Gestural training benefits L2 phoneme acquisition: Findings from a production and perception perspective

Marieke Hoetjes, Lieke van Maastricht, and Lisette van der Heijden

Centre for Language Studies, Radboud University, The Netherlands

m.hoetjes@let.ru.nl, l.vanmaastricht@let.ru.nl, l1.vanderheijden@student.ru.nl

Abstract

This paper aims to study whether training with gestures benefits L2 phoneme acquisition from both a production and perception perspective. In the production study, Dutch learners of Spanish received pronunciation training for the phonemes /u/ and /θ/ in one of four conditions: audio-only, audio-visual, audio-visual with pointing gestures, or audio-visual with iconic gestures. Results show that in general, gestural training benefits L2 phoneme acquisition, but different gestures benefit the acquisition of different phonemes, possibly depending on their complexity. The perception study, in which L1 speakers of Spanish judged the L2 Spanish material on accentedness and comprehensibility, corroborate the findings from the production study: Including visual information in training generally lowered the perceived accentedness and increased the perceived comprehensibility of speech, but the type of phoneme matters. Together, these studies suggest that gestural training can benefit L2 phoneme acquisition, yet certain gestures work better for certain phonemes than others.

1. Introduction

It has been established that speech and gesture are closely related (Kendon, 2004; McNeill, 1992), evidenced, for example, by the semantic and temporal coordination of speech and gesture (see e.g. Gullberg, 2006, for an overview). Moreover, previous studies have shown that gesture is relevant in language development, as children produce pointing gestures for objects they do not yet have labels for (e.g., Iverson & Goldin-Meadow, 2005), and these gestures predict the words that are to appear next in their vocabulary. Gestures have also been shown to facilitate learning, both in non-linguistic (Yeo, Wagner Cook, Nathan, Popescu, & Alibali, 2018), and linguistic contexts. For example, gestures have been shown to benefit the acquisition of novel L2 words (Kelly, McDevitt, & Esch, 2009; Tellier, 2008). Regarding L2 phonemic contrasts specifically, prior studies have demonstrated that seeing the speaker benefits the production of phonemes by L2 learners (Hardison, 2003; Hazan, Sennema, Iba, & Faulkner, 2005), yet studies on the role of gestures in the perception of L2 tonal and phonemic contrasts report contrasting findings: Hannah, Wang, Jongman, and Sereno (2017) and Kelly, Bailey, and Hirata (2017) revealed that training with gestures significantly improves the perception of non-native phonemic tones and intonation contours, while work by Hirata, Kelly, and colleagues (Hirata & Kelly, 2010; Hirata, Kelly, Huang, & Manansala, 2014; Kelly et al., 2017; Kelly, Hirata, Manansala, & Huang, 2014) revealed no significant improvement in the perception of non-native phonemic vowel length distinctions after gestural training. Kelly et al. (2017) concluded that “gestures help with some – but not all – novel speech sounds in a foreign language” (p. 1).

Thus, while gestures are known to be a common and effective resource in L1 communication, less is known about their possibly beneficial effects in the context of L2 acquisition, especially concerning the educational value of different types of gestures. Previous studies on gestures in L2 pronunciation training have used varying gesture types and hand movements, e.g., beats (e.g. Gluhareva & Prieto, 2017), metaphoric gestures (e.g. Kelly et al., 2014), clapping (e.g. Zhang, Baills, & Prieto, 2018), with varying methods and results, which complicates the process of determining which kinds of gestures facilitate L2 phoneme acquisition and under which conditions. In addition, most studies on the effects of gesture on L2 phoneme acquisition rely only on perception measures, either by analysing on-target phoneme production through L1 perception measures (e.g. Gluhareva & Prieto, 2017), or by testing whether L2 learners can discriminate between different L2 phonemes in a perception task (e.g. Kelly et al., 2017). Therefore, we aim to

 doi:10.17619/UNIPB/1-808

determine whether instruction modality affects L2 phoneme acquisition, distinguishing between four training conditions: 1) training in which examples are presented as audio fragments only; 2) training in which examples are presented as video fragments but the trainer does not gesture; 3) training in which examples are presented as video fragments and the trainer produces a pointing gesture towards her mouth when producing the target phoneme; and 4) training in which examples are presented as video fragments and the trainer produces an iconic gesture visualizing the position and/or form of the relevant articulators near her mouth when producing the target phoneme. As measures of successful L2 phoneme acquisition, we analyse both the phonetic characteristics of L2 speech (Study 1) and L1 listeners' ratings of foreign accentedness and comprehensibility (Study 2).

