MULTIMODAL TRAINING FACILITATES L2 PHONEME ACQUISITION: AN ACOUSTIC ANALYSIS OF DUTCH LEARNERS' SEGMENT PRODUCTION IN SPANISH

Lieke van Maastricht, Marieke Hoetjes & Lisette van der Heijden[1]

Centre for Language Studies, Radboud University, Nijmegen
l.vanmaastricht@let.ru.nl, m.hoetjes@let.ru.nl, l1.vanderheijden@student.ru.nl

ABSTRACT

Given that co-speech gestures affect language perception in both the L1 and L2, this paper aims to determine whether they can also lead to improved L2 production. To this end, 51 native speakers of Dutch received training focused on the target-like pronunciation of the Spanish phonemes /θ/ and /u/, which are typically difficult to acquire for native speakers of Dutch. Participants were allocated to one of four training conditions: audio-only, audio-visual, audio-visual with pointing gestures, or audio-visual with iconic gestures. Before and after training, participants read aloud Spanish sentences that included words with /θ/ and /u/. Acoustic analysis revealed that /u/ is easier to acquire than /θ/ and that training modality affects on-target production. More specifically, all training conditions that included the visual modality lead to more on-target productions than the audio-only training. Interestingly, the effectiveness of the different types of multimodal training varies between the two phonemes.

Keywords: multimodality; phoneme acquisition; gesture; Dutch; Spanish.

1. INTRODUCTION

Language is generally viewed as an embodied, multimodal system in which the motor, visual, and speech modalities are integrated to convey meaning [18, 21, 22, 30]. Several studies on L1 acquisition have reported that children start communicating by pointing at objects they do not yet have labels for [4, 5, 12, 28]. Moreover, these early pointing gestures seem to predict the lexical items appearing in the child's vocabulary [12, 24]. Thus, gestures seem to pave the way for language development. Similar findings are reported for L2 acquisition, where studies have demonstrated that novel words learned with iconic gestures are better memorised than words learned without iconic gestures [17, 25]. The gestures are thought to enrich the sensorimotor memory trace and therefore facilitate the recall of novel words [19]. Visual and gestural input is also known to affect L1 comprehension at the phonetic level. Listeners are reported to use face and mouth movements as well as pointing gestures to disambiguate speech [20, 26, 27]. As phonemic accuracy is also crucial to L2 learners' intelligibility, comprehensibility and accentedness [1, 3, 23], research on the interplay between gestural and phonemic input is especially relevant.

Recent work has demonstrated that seeing the speaker helps in the acquisition of L2 phoneme contrasts [8, 9], yet studies on the role of gestures in the perception of non-native tonal and phonemic contrasts report contrasting findings: Hannah, Wang, Jongman, and Sereno [7] and Kelly, Bailey, and Hirata [14] revealed that gestural training significantly improves the perception of L2 phonemic tones and intonation contours, but work by Hirata, Kelly and colleagues [10, 11, 14, 15] revealed no significant improvement in the perception of non-native phonemic vowel length distinctions after gestural training. Kelly et al. [14] concluded that "gestures help with some – but not all – novel speech sounds in a foreign language" (p. 1). Thus, while gesture and speech are clearly integrated at the semantic and suprasegmental phonetic level [13], prior work shows that it is less clear whether gestures also contribute to perception at the phonemic level. Moreover, while prior studies focused on the *perception* of non-native phoneme contrasts by L2 learners, the role of gesture in the acquisition of non-native phoneme *production* remains unknown. Finally, earlier work did not compare the effect of different types of gestures on L2 learning. Distinguishing between, i.e., pointing gestures that only serve to draw attention to the mouth and iconic gestures that also give information on what the speaker should do to achieve on-target pronunciation might lead to more specific conclusions about the relevance of gestures in L2 classrooms.

