A DYNAMIC APPROACH TO PHONETIC CHANGE

Stephen Tobin

Department Linguistik, Universität Potsdam
tobin@uni-potsdam.de

ABSTRACT

Investigations into phonetic accommodation reveal that convergence in interlocutors' phonetic forms is typical. However, there is considerable between-speaker variation in the extent of this convergence. In this investigation, I propose that some of this variation can be attributed to language-specific differences in articulatory timing stability (Voice-Onset-Time) and that a dynamical systems approach gives a better account of such findings than extant theories/models. Spanish-English and Korean-English bilinguals, and monolingual English controls completed word shadowing tasks to induce phonetic accommodation and word reading tasks to measure accommodation. The results confirm that between-language differences in articulatory timing stability influence the likelihood of phonetic change and that a dynamical approach provides the best account of the effect.

Keywords: Articulatory Timing, Bilingualism, Dynamical System, Phonetic Accommodation, VOT.

1. INTRODUCTION

Phonetic accommodation is a pattern of change in speech, whereby speaker-hearers' production patterns are influenced by those of their interlocutor(s) or ambient language (cf. [2], [12], [15], [20], [25]). Typically, speech patterns are reported to become more similar, but there is considerable between-speaker variation in the extent of this convergence (cf. [21]). The claims of this investigation are twofold—first, that some of the between-speaker variation can be attributed to language-specific patterns of articulatory timing, and second, that these patterns of phonetic change are best accounted for within a dynamical systems approach.

With respect to patterns of articulatory timing, the observation that pairs of actions performed synchronously are easier to execute and to maintain than those performed with some particular asynchronous phase-timing is well established in the literature on motor coordination (cf. [1], [14]) and in speech specific investigations (cf. [6], [7]). A number of investigations suggest that this observation also applies to VOT (cf. [8], [10], [19], [26]). In the context of VOT, the relevant actions are laryngeal abduction and oral constriction—near-zero VOTs result from near-synchronous timing of oral and laryngeal gestures, while long VOTs result from asynchronous timing.

Spanish voiceless stops exemplify the near-synchronous timing pattern (VOT ~20ms), while Korean aspirated stops exemplify the asynchronous timing pattern (~120ms). Because the Korean VOTs demonstrate a less stable pattern of coordination than the Spanish VOTs, I hypothesized that Korean-English bilinguals would more readily converge towards intermediate monolingual English VOT (~70ms) than native Spanish speakers (cf. [18]).

The observations upon which these hypotheses are based are inherently dynamical. A dynamical system consists of a vector describing the current state of the system and a function whereby future states of the system are predicted (cf. [27]). In a geometric map of a dynamical system, variations in the surface of the map represent variations in the stability and attractive or repelling force of the various states of the system. Thus, stable areas/points in the system (states in which little to no change is likely) appear as flat surfaces or as basins of attraction, representing states towards which the system is likely to be drawn. Likewise, repellers appear as hills, representing states from which the system is likely to be repulsed, the depth of the basin or the height of the hill reflecting the strength of the force or likelihood of change. In this context, zero VOT constitutes a stable attractor basin, while longer VOTs constitute unstable points on a slope leading towards the zero-VOT attractor basin.

Existing theories of speech category learning account for various aspects of between-language phonetic influence but none of them would predict the language-specific patterns of accommodation that the dynamical approach predicts. Flege's Speech Learning Model (cf. [9]) accounts for influences between the phonetic categories of L2 learners' first and second languages because in this model both sets of phonetic categories share the same acoustic-phonetic space. However, the model would not predict any particular difference between Korean- and Spanish-speakers' accommodation towards English voiceless stop VOT. Kuhl's Native Language Magnet (NLM, cf. [16], [17]) and Best's Perceptual Assimilation Model (PAM, cf. [4], [5]) both focus on perception and as such they do not make particular

predictions about production, though PAM was conceived with articulatory constraints in mind. The explicit representation of phonetic variables in exemplar approaches (cf. [13], [22], [23]) closely resembles that of a dynamical system. However, in an exemplar approach the dynamic forces of articulatory constraints would have to be stipulated rather than being an inherent element of the model, as is the case in a dynamical approach.

