ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

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

2aSC43. Phonetic accommodation in Spanish-English and Korean-English bilinguals
Stephen Tobin*

*Corresponding author's address: Psychology, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, Connecticut 06269-1020, stephen.tobin@uconn.edu

Preliminary results from a cross-linguistic investigation of phonetic accommodation in speech production and perception are presented. The finding that synchronous actions are more stable than asynchronous ones has been reported in studies of general (Kelso, 1981) and speech-specific (Browman & Goldstein, 1992; Byrd, Tobin, Bresch & Narayanan, 2009) motor control. With reference to glottal-oral timing, near-zero VOTs (voice onset times) are representative of near-synchronous timing, whereas long-lag VOTs are representative of asynchronous timing (Sawashima & Hirose, 1980; Dixit, 1984; Löfqvist & Yoshioka, 1989; Fuchs, 2005). These observations served as a basis for the prediction that native speakers of Korean, with its long-lag aspirated stops (~120 ms), would more readily accommodate to typical English voiceless stop VOT (~70 ms) than native speakers of Spanish, with its short-lag voiceless stops (~20 ms). Spanish-English and Korean-English bilinguals were recorded reading voiceless stop-initial English words, before and during a task in which participants shadowed recorded productions of a native speaker of American English. Preliminary analysis of the production data provides some support for these hypotheses. The results contribute to our understanding of the conditions that promote phonetic accommodation.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Recent studies of phonetic accommodation, that is, perceptually guided changes in speech categories, have to a great extent focused on social and cognitive influences on the extent and direction of accommodation (Babel, 2012; Kim, 2011; Nielsen, 2011; Abrego-Collier et al., 2011). In this study, we consider the effects that the stability or instability of articulatory timing might have on speaker-hearers’ patterns of accommodation.

One line of research on speech articulation has provided evidence that vocal tract gestures that are planned to occur synchronously are more stable than those that are planned to occur asynchronously (Brown & Goldstein, 1992; Nam, Goldstein & Saltzman, 2009). At the syllable level, the constituent gestures of onset consonants have been shown to be produced in a more tightly coordinated fashion than the constituent gestures of coda consonants (Byrd, Tobin, Bresch & Narayan, 2009). Although these patterns have been observed in speech articulation, they are not specific to speech. The finding that synchronous actions are more stable than asynchronous ones has been reported in studies of general motor control. A series of studies on finger-tapping (Kelso, 1981; Haken, Kelso, Bunz, 1985) and leg-swinging (Schmidt, Carello, Turvey, 1990) have revealed strong preferences for particular patterns of coordination, the strongest preference, and most stable pattern is synchronous, or in-phase, timing. Both fingers, legs, or articulators progress through their cyclical movements approximately synchronously.

With reference to oral-laryngeal timing in the articulation of stop consonants, near-synchronous timing patterns would yield near-zero voice onset times (VOTs), whereas a large degree of asynchronicity between oral and laryngeal gestures would yield longer positive or negative VOTs. Several articulatory investigations of oral-laryngeal timing provide support for this assertion. In his photoelectric glottographic investigation of oral-laryngeal timing in Hindi, Dixit's (1984) shows that in the voiceless unaspirated bilabial stop, [p], peak glottal opening occurs during the oral closure. This is quite consistent with the synchronicity, or in-phase timing, of peak labial closure and peak glottal opening. In the voiceless aspirated bilabial stop [pʰ], however, peak glottal opening occurs at the end of the oral closure. In this case we observe asynchronicity between maximal labial constriction and the subsequent peak glottal opening. Löfqvist & Yoshioka (1984) report that the timing and duration of American English speakers' voiceless stops and fricatives remains relatively constant across speaking rates. While this investigation does not provide direct evidence for the relation I propose between articulatory phase timing and VOT, their results are consistent with it. Likewise, although Sawashima & Hirose (1983) and Fuchs (2005) do not provide direct evidence for the relation either, the cross-linguistic patterns of oral-laryngeal timing that they report are consistent with it.

Based on these observations, I predict that, given the long-lag aspirated stops (~120 ms) of Korean, Korean-English bilinguals will more readily converge towards typical English voiceless stop VOTs (~70 ms) than Spanish-English bilinguals, given their short-lag voiceless stops (~20 ms). An experiment was designed to test this hypothesis. Although the stops of Korean are divided into three categories of glottal behavior (tense, lax and aspirated), any of which could, in principle correspond to English voiceless stops in an investigation of the mapping between Korean speakers' Korean and English perceptual categories, Park and DeJong (2008) found that participants reliably identified English voiceless stops ([pʰ] and [tʰ]) as their Korean aspirated counterparts.

METHOD

Participants

Ten Spanish-English bilinguals and ten Korean-English bilinguals participated in the experiment (N=20). Participants were recruited from the University of Connecticut, Haskins Laboratories and Yale University international communities. Participants received course credit or $10 per hour of participation.

Stimuli

Forty monosyllabic [k]-initial words of English were selected and recorded by a female native speaker of American English in Haskins Laboratories’ anechoic chamber. The mean voice onset time (henceforth VOT) of these tokens was 110.45 ms (SD = 13.73).
Procedure

Participants began with a baseline word-reading task. Ten repetitions of the series of 40 monosyllabic English words were presented to participants on a computer screen using E-Prime software (Schneider, Eschman & Zuccolotto, 2002), yielding 400 word tokens per participant. A new word appeared on the screen every two seconds. In all of the experimental tasks, the program paused after every 22 trials and participants pressed a key to continue. Participants took a short break after this task.