Therefore, our research question is: Does instruction modality affect L2 learners' production of non-native phonemes? We hypothesize: 1) that adding audio-visual information to language training will be beneficial for phoneme acquisition compared to

[1] The authors are grateful to Núria Domínguez, Hien Huynh, and Judith Peters for their assistance in material preparation and data collection and analysis.

providing only audio information [8, 10, 29]. 2) Given that the use of gestures is helpful in the acquisition of certain segments [7, 14], as well as suprasegments [6], using gestures in the audio-visual training will be more beneficial than not including them. As prior work has not yet compared the effect of different types of gestures, no predictions can be made on whether iconic gestures will facilitate phoneme acquisition more than pointing gestures (the two types of gestures used in the current study).

2. METHOD

2.1. Design

This study had a between-subjects design and included a pre-test (T1) and a post-test (T2). Participants took part 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). The dependent variable was the pronunciation of the target phonemes, coded as either on-target or not.

2.2. Subjects

Fifty-one adult L1 speakers of Dutch (28 female, 23 male), with an average age of 25 years old (range 18-61 years old) took part in the study. Participants did not speak Spanish and had no auditory or visual impairments that could affect their participation. They were recruited via the Radboud University research participation system and received either credits or a small financial reward for their participation.

2.3. Materials

2.3.1. Sentences

In this study, we focused on the L2 production of the Spanish phonemes /θ/ and /u/, as read aloud by participants in the same set of four-word sentences at T1 and T2 in one of two randomised orders. /θ/ and /u/ were chosen, as on-target production of these phonemes was expected to be complicated by two factors: 1) The difference in grapheme-to-phoneme conversion between Dutch and Spanish. The grapheme ‘u’ should be pronounced as /u/ in Spanish, while in Dutch it is usually pronounced as /y/, /ə/, or /ʏ/. Similarly, the grapheme ‘z’ is pronounced as /θ/ in Spanish, but as /z/ or /s/ in Dutch. 2) The possible absence of L2 segments in the L1 inventory. While the /u/ is part of the Dutch phoneme inventory, /θ/ is not. The sixteen sentences read by the participants were presented on separate PowerPoint slides. Each sentence was accompanied by a picture illustrating the meaning of the sentence, to help participants understand the semantic meaning of the sentence, and to make the task more interesting (see Figure 1). Half of the sentences had a word containing the target phoneme as the second word of the sentence. The target phoneme always occurred in the first syllable of this two-syllable word (e.g., *La nube es blanca, La zeta es verde*). Each target phoneme occurred in four target words. The remaining eight sentences were fillers, containing the phonemes of interest, but at a different position either within the word or within the sentence. For this paper, the filler items were not analysed.

Figure 1: Example of an experimental item containing the target phoneme /u/.

La nube es blanca

2.3.2. Training

After T1, the participants received a short training focusing on the pronunciation of /θ/ and /u/. The order of the two phonemes within the training was counterbalanced across participants. During training, participants studied PowerPoint slides on which information was given about the way in which each target phoneme is pronounced in Spanish. Specifically, they were informed that the Spanish pronunciation of both graphemes differs from the Dutch pronunciation, and were given information about which articulatory gestures are necessary for on-target pronunciation (e.g., “when pronouncing the letter ‘u’ in Spanish, you need to round your lips”). The training included several examples, produced by an L1 speaker of Spanish; one example segment was given on the same slide as the written information about the phoneme, and two example sentences were given on subsequent slides.

The manipulation of the training modality consisted of the fact that the examples were presented in either the AO, AV, AV-P, or AV-I condition. The same audio (from the L1 speaker seen in the video) was dubbed over all conditions, but they differed with regard to the video material that was presented: In the AO condition, participants heard the examples, but did not see the speaker. In the AV condition, a video of the speaker was shown, but the speaker did not move her body, apart from her mouth. In the AV-P condition, the speaker made a pointing gesture towards her mouth while producing the target phoneme. In the AV-I condition, the speaker made an iconic gesture as she produced the target phoneme. This iconic gesture represented the articulatory gesture needed for on-target segment production, as explained in the training. For the /u/, it was a one-handed gesture indicating the rounding of the lips (see Figure 2), and for the /θ/ it was a one-handed gesture indicating that the speaker should push their tongue out between their teeth (see Figure 3).

Figure 2: Still from training video in AV-I condition showing the articulatory gesture needed for /u/.