Based on prior research, we hypothesize that adding audio-visual information to language training will be beneficial for phoneme acquisition compared to providing only audio information (Hardison, 2003; Hirata & Kelly, 2010; Wang, Behne, & Jiang, 2008). Given that the use of gestures is helpful in the acquisition of certain segments (Hannah et al., 2017; Kelly et al., 2017), 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 regarding comparisons between iconic and pointing gestures. Based on Zhang, Baills and Prieto (2018), we predict that the findings for our production and perception measures will be congruent, with possibly a stronger effect of gestural training on perception than on production.

2. Study 1: Production

This study was set up using a pre-test (T1) – training – post-test (T2) design. Fifty-one L1 speakers of Dutch (30 female, mean age 25 years old, range 18-61 years old), who did not speak Spanish, 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).

2.1. Materials

We focused on the acquisition of the Spanish phonemes /θ/ and /u/, since their nativelike production by L2 learners is often 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, whereas in Dutch it is generally pronounced as /y/, /ə/, or /ʏ/. Likewise, the grapheme 'z' is pronounced as /θ/ in Spanish, yet as /z/ or /s/ in Dutch. 2) The possible absence of L2 segments in the L1 inventory. While the /u/ exists in the Dutch phoneme inventory, /θ/ does not.

The phonemes were embedded in 16 four-word sentences, which were read aloud by participants at T1 and T2 in one of two randomised orders. Each sentence was presented on a separate PowerPoint slide. Above each sentence, a picture illustrated its meaning. Half of the sentences were experimental items, which 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 of the two target phonemes occurred in four target words. The remaining eight sentences were fillers, which were not currently analysed.

After T1 and before T2, the participants received a short training focusing on the Spanish pronunciation of /θ/ and /u/ (in counterbalanced order). Training consisted of a set of PowerPoint slides on which information was given about how each target phoneme is pronounced in Spanish. Specifically, participants were told that the Spanish pronunciation of both graphemes is different from the Dutch pronunciation of these graphemes, and it was explained which articulatory gestures are necessary for nativelike pronunciation (e.g., "when pronouncing the letter 'u' in Spanish, you round your lips"). The training included various examples, produced by an L1 speaker of Spanish; one example was given on the same slide as the written information about the respective target phoneme, and two examples were given on successive slides.

To manipulate the training modality, the visual information that was presented to participants in each condition was varied while the audio (from the L1 speaker seen in the video) was dubbed over all conditions: In the AO condition, participants heard the audio example but did not see the video. In the AV condition, a video of the speaker was shown, but the speaker did not gesture. In the AV-P condition, the speaker pointed towards her mouth while producing the target phoneme. In the AV-I condition, the speaker made an iconic gesture representing the articulatory gesture needed for on-target segment production as she produced the target phoneme. For the /u/, this was a one-handed gesture indicating the rounding of the lips, and for the /θ/, this was a one-handed gesture indicating that the speaker should push their tongue out between their teeth (see Figure 1).

Figure 1. Stills from training video in AV-I condition showing the articulatory gesture needed for /u/ (left) and /θ/ (right).

2.2. Procedure

The experiment took place in Dutch (with the exception of the Spanish sentences) in a soundproof booth. After receiving instructions, participants read out the 16 Spanish sentences into a microphone (T1). After T1, participants completed a language background questionnaire, followed by one of the four types of training. Participants then read a reordered version of the same sentences (T2). The audio produced during T1 and T2 was recorded, and production of the target phonemes was analyzed using Praat (Boersma & Weenink, 2018). The target phonemes were annotated by two phonetically trained coders, distinguishing between nativelike production (i.e., as an L1 speaker of Iberian Spanish would do) and several non-nativelike options (for /θ/ these were /s/, /z/, or 'other'; for /u/: these were /y/, /ə/, /ʏ/, or 'other'). For the present analysis, on-target productions were distinguished from non-target productions, collapsing data across the non-target options. There was an overlap in coding of 50%, with a good interrater reliability, $\kappa = .900$, $p < .001$. Annotations for the same sentences were compared between T1 and T2, resulting in 4 options: 1) the participant was able to produce the target phoneme at T1, but not anymore at T2; 2) the participant was not able to produce the target phoneme at either T1 or T2; 3) the participant was able to pronounce the target phoneme at both T1 and T2; or 4) the participant was unable to produce the target phoneme at T1, but able to do so at T2. In the current analysis, we distinguish between progress (i.e., option 4), and no progress (i.e., options 1-3) and conduct chi-square analyses to analyse whether training modality affected target phoneme production.

