups, we could test whether phonetic convergence occurred in response to auditory exposure to a model speaker. Additionally, we could also test generalization of convergence to unexposed items.

Figure 1: Schematic description on the experiment procedure for each experimental condition.

2.1.3. Participants

Fifty female monolingual native American-English speakers participated in the experiment with normal speech and hearing. Out of the fifty participants, five groups of ten participants were

randomly assigned to each of the four model speaker conditions and to the control condition. These groups of ten were then each sub-divided into two groups with five participants in each group (a Group 1 and a Group 2 as shown in Fig 1 above). In total, there were 1x2 control groups and 4 x 2 experimental groups (a Group 1 and a Group 2 for each of the four model speakers, of which 2 were native and 2 were nonnative English speakers).

2.1.4. Analyses

Praat was used for acoustic analyses on the monosyllabic words read by the model speakers and participants (pretest and posttest readings). With the word recordings, durations were measured from the burst of the initial consonant until the formant structure of the vowel ended (CV duration, henceforth).

The data were submitted to a linear mixed effects regression model [1, 4] with CV duration as the dependent measure. Phonetic convergence was assessed by comparing effects from experimental groups to effects from control groups; if the difference between pretest and posttest readings is significantly larger in experimental groups than in control groups, this indicates phonetic change, whether in a positive direction (convergence towards the model value) or in a negative direction (divergence away from the model value).

Specifically, the fixed effect factors were timing (model speakers, pretest, and posttest), experiment conditions (control condition, native model speaker, nonnative model speaker), exposure condition (Set 1, Set 2), and word set (Set 1, Set 2). For unexposed items during the exposure phase (either Set 1 or Set 2 in experimental groups and both Set 1 and Set 2 in control conditions), the model speaker level of the timing factor was filled with pretest level values. This decision was done to conduct a unified model on the total dataset with all fixed effect factors. The reference level for timing was pretest; model speaker values and posttest values were each compared to pretest values. The reference level for experimental condition was control conditions, so that each of native model speaker condition and nonnative model speaker condition were compared to control conditions. The reference level for exposure condition and word sets were both Set 2. Interactions of all fixed effect factors were also included to the model. Participants, words, and model speakers (two natives and two nonnatives) were included as random effect factors.

2.2. Results

None of the fixed effect factors, timing, experimental conditions, exposure condition, and word sets, had significant main effects. The critical interactions to assess phonetic convergence are interactions with timing and experiment condition. Figure 2 summarizes the critical results. The interaction of experiment condition and timing was significant. Specifically, when the model speaker was a native English speaker, participants maintained their CV durations after the exposure phase as much as control participants ($\widehat{\beta} = -6.91$, $SE = 4.86$, $t = -1.42$, $p = 0.15$), and their pretest readings were marginally different from the model speaker readings ($\widehat{\beta} = -9.02$, $SE = 4.86$, $t = -1.85$, $p = 0.06$). In contrast, when the model speaker was a nonnative speaker of English, participants reduced their CV durations after exposure marginally more than control participants ($\widehat{\beta} = -8.79$, $SE = 4.60$, $t = -1.91$, $p = 0.056$). Because model speech values were significantly smaller than pretest values in nonnative model speaker conditions ($\widehat{\beta} = -19.37$, $SE = 4.6$, $t = -4.20$, $p < 0.01$), we can see that the change after exposure in the nonnative model speaker condition was phonetic convergence.

Figure 2: Duration of the initial consonant and vowel (CV duration) of monosyllabic words, spoken by the model speakers (dark grey bars) and participants in the pretest (white bars) and posttest (light grey bars) recordings in each experimental condition (control conditions, native model speaker, and nonnative model speakers). Error bars depict 95 % confidence intervals. ** p < 0.01, * p < 0.1.
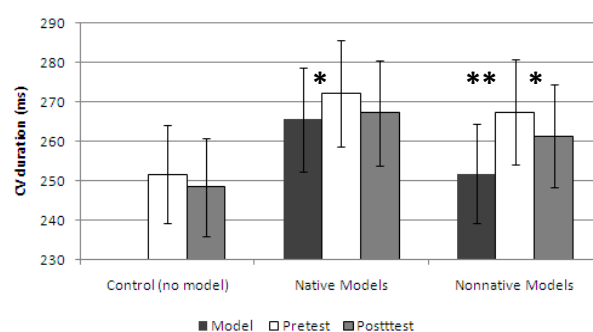
The interaction among timing, experiment condition, and exposure condition was not significant. Also, the interaction among timing, experiment condition, and word set was not significant. This means that neither of the two exposure conditions and the two word sets differed

in their effect on the critical phonetic convergence patterns described above.

Finally, there was no interaction among timing, experiment condition, exposure condition, and word sets. This indicates that the phonetic change found after exposure to Set 1 did not appear differently on Set 1 and Set 2. In other words, participants generally applied the changes they made on exposed items to unexposed items during the exposure phase.

3. GENERAL DISCUSSION

These results differ from the previous work which showed that phonetic convergence is facilitated by closer interlocutor language distance [9]. In terms of one specific acoustic-phonetic measurement, CV duration, participants in the present study showed larger phonetic convergence towards nonnative model speakers than towards native model speakers. There might be two reasons for this discrepancy between studies. First, phonetic convergence patterns might occur differently in spontaneous conversations and after perceptual exposure to pre-recorded read speech. Second, the observance unit might matter; phonetic convergence observed by holistic perceptual judgments on phrases might pattern differently from phonetic convergence observed by fine-grained acoustic measurements on CV durations of monosyllabic words.

The results support the second prediction that phonetic convergence patterns to exposed items would generalize to unexposed items. This finding is in line with the previous finding in the literature that speakers generalized their VOT imitation on a bilabial stop to a velar stop [10].

We note that all significant phonetic changes between pretest and posttest readings observed in the present study were duration reduction. Because the model speech samples were not significantly longer than the pretest speech samples, we cannot exclude the possibility that the phonetic changes found in this study might be an effect of second mention reduction [3]. Ongoing analyses on the data of disyllabic words and sentences might help resolve these questions.

4. REFERENCES

[1] Baayen, R.H. 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/web/packages/languageR/*

[2] Babel, M.E. 2009. *Phonetic and Social Selectivity in Speech Accommodation*. Dissertation of University of California, Berkeley.

[3] Baker, R.E., Bradlow, A.R. 2009. Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech* 52(4), 391-413.

[4] Bates, D., Martin M., Ben, B. 2010. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-35. *http://cran.r-project.org/web/packages/lme4/*

[5] Brysbaert, M., New, B. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41(4), 977-990.

[6] Delvaux, V., Soquet, A. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64, 145-173.

[7] Giles, H. 1973. Accent mobility: A model and some data. *Anthropological Linguistics* 15, 87-109.

[8] Goldinger, S.D., Azuma, T. 2004. Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review* 11, 716-722.

[9] Kim, M., Horton, W., Bradlow, A.R. In press. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*.

[10] Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *J. Phonetics*. doi:10.1016/j.wocn.2010.12.007

[11] Pardo, J.S. 2006. On phonetic convergence during conversational interaction. *J. Acoust. Soc. Am.* 119, 2382-2393.

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