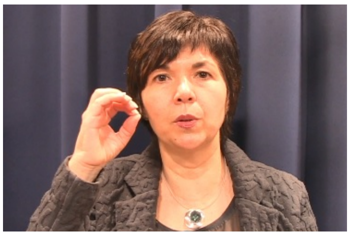
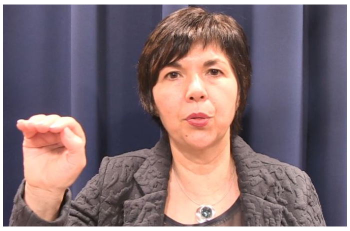


Figure 3: Still from training video in AV-I condition showing the articulatory gesture needed for /θ/.



2.4. Procedure

The experiment took place in a sound-proof booth to minimize distractions. Participants signed a consent form, and then took part in the experiment, which was self-paced. Written and verbal instructions were given, followed by a practice sentence and the opportunity to ask questions. Participants were then asked to read out the sixteen Spanish sentences into a microphone (T1). They were allowed to repeat the sentence until they were satisfied with their production; the final attempt was used for analysis. After T1, a language background questionnaire was administered and after receiving one of the four types of training, participants reread the same sentences as during T1, albeit in a different order (T2). The audio produced by participants during T1 and T2 was recorded, and production of the target phonemes was analysed with Praat (version 6.0.43, [2]). The entire experiment, with the exception of the Spanish sentences, took place in Dutch.

2.5. Analyses

First, the words containing the target phonemes (8 T1 + 8 T2 words × 51 participants = 816 segments) were extracted from the sound files. The target phonemes were then annotated phonetically. In the annotation, coders distinguished between target-like production (i.e., as an L1 speaker of Iberian Spanish would do) and several non-target options (for /θ/: /s/, /z/, or ‘other’; for /u/: /y/, /ə/, /ʏ/, or ‘other’). All extracted phonemes were annotated by two phonetically trained coders, with an overlap in coding of 50%. The interrater reliability was good, $\kappa = .820$, $p < .001$. For the current analysis, we only distinguish between on-target versus non-target productions, collapsing data across the non-target options. This further improved interrater reliability, $\kappa = .900$, $p < .001$.

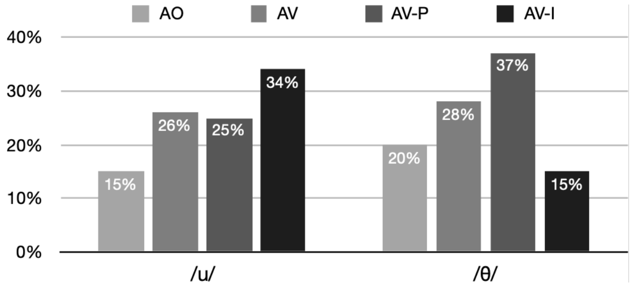
Annotations for the same items were then compared between T1 and T2, and we coded whether the participant was able to produce the target phoneme at T1, but not anymore at T2 (1), was not able to produce the target phoneme at either T1 or T2 (2), was able to pronounce the target phoneme at both T1 and T2 (3), or was unable to produce the target phoneme at T1, but able to do so at T2 (4). We distinguished between progress (i.e., (4)), where the participant learned to produce the target phoneme), and no progress (i.e., (1), (2), and (3)). Chi-square analyses were used to analyse whether training modality affected target phoneme production.

3. RESULTS

The analysis of the results for on-target /u/, i.e., using only those productions coded as (4), revealed no significant association between training condition and progress, $\chi^2(3) = 6.679$, $p = .083$. Yet, the highest proportion of learning was obtained in the AV-I training, especially compared to the AO condition, suggesting that for acquiring /u/ the AV-I condition is most helpful (Figure 4). Inspection of the frequencies of the results coded as (1), (2) or (3) show that in 64.6% of all cases, participants already produced the /u/ correctly at T1, continuing to do so at T2 (vs. the 31.3% of all cases in which participants produced the /u/ off-target at T1 and on-target at T2).

