Abstract for Speech Imitation: Improvement with Practice

Kuilin Li

Department of Linguistics, McGill University (Canada)

Within the framework of motor control theory, speech production is similar to nonverbal motion [1] in that both contain temporal and spatial information. Temporal information in speech production refers to duration, rhythm, speed, etc., while spatial information refers to vowel quality, consonant quality, pitch, etc. Whether spatial and temporal information are integrated or separated in speech production has been under debate. Articulatory Phonology developed in the Task Dynamics framework (AP/TD) [2] proposes that the phonological representation is spatiotemporal, that is, the timing of speech movement is intrinsic to the phonology [3]. In contrast, the Phonology-Extrinsic-Timing-Based Three-Component model-version 1 (XT/3C-v1) [4] posits that the phonological representation is symbolic and does not contain quantitative information about the movement, which is planned phonetically at a later stage [5] [6].

This study used a learning novel verbal movement task, that is, a second language imitation task to examine whether temporal and spatial information in speech production is integrated or separate. The prediction, based on AP/TD, is that spatial and temporal improvement should be synchronized, whereas according to XT/3C, we predict that temporal and spatial improvement occur at different rates.

We recruited 16 native English monolinguals to imitate 12 Mandarin phrases 30 times each in the laboratory. After the experiment, we acoustically measured the temporal information (speech rate and rhythm) and spatial information (vowel quality, fricative quality, and tone), and compared their stabilization time points in order to see whether these two types of information improved in tandem over the course of the imitation practice. In addition, this study examined the effects of phrase length, segmental complexity, and tone patterns on the spatial and temporal improvements.

The findings on imitation patterns varied considerably across speakers and phrases, but in general the spatial and temporal features improved at different rates. Some speakers were able to achieve native-like stability on the first imitation for short and simple phrases and were able to maintain this stability over many repetitions. However, in most speech productions, there was a decoupling of the improvements of temporal and spatial information in two cases: 1) spatial information stabilized before temporal information, especially in simple phrases; and 2) temporal information stabilized before spatial information, especially in longer phrases. The sequence of imitating novel utterances might follow a spatial (easy) - temporal - spatial (difficult) pattern, with the acquisition and adaptation of temporal information dependent on the initial acquisition of spatial information. This is consistent with the results of previous experiments on nonverbal motor learning [7] [8] and neural evidence [9]. Our findings support the separation of spatial and temporal information in speech production.

References

[1] Kelso, J. A., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, *14*(1), 29-59.

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

[3] Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of phonetics*, *8*(1), 113-133.

[4] Turk, A., & Shattuck-Hufnagel, S. (2020). *Speech timing: Implications for theories of phonology, speech production, and speech motor control* (Vol. 5). Oxford University Press, USA.

[5] Lee, D. N. (1998). Guiding movement by coupling taus. *Ecological psychology*, *10*(3-4), 221-250.

[6] Lee, D. N. (2009). General Tau Theory: evolution to date. *Perception*, *38*(6), 837-850.

[7] Kornysheva, K., Sierk, A., & Diedrichsen, J. (2013). Interaction of temporal and ordinal representations in movement sequences. *Journal of Neurophysiology*, *109*(5), 1416-1424.

[8] Torres, E., & Andersen, R. (2006). Space–time separation during obstacle-avoidance learning in monkeys. *Journal of neurophysiology*, *96*(5), 2613-2632.

[9] Kornysheva, K., & Diedrichsen, J. (2014). Human premotor areas parse sequences into their spatial and temporal features. *Elife*, *3*, e03043.