2.3. Results

The analysis of the results for on-target /u/, i.e., using only those productions coded as option 4, did not reveal a significant association between training condition and progress, $\chi2(3) = 6.679$, $p = .083$. However, the highest proportion of learning was obtained after the AV-I training, which is substantially higher than the proportion of learning after AO training (see Figure 2). The frequencies of the results coded as options 1-3 showed that in 64.6% of all cases participants already produced the /u/ correctly at T1, and continued to do so at T2 (vs. the 31.3% of all cases coded as option 4).

The 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 the proportion of cases with progress in the AV-P condition (37%) was significantly higher, and the proportion of cases with progress in the AV-I condition (15%) was significantly lower than expected. In other words, for the acquisition of /θ/, the AV-P condition is particularly helpful but the AV-I condition is particularly harmful (see Figure 2). Interestingly, inspection of the frequencies of the results for /θ/ show that in the majority of all cases (64.5%) participants never learned to produce the /θ/ correctly. This suggests that /θ/ is particularly challenging for L2 learners, in contrast to /u/, which appears to be substantially less challenging.

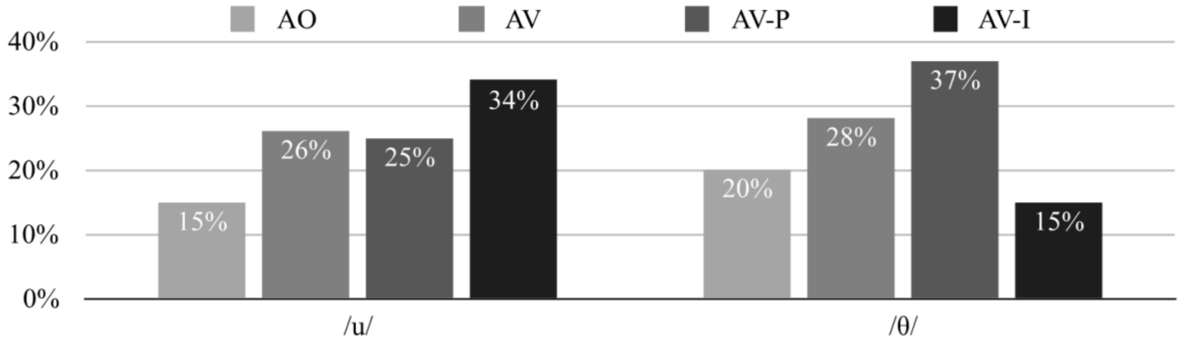

Figure 2. Percentages of /u/ (left) and /θ/ (right) acquisition, separated by training condition.

3. Study 2: Perception

In this within-subjects design, 46 L1 Spanish speakers (19 females, mean age 31 years old, range 19-70 years old) listened to a selection of target words, produced at T1 and at T2 after AV, AV-P, or AV-I training, and rated them on accentedness (21 subjects) or comprehensibility (25 subjects).

3.1. Materials and Procedure

In order to keep the length of the experiment acceptable, we used 8 items (2 with /θ/ and 2 with /u/ from T1 and T2) from 21 speakers (7 randomly chosen speakers from 3 types of training condition: AV, AV-P, and AV-I) from the production study, resulting in 168 items. The AO condition was left out as it represents the least realistic learning context. Accentedness was measured with the statement "This speaker speaks …", followed by a 7-point semantic differential anchored by "without a foreign accent - with a strong foreign accent" (based on Jesney, 2004). Comprehensibility was measured with the statement "I find this speaker easy to understand" followed by a 7-point Likert scale anchored by "totally disagree - totally agree" (based on Derwing & Munro, 1997). Before participants rated the items, a brief explanation of either accentedness or comprehensibility was given. The entire experiment took place in Spanish. Subjects were requested to wear headphones.

