

Multimodal training can facilitate L2 phoneme acquisition

Marieke Hoetjes¹, Lieke van Maastricht¹, Lisette van der Heijden¹

¹ Centre for Language Studies, Radboud University Nijmegen

Abstract

Speech and gesture are closely related (Kendon, 2004; McNeill, 1992), and gestures have been shown to contribute to language development. Previous work on L1 acquisition showed that children start communicating by pointing at objects, and that these gestures predict the lexical items appearing in the child's vocabulary (Goldin-Meadow, 2007). In L2 acquisition, it has been demonstrated that in vocabulary acquisition, novel words are better memorised when learned with a gesture (Tellier, 2008). At the phonetic level, the evidence on the facilitatory role of gestures in L2 acquisition is less consistent. Some studies have shown that seeing gestures helps in perceiving L2 phonemic contrasts (Hannah, Wang, Jongman, & Sereno, 2017; Kelly, Bailey, & Hirata, 2017), whereas others found the opposite (e.g., Kelly, Hirata, Manansala, & Huang, 2014). Moreover, existing studies on gestures in L2 phoneme acquisition focused on *perception*, but it is unknown whether gestures can also improve L2 phoneme *production*. An imperative question, as many L2 learners aim to sound native-like in their L2.

To determine whether including gesture in language training facilitates L2 phoneme production, 51 native speakers of Dutch received training on the target-like pronunciation of the Spanish phonemes /θ/ and /u/, which are typically difficult for native speakers of Dutch. Participants were asked to read aloud Spanish sentences including words with /θ/ and /u/ before (T1) and after training (T2). Participants received training in one of four training conditions: audio-only (AO), audio-visual (AV), audio-visual with pointing gestures (AV-P), or audio-visual with iconic gestures (AV-I). They received a written explanation in Dutch of how to pronounce the target phonemes in Spanish, accompanied by two example phrases produced by a Spanish native, in one of the four conditions. In the AV-P condition, the speaker pointed at her mouth as she produced the target phoneme. In the AV-I condition, she produced an iconic gesture illustrating how to produce the target phoneme.

The target phonemes were extracted from the data and annotated phonetically. Reliability (50% of data) between 2 coders deciding whether the target phonemes that were produced were on target, or not, was high ($\kappa=.900$, $p<.001$). Annotations for the same items were then compared between T1 and T2, and coded for whether there was progress after learning, i.e., whether the target phoneme was not produced on-target at T1, but on-target at T2. Cases in which the participant was always able or never able to produce the phoneme, or not able to produce the target phoneme anymore at T2, were coded as no progress.

The results showed that /u/ is easier to acquire than /θ/ and that training modality can affect on-target production. Specifically, the descriptives showed that all training conditions with a visual component lead to more on-target productions than the AO training. However, the statistical analyses showed that the effectiveness of the type of multimodal training varied between the two phonemes. For /u/, we found no significant association between training condition and progress. Most participants were always able to produce this phoneme correctly. For /θ/, we found a significant association between training condition and progress. The AV-P condition was particularly helpful for learning, while the AV-I condition caused progress to be less than expected. Interestingly, in the majority of cases, /θ/ was not produced target-like at T1 and on-target pronunciation was never learned, suggesting that this phoneme is particularly challenging for L2 learners.

This study shows that multimodal training can facilitate L2 phoneme production. Pointing seems helpful, especially when learning challenging L2 sounds, possibly because processing pointing gestures requires less cognitive resources than the processing of iconic gestures. These resources can subsequently be used in the actual production of the L2 phoneme.