The chi-square analysis for target production of /θ/ showed a significant association between training condition and progress, $\chi^2(3) = 9.155$, $p = .027$. The progress in the AV-P and AV-I conditions differed significantly from the expected values. The analysis revealed that within the AV-P condition the proportion of cases with progress (37%) was significantly higher than the proportion of cases without progress (20%). In other words, it appears that for the acquisition of /θ/, the AV-P condition is particularly helpful, while the AV-I condition is particularly harmful. Interestingly, inspection of the frequencies of the results coded as (1), (2) or (3) show that in the case of /θ/, in the majority of all cases (64.5%), participants never learned to produce the /θ/ correctly (vs. the 34.5% of all cases in which participants did learn to produce the /θ/ on-target between T1 and T2). This suggests that this phoneme is particularly challenging for L2 learners, in contrast to /u/, which appears to be substantially less challenging.

Figure 4: Percentages of successful /u/ (left) and /θ/ (right) acquisition, separated by training condition.



4. DISCUSSION

The aim of this study was to determine whether instruction modality affects L2 learners' production of non-native phonemes during a reading task. We expected that audio-visual information would be more beneficial than only audio during training (H1), that showing gestures in training would be more beneficial than providing audio-visual information without gestures (H2), and we aimed to determine whether the type of gesture used during training (iconic or pointing) matters. Results show that adding the visual modality to L2 phoneme training generally leads to more on-target productions by L2 learners, corroborating H1. Although there was no significant association between training condition and progress for the /u/, the descriptives suggest that all visual conditions yield higher proportions of learning than the AO condition. For the /θ/, there was a significant association between training condition and progress, and the descriptives imply that both the AV and AV-P condition generate higher proportions of learning than the AO condition but the AV-I does not. Concerning H2, results differ across phonemes: For the /u/, the descriptives suggest that using an iconic gesture in AV training improves phoneme production while using a pointing gesture does not. Conversely, for the /θ/, adding a pointing gesture to the AV training improves phoneme production, while using an iconic gesture actually generates lower proportions of learning. These findings support the idea that while audio-visual information in general, and gestures in particular, are beneficial in grapheme-to-phoneme training, the type of gesture that is most beneficial is phoneme-dependent.

These findings might be explained by the fact that /θ/ appears to be more challenging for L2 learners than /u/, irrespective of the training learners received. Unexpectedly, learners often already produced the /u/ on-target at T1, making it impossible for progress to take place. Also, the /θ/ does not exist in the Dutch phoneme inventory, which could have obstructed its successful acquisition by L2 learners after only one training session. The fact that the use of an iconic gesture in phonemic training is not beneficial when the target segment is particularly difficult for learners corroborates previous work reporting that the use of iconic gestures in phonetic/ semantic training benefits L2 word learning, but only when the phonetic demands of the target words are low [16]. Similarly, prior work showed that seeing lip movements with speech helped L2 learners to make phonemic contrasts, but adding (here metaphoric) gestures to audio-visual training actually impaired learners [10]. In addition, it should be noted that as the /u/ is present in the Dutch phoneme inventory, while the /θ/ is not, progress in the /u/ context might reflect participants' knowledge of the grapheme-to-phoneme conversions rather than the target phoneme acquisition.

The chi-square analysis shows that the use of pointing gestures in training is just as beneficial (in the case of /u/), if not more beneficial (in the case of /θ/), for L2 segment production than the use of iconic gestures. Thus, maybe providing a gesture that merely directs the learner's attention to the phoneme's articulation is more beneficial than providing a gesture that also provides information about specific details of the phoneme articulation. This may not be that surprising given that iconic gestures are typically related to speech at the semantic, and not the phonetic, level [13]. Having said this, future work might focus on other phonemic contrasts, while also controlling whether the L2 segment exists in the L1 inventory.

In sum, gestural training appears to stimulate L2 phoneme learning and can thus be considered a useful tool in L2 classrooms, but more research is needed to determine which type of gesture is to be used depending on the segment under acquisition. In addition, it remains to be determined whether non-target acoustic realisation of phonemes also affects L1 perceptions of L2 learners' speech, e.g., with respect to measures of accentedness, comprehensibility, and/or intelligibility. If instruction modality affects not only phoneme production, but changes in phoneme production in turn also affect the way in which an L2 learner is perceived by L1 speakers, this would lend support to using a multimodal approach, one including gestures, in the L2 classroom.