3.2. Results

Accentedness ratings were transformed to reflect the same direction of effect as comprehensibility ratings, i.e., a higher rating always reflects more nativelike speech. A repeated measures analysis for accentedness with Type of training (4 levels: T1, AV, AV-P, AV-I) and Target sound (2 levels: /u/ or /θ/) as within-subject factors showed a significant main effect of Type of training (F (3, 72) = 16.17, $p < .001$, $\eta p^2 = .40$), no main effect of Target sound (F (1, 24) < 1 , $p = .606$), and a significant interaction between Type of training and Target sound (F (3, 72) = 12.94, $p < .001$, $\eta p^2 = .35$). Pairwise comparisons between training conditions within individual phonemes revealed that for words containing /u/, scores for identical items increased significantly between T1 and T2 after AV-P ($p < .001$) and AV-I training ($p < .001$), but not after AV training. For words containing /θ/, scores for identical items increased significantly between T1 and T2 after AV-P training ($p < .001$) but not after AV or AV-I training, implying that L2 speakers only benefitted from AV-P training.

The analysis for comprehensibility revealed a significant main effect of Type of training (F (2.04, 40.72) = 10.26, $p < .001$, $\eta p^2 = .34$), no significant main effect of Target sound (F (1, 20) < 1, $p = .749$), and a significant interaction between Type of training and Target sound (F (3, 60) = 10.74, $p < .001$, $\eta p^2 = .35$). Pairwise comparisons between training conditions within individual phonemes revealed that for words containing /u/, scores were significantly higher at T2 than at T1 after both AV-P ($p = .003$) and AV-I training ($p = .001$), but not after AV training. Conversely, for words containing /θ/, scores for identical items increased significantly between T1 and T2 after AV-P training ($p = .001$) but decreased significantly after AV-I training ($p = .044$). This suggests that, for comprehensibility, L2 speakers always benefitted from AV-P training, but were actually hindered by the AV-I training for /θ/, see Figure 3, which shows the mean ratings of accentedness and comprehensibility for both /u/ and /θ/.

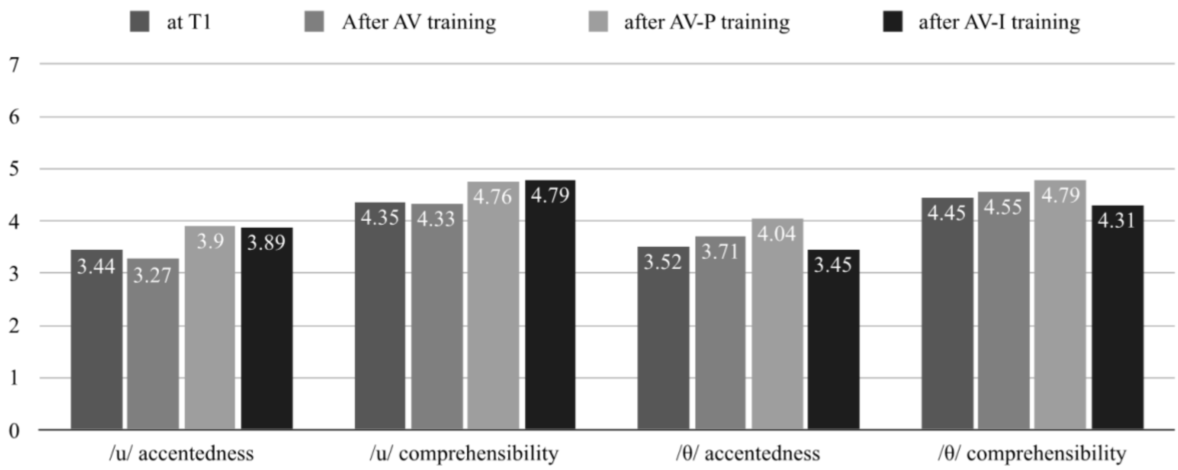

Figure 3. Mean ratings of accentedness (left) and comprehensibility (right) for /u/ and /θ/.