2. METHOD

2.1. Participants

Ten bilingual speakers of Spanish and English (henceforth *Spanish group*), 10 bilingual speakers of Korean and English (henceforth *Korean group*), and 10 monolingual native speakers of American English (henceforth *English group*) participated in the experiment. All participants were recruited from the University of Connecticut, Haskins Laboratories and Yale University. Participants were compensated with course credit or $10 for each hour of their participation.

2.2. Stimuli

Forty [k]-initial monosyllabic words of English (e.g., *keen, cab, cot, coat*) were selected for presentation to participants. The vowels of the set of words were evenly distributed among five areas of the vowel quadrilateral (high front, low front, low back high back and mid central). The words were presented visually as part of a word reading task and auditorily, as recorded by a female native speaker of American English in Haskins Laboratories' anechoic chamber.

2.3. Procedure

Participants completed the informed consent process and then completed a series of reading and shadowing tasks. Baseline and test reading tasks were included so that any task-specific confounds could be excluded. A shadowing task was used to induce phonetic change. The shadowing task was divided into blocks and interleaved with blocks of the test reading task in order to maximize the potential influence of the shadowing task on the test task VOTs.

2.3.1. Baseline word reading

Participants completed the informed consent process and were instructed to read aloud words presented on a monitor. Ten randomized repetitions of the 40 stimulus words were elicited, yielding 400 baseline tokens. Participants were recorded with a portable solid state sound recorder (Marantz PMD 661). The mean VOT of the stimuli was subsequently found to be longer than expected in conversation speech (M=110.5ms, SD=13.7) because the speaker was using clear speech.

2.3.2. Word shadowing

Participants were then instructed to repeat words they heard over headphones as quickly and as accurately as possible. Two randomized repetitions of the 40 words were presented in five separate blocks, which were interleaved with the five blocks of the test word reading block. These two tasks were interleaved in order to maximize the likelihood that the accommodation induced during shadowing would still be present during the test word reading phase.

2.3.3. Test word reading

The test word reading task was identical to the baseline word reading task, except that the task was divided into five blocks of two randomized repetitions, and these five blocks were interleaved with the word shadowing task. Thus, participants produced 400 test word tokens.

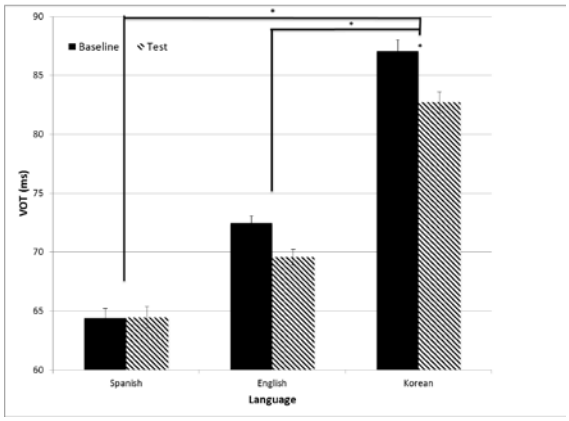
2.3.4. Acoustic measurements

Audio files were saved to a computer, segmented into individual word files and then VOT measurements were made. Seventy percent of the VOTs (16,800) were measured by the author and the remaining 30% (7,200) were measured by undergraduate research assistants. The author measured an additional 10% (2,400) of the tokens measured by research assistants to test for reliability. An intra-class correlation analysis yielded a coefficient (ICC) of 0.96.

3. RESULTS

Analysis of the data was conducted using linear mixed models (LMMs) in the R statistical environment. LMMs allow for the inclusion of both fixed (intended manipulations) and random (not manipulated or experimentally controlled) effects in the analysis (cf. [3]). Native Language (Spanish, English, Korean) and Task (Baseline, Test) were the fixed effects in this analysis, while Word, Participant and the Participant x Task interaction were included as random effects. A significant effect of Native Language obtained—the Korean group's VOTs were longer than those of the English ($t(29.96)=2.613$, $p<.05$) and Spanish ($t(29.96)=4.062$, $p<.05$) groups, though these latter two did not differ significantly (see Figure 1).

Figure 1: Effect of shadowing on VOT.



Although the model speakers' mean VOT was longer than those of the participants, the Korean participants' VOTs still decreased, while no significant change in VOT was observed in the other groups.