5. REFERENCES

[1] Anderson-Hsieh, J., Johnson, R., Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*(4), 529–555.

[2] Boersma, P., Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved 8 September 2018 from http://www.praat.org/

[3] Caspers, J., Horłoza, K. (2012). Intelligibility of non-natively produced Dutch words: Interaction between segmental and suprasegmental errors. *Phonetica, 69*(1-2), 94–107.

[4] Goldin-Meadow, S. (2007). Pointing sets the stage for learning language – and creating. *Child Development, 78* (3), 741–745.

[5] Goldin-Meadow, S. (2014). Widening the lens: what the manual modality reveals about language, learning and cognition. *Philosophical Transactions of the Royal Society, 369*, 1–11.

[6] Gluhareva, D., Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, *21*(5), 609–631.

[7] Hannah, B., Wang, Y., Jongman, A., Sereno, J. A. (2017). Cross-modal association between auditory and visual-spatial information in Mandarin tone perception. *The Journal of the Acoustical Society of America, 140*(4), 3225–3225.

[8] Hardison, D. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24*(4), 495–522.

[9] Hazan, V., Sennema, A., Iba, M., Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication, 47*(3), 360–378.

[10] Hirata, Y., Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, *53*(2), 298–310.

[11] Hirata, Y., Kelly, S. D., Huang, J., Manansala, M. (2014). Effects of hand gestures on auditory learning of second-language vowel length contrasts. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2090–2101.

[12] Iverson, J. M., Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological science, 16*(5), 367–371.

[13] Kelly, S. (2017). Exploring the boundaries of gesture-speech integration during language comprehension. In R. B. Church, M. W. Alibali, S. D. Kelly (Eds.), *Why Gesture?* (pp. 243–256): John Benjamins.

[14] Kelly, S., Bailey, A., Hirata, Y. (2017). Metaphoric gestures facilitate perception of intonation more than length in auditory judgments of non-native phonemic contrasts. *Collabra: Psychology, 3*(1), 7.

[15] Kelly, S. D., Hirata, Y., Manansala, M., Huang, J. (2014). Exploring the role of hand gestures in learning novel phoneme contrasts and vocabulary in a second language. *Frontiers in Psychology, 5*, 1–11.

[16] Kelly, S., Lee, A. (2012). When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high. *Language and Cognitive Processes*, 27(6), 793–807.

[17] Kelly, S.D., McDevitt, T. Esch, M. (2009). Brief training with co-speech language lends a hand to word learning in a foreign language. *Language and Cognitive Processes, 24*(2), 313–334.

[18] Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press.

[19] Macedonia, M., Klimesch, W. (2014). Long-term effects of gestures on memory for foreign language words trained in the classroom. *Mind, Brain, and Education, 8*(2), 74–88.

[20] McGurk, H., MacDonald, J. (1976). Hearing lips and seeing voices. *Nature 264*, 746–748.

[21] McNeill, D. (1992). *Hand and mind. What gestures reveal about thought*. Chicago: University of Chicago Press.

[22] McNeill, D. (2008). *Gesture and Thought*. Chicago: University of Chicago Press.

[23] Munro, M. J., Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *49*, 285–310.

[24] Rowe, M. L., Özçalışkan, Ş., Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language, 28*(2), 182–199.

[25] Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture, 8*(2), 219–235.

[26] Thompson, L. A., Massaro, D. W. (1986). Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology*, *42*(1), 144–168.

[27] Thompson, L. A., Massaro, D. W. (1994). Children′ s integration of speech and pointing gestures in comprehension. *Journal of Experimental Child Psychology*, *57*(3), 327–354.

[28] Tomasello, M. (2008). *Origins of Human Communication*. Cambridge: MIT Press.

[29] Wang, Y., Behne, D., Jiang, H. (2008). Linguistic experience and audio-visual perception of non-native fricatives. *Journal of the Acoustical Society of America, 124*, 1716–1726.

[30] Zwaan, R. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of Learning and Motivation 44*, 35–62.