4. General Discussion

The aim of the current studies was to determine whether training with gestures can facilitate L2 phoneme acquisition. Specifically, we aimed to determine whether instruction modality affects L2 phoneme acquisition. We analysed the phonetic characteristics of speech produced by L2 speakers of Spanish (Study 1), and L1 listeners' ratings of accentedness and comprehensibility for a selection of these same items (Study 2). Our first hypothesis was that adding audio-visual information to training would be more beneficial than providing audio only information during training. In addition, we specifically expected that including gestures in the audio-visual training would be more beneficial than not including gestures. Whether there would be differences concerning the facilitatory effects of the different types of gestures was an open question. Moreover, we expected findings from the perception study to confirm those of the production study.

The results from Study 1 showed different effects for the two phonemes under investigation: For /u/, there was no significant association between training condition and progress, even though more cases of native-like phoneme production occurred in all types of audio-visual training in comparison to AO training, and the AV-I training appeared most beneficial. It is important to note though, that /u/ was relatively easy for participants to produce, which was apparent by the fact that in many cases, native-like production already took place at T1. For /θ/, however, results showed that many speakers never acquired a native-like production. For those speakers who did acquire native-like production after training, our results showed that AV-P training was helpful, but AV-I training was harmful. Thus, the findings from Study 1 suggest that in general, including visual information in phoneme training helps, but also that the combination of type of gesture and type of phoneme matters: iconic gestures, which typically provide more semantic information than pointing gestures, seem helpful only when the phoneme is relatively easy to acquire. For a phoneme that learners find difficult to acquire, including an iconic gesture in training is not helpful for acquisition. However, for this difficult phoneme, including a pointing gesture, which mainly served to point the listener's attention to the mouth of the trainer, does facilitate acquisition. This is in line with prior research stating that the use of iconic gestures in training benefits L2 word learning, but only when the cognitive demands of the target words are low (Kelly & Lee, 2012). Similarly, seeing lip movements with speech made it easier for L2 learners to discriminate between phonemes, but adding (here meta-phoric) gestures to audio-visual training actually hindered them (Hirata & Kelly, 2010). The results from Study 2 corroborated the findings from Study 1 in the sense that including gesture in training generally led to speech that was perceived as less accented and more comprehensible. Similarly, in the perception study, this effect differed between the two types of phonemes: Production of /θ/ was perceived as less accented and more comprehensible after AV-P training, but not after AV-I training. For /u/, both types of gesture training led to equal improvement in perception, with lower perceived accentedness and higher perceived comprehensibility.

Taken together, these studies suggest that gestures can benefit L2 phoneme acquisition. Not only can gestures help L2 speakers produce native-like phonemes, but this progress in phoneme production in turn also has a positive effect on speakers' perception with regard to accentedness and comprehensibility. However, the differing findings for the two phonemes under investigation also indicate that this process cannot easily be generalized and that more research is required comparing the use of different gesture types in more and less challenging acquisition contexts.

Acknowledgments
Many thanks to Núria Domínguez, Judith Peters and Nick Theelen for their help in creating the material, data collection, and data analysis.

References

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer (Version 6.0.49) [Computer program]. Retrieved March 2, 2019, from http://www.praat.org/

Derwing, T., & Munro, M. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in second language acquisition, 20*, 1-10.

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

Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (hommage a Adam Kendon). *International Review of Applied Linguistics, 44*, 103-124.

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.

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

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*, 360-378.

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

Hirata, Y., Kelly, S., 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*, 2090-2101.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*(5), 367-371.

Jesney, K. (2004). The use of global foreign accent rating in studies of L2 acquisition. *Calgary, AB: University of Calgary Language Research Centre Reports*, 1-44.

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. https://doi.org/10.1525/collabra.76

Kelly, S., 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*(673).

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.

Kelly, S., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and cognitive processes, 24*, 313-334. doi:10.1080/01690960802365567

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

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

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

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.

Yeo, A., Wagner Cook, S., Nathan, M. J., Popescu, V., & Alibali, M. (2018). Instructor gesture improves encoding of mathematical representation. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 2723-2728). Austin, TX: Cognitive Science Society.

Zhang, Y., Baills, F., & Prieto, P. (2018). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words. *Language Teaching Research*.