In the second part of the experiment, participants were recorded completing a series of interleaved word-shadowing and test word-reading tasks. The word-shadowing task was intended to induce accommodation, and the test word-reading task was intended to provide a measure of the induced accommodation that was directly comparable to the baseline word-reading task. In the word-shadowing task, two randomized repetitions of the 40 English monosyllabic stimulus words were presented to participants over headphones. Participants were instructed to repeat each word they heard as quickly and as accurately as possible. After each word-shadowing task, participants completed a test word-reading task, in which two randomized repetitions of the 40 English monosyllabic words were presented to participants. Each participant completed five repetitions of this pair of tasks, yielding 400 tokens from the test word-reading task. The word-shadowing and test word-reading tasks were interleaved in order to quickly measure the effects of accommodation before they had subsided. Finally, participants completed the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld & Kaushanskaya, 2007).

Acoustic Analysis

Audio files of each participant's baseline and word-reading tasks were divided into individual word segments (20 x 800 = 1600) using Audacity and the voice onset time of each word was measured using Wavesurfer (Sjölander & Beskow, 2000). The words were subsequently concatenated using Praat (Boersma & Weenink, 2009) and submitted to the UPenn Forced Aligner (Yuan & Liberman, 2008) in order to obtain durations of the vowel portions of the audio words for an additional analysis.

RESULTS

The VOT data were submitted to a 2 x 2 repeated-measures ANOVA with Language (Korean, Spanish) as a between-subjects factor and Task (Baseline, Test) as a within-subjects factor. Given that there was considerable variation in participants' experience living in an English-language community, we included the number of years participants had spent in the US (Mean = 8.90, SD = 6.59) as a covariate. The analysis yielded a significant interaction of Language x Task (F(1,17) = 4.52, p < .05). A Fisher's LSD post hoc analysis revealed a significant effect of Task on the Korean VOTs (Baseline = 80.50 ms, Test = 74.43 ms (p < .05)) but not on the Spanish VOTs (Baseline = 60.93 ms, 62.27 ms; see Figure 1).

I observed that the model speaker's VOTs were, in fact, longer than those of both the Spanish and the Korean groups' baseline VOTs (Mean = 110.45 ms, SD = 13.73), such that convergence towards the target would be instantiated by longer VOTs in the test condition. I attribute these longer than typical VOTs to overall consonant lengthening, which is a characteristic of clear speech (Smiljanic & Bradlow, 2008).

In contrast to the expected pattern of convergence, the significant change in the Korean group's VOTs was a decrease of 6.07 ms. One possible explanation for this decrease in VOT is that the duration of each whole word may have decreased in subsequent repetitions after the initial utterance of the word (Fowler & Housum, 1987). If whole word durations decreased over the course of the task, this might mask an increase in VOTs relative to the rest of the word.

In order to obtain a measure of VOT relative to the rest of the word, each VOT was divided by the duration of the following vowel, as measured by the Penn Phonetics Lab Forced Aligner (Yuan & Liberman, 2008). An additional 2 x 2 repeated-measures ANOVA was run with the same design as the first, but with the dependent measure of VOT/Vowel Duration. This analysis also yielded a significant interaction of Language x Task (F(1,17) = 8.27, p < .05). Again, a Fisher's LSD post hoc analysis revealed a significant effect of Task on the Korean VOT/Vowel ratios (Baseline = 0.56, Test = 0.46 (p < .05)) but not on the Spanish VOT/Vowel ratios (Baseline = 0.39, Test = 0.37; see Figure 2). As with the raw VOTs, the model speaker's VOT/Vowel ratios were greater (Mean = 0.58, SD = 0.22) than those of the Spanish and the Korean groups. This confirms the VOT shortening observed in the initial ANOVA conducted with raw VOTs.
FIGURE 1. Effect of shadowing task on VOTs of [k]-initial words read aloud.

FIGURE 2. Effect of shadowing task on VOT/Vowel ratios of [k]-initial words read aloud.
DISCUSSION

The observed results show a pattern of accommodation in which Korean-English bilinguals' VOTs diverge from the longer VOTs of a model speaker. Although studies of phonetic accommodation have more typically reported convergence, a number of studies have reported divergence. Kim, Horton and Bradlow (2011) found trends in their data suggesting that non-native speaker interlocutors with very heavy or very light accents were more likely to diverge away from a native speaker than those in the intermediate range. They also found some indication that interlocutors who are speakers of different dialects of a language may diverge away from each other.

This significant reduction in VOT of the Korean-English bilingual group is also consistent with the original hypothesis. It was claimed that the attractive force of the stable oral-laryngeal coordination pattern of zero VOT and that longer VOTs would reduce more readily than shorter VOTs increase. The Spanish-English bilingual group's VOTs may have been sufficiently close to the stable pattern of near-zero VOT to inhibit phonetic accommodation.

Further analysis of individual participant level patterns and data from the LEAP-Q (Marian, Blumenfeld & Kaushanskaya, 2007) may yield further insight into the specific causes of accommodation. Although neither group converged towards the model speaker's VOTs, some of the individuals did.

The question of whether shadowing of English VOTs falling between the Spanish-English bilingual (~60 ms) and Korean-English bilingual (~75-80 ms) ranges would yield greater accommodation from Korean-English bilinguals, remains open.

ACKNOWLEDGMENTS

I would like to thank Carol Fowler, Emily Myers, Letty Naigles, Rachel Theodore, Hosung Nam, Eliot Saltzman, Christian DiCanio, Mark Tiede, Tine Mooshammer and Jelena Krivokapic for guidance and feedback on this project.

REFERENCES


Fuchs, S. (2005). “Articulatory correlates of the voicing contrast in alveolar obstructed production in German.” ZASPIL.


rhythmic movements between people.” JEP.HPP 16, 227-247.