4. DISCUSSION

The pattern of results is consistent with the hypothesis. The longer, less stably timed VOTs of the Korean group decreased following the English word shadowing task, while the shorter, more stably timed VOTs of the Spanish group did not change. Given that the model speakers' mean VOT was longer than those of the participant groups' means, the observed reduction in VOT could be interpreted as a more general pattern of VOT reduction due to task-related fatigue, rather than as an instance of phonetic convergence. However, the model speaker's tokens were produced in clear speech style, while the participants' tokens were produced in casual speech style. Participants almost certainly all have experience with clear speech style, whether from hearing individual words uttered in citation form, or from hearing the speech of teachers in the English language classroom. This presents the possibility that on hearing the clear speech English VOTs of the model speaker participants were able to normalize for speech style/register and converged towards a mean VOT appropriate for the casual register in which they were speaking. This latter interpretation depends on participants having mental representations of casual and clear speech VOTs.

While it is not possible to determine unequivocally which of these interpretations most accurately describes the pattern of VOT change, the basic hypothesis that the longer VOTs of Korean speakers are less stably phased and more susceptible to change than the shorter, more stably phased VOTs of Spanish speakers is supported. Thus, language-specific patterns of articulation can account for variability in the likelihood that speakers' phonetic categories will change, and given that this is attributable to the phase relation between laryngeal and oral gestures, the finding is best accounted for within a dynamical systems framework (cf. [11], [24]).

5. REFERENCES

[1] Amazeen, P. G., Amazeen, E. L., & M. T. Turvey. (1998). Dynamics of human intersegmental coordination: Theory and research. In D. A. Rosenbaum & C. E. Collyer, (Eds.), *Timing of behavior: Neural, psychological, and computational perspectives* (pp. 237-259). Cambridge, MA: MIT Press.

[2] Babel, M. E. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*(4), 437-456.

[3] Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

[4] Best, C. T. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In J. Goodman & H. Nusbaum (Eds.). *The development of speech perception: The transition from speech sounds to spoken words* (pp.167-224). Cambridge, MA: MIT Press.

[5] Best, C. & Tyler. M. (2007). Non-native and second language speech perception. In Bohn, O-S. & M. J. Munro (Eds.), *Language experience in second language speech learning. In honor of James Emil Flege* (pp. 15-34). Amsterdam: John Benjamins.

[6] Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*(3-4), 155-180.

[7] Byrd, D., Tobin, S., Bresch, E., & Narayanan, S. (2009). Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *Journal of Phonetics, 37*(1), 87-110.

[8] Dixit, R. P. (1989). Glottal gestures in Hindi plosives. *Journal of Phonetics, 17*(3), 213-237.

[9] Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*(1), 47-65.

[10] Fuchs, S. (2005). Articulatory correlates of the voicing contrast in alveolar obstruent production in German [Monograph]. *Zentrum für Allgemeine Sprachwissenschaft Papers in Linguistics 41*, 1-238.

[11] Gafos, A. I., & Benus, S. (2006). Dynamics of phonological cognition. *Cognitive Science: A Multidisciplinary Journal, 30*(5), 905-943.

[12] Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, N. Coupland & J. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1-68). Cambridge: Cambridge University Press.

[13] Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K.

Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145-165). San Diego: Academic Press.

[14] Kelso, J. A. S. (1981). On the oscillatory basis of movement. *Bulletin of the Psychonomic Society, 18*(2), 62-63.

[15] Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology, 2*(1), 125-156.

[16] Kuhl, P. K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*(2), 93-107.

[17] Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*, 979-1000.

[18] Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384-422.

[19] Löfqvist, A., & Yoshioka, H. (1984). Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. *Speech Communication, 3*(4), 279-289.

[20] Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*(2), 132-142.

[21] Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America, 119*(4), 2382-2393.

[22] Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee, P. Hopper & J. Bybee (Eds.), *Frequency and the emergence of linguistic structure.* (pp. 137-157). Amsterdam: John Benjamins.

[23] Pierrehumbert, J. B. (2002). Word-specific phonetics. *Laboratory Phonology, 7*, 101-139.

[24] Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research Series, 15*, 129-144.

[25] Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*(4), 421-436.

[26] Sawashima, M., & Hirose, H. (1980). Laryngeal gestures in speech production. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, 14*, 29-51.

[27] Strogatz, S. H. (2001). *Nonlinear dynamics and chaos.* Boulder, CO: Westview Press.