



The Effectiveness of Gestures in Learning Novel Sounds

*An exploration of the Benefit of Seeing Gestures for Learning Novel Sounds:
Native Spanish Perceptions of Dutch L2 Spanish pronunciation*

Date: 28-12-2018

Course: LET-CIWM402

Task: Master's Thesis

Supervisor: Dr. M. Hoetjes

Second reader: Dr. F. van Meurs

Student: Nick Theelen

Student number: 4247469

Abstract

The purpose of the present study is to investigate the role of gestures in the acquisition of novel sounds in a foreign language. A large number of studies provide evidence that there is a strong relation between language and gestures, which gives reason to believe that gestures can help in second language (L2) acquisition. On a semantic level, various studies have pointed out that using gestures in L2 acquisition has several beneficial outcomes (e.g. the ability to reproduce words or expressions). However, few studies have focused on the lowest level of language processing, i.e. the phonetic level. The studies that have been conducted focused on the recognition of small distinctions in a foreign language (e.g. vowel length contrasts in Japanese). The present study focuses on the production of the phonemes /θ/ and /u/ in L2 Spanish. More specifically, it was studied whether receiving one of different gesture trainings (audio-visual training, audio-pointing gesture training, or audio-iconic gesture training) has an effect on the phoneme pronunciation of Dutch L2 speakers of Spanish. In an experiment sound fragments of Dutch speakers of Spanish were judged on *comprehensibility* and *accentedness* by native Spanish listeners ($N = 46$). The main findings revealed that the more visual information is provided in the pronunciation training, the more comprehensible and less accented the Dutch L2 speakers of Spanish were judged by native Spanish listeners. One exception was for the Dutch L2 speakers of Spanish who received audio-iconic gesture training of the target phoneme /θ/; native Spanish listeners judged these sounds as least comprehensible and most accented.

1. Introduction

Learning a foreign language is a difficult process. A second language (L2) learner will have to learn an entirely new vocabulary, new conjugations, different sentence structures and often a whole new set of sounds. Some of these aspects are rather easy and a matter of studying and reproducing. For example, conjugations of regular verbs; when an L2 learner remembers the conjugations that belong to specific regular verbs, he/she can apply this to all the regular verbs of that kind. Other aspects can be rather difficult. For example, every language has its unique set of sounds, which is one of the aspects that make one language different from another. Some of these sounds can be similar to existing sounds that a person might have in their vocal inventory, but other sounds are completely new and hence, difficult to pronounce. However, it is important to master these novel sounds. Sato (1991) states that in some circles, particularly employers, an intolerance for foreign accents exists, addressing the importance of mastering the sounds. Additionally, several studies have shown that native speakers tend to downgrade non-native speakers for the sole reason of their foreign accents. Negative consequences could involve being taken less seriously (Clark & Paran, 2007); being perceived as less positive (Hendriks, van Meurs, & Hogervorst, 2016); and less credible (Lev-Ari & Keysar, 2010).

Whilst learning a second language, the ultimate goal of an L2 learner is to communicate successfully with people that are speaking that language. Often, communication is assumed to only consist of verbal aspects. However, non-verbal communication represents a large part of communication as well. It is estimated that around two-thirds (66%) of human interaction is derived from non-verbal cues (Burgoon, Guerrero, & Floyd, 2016). One of those non-verbal aspects of communication is gesture: the symbolic movements that people make with their arms and hands. McNeill describes gestures as “the movements of the hands and arms that we see when people talk” (1992, p. 1). It was also McNeill (1985) who argued that speech (spoken language) and gestures form a tightly integrated system of demonstrating expression (think of the gesture that is used in the sentence “the fish was [this] big”). McNeill’s view got widely accepted - from linguistics to psychology to neurosciences - in scientific research.

Speech-associated gestures (produced whilst talking) are systematically closely related to language and speech (Gullberg, 2006). This is somewhat paradoxical: due to their lack of convention speech-associated gestures appear to be the least language-like of all forms of gesture (the opposite would be sign language which is highly conventionalized). Nonetheless, there is a tight semantic and temporal relation between speech and co-speech gesture. For example, gesture often tends to convey the same (or closely related) meaning to an utterance,

which demonstrates a semantic relation (Gullberg, 2006). Also, most gestures are executed with a timely precision in speech, which demonstrates a temporal relation (Gullberg, 2006). The tight link between speech and gesture is also reflected in a study by Tomasello, Carpenter and Liszkowski (2007) who argue that pointing gestures (part of speech-associated gestures) form the basis of learning a language. Children, at a very young age (even before they can talk), will execute pointing gestures at objects for which they have not yet produced labels. Goldin-Meadow (2007) also provides evidence that pointing gestures do not only form the foundation of language learning, but that these gestures also set the stage for creating language for young deaf children. The studies by Tomasello et al. (2007) and Goldin-Meadow (2007) are based on children learning a native language. One might ask whether gestures could also serve as a beneficial tool for second language (L2) acquisition.

Recently, researchers started investigating the role that gestures could have in L2 acquisition. The focus of most of these studies was on learning vocabulary. Results show that words accompanied by iconic gestures were memorized better than when they were presented without iconic gestures (Kelly, McDevitt, & van Esch, 2009; Tellier, 2008; Allen, 1995). Researchers have also started investigating whether visual information could be beneficial for learning novel L2 sounds, which puts a focus on the phonetic level. Engwall (2012) developed a virtual articulation teacher to improve L2 phoneme pronunciation. The virtual trainings would include an animation of how the novel sounds should be shaped with your lips. The results of his study revealed that pronunciations of certain novel sounds improved with the virtual articulation teacher system. The virtual articulation teacher of Engwall (2012) contained visual information provided by movements made by the lips. One can argue that another type of visual information is provided by gestures. This is where the present study attempts to continue: to explore the role of gestures in L2 phoneme acquisition.

Most previous studies on gesture and L2 learning focused on semantics (learning words). Recently, some work has been done on other aspects of L2 acquisition (mainly perceptions of intonation, vowel length contrasts, and phonemes). For example, Hirata and Kelly (2010) concluded that seeing lip movements significantly helps learners to perceive difficult second-language vowel length contrasts, but visual gestures did not. The findings of Kelly, Hirata, Mansala and Huang (2014) also suggest that gestures may not be well-suited for learning novel phonetic distinctions and the authors argue that the gesture-speech integration may break down at the lowest level of language processing and learning, i.e. at the phonetic level. However, Kelly, Bailey and Hirata (2017) conclude that metaphoric gestures do help in perceiving foreign language speech sounds varying in vowel length and intonation.

The findings of these previous studies on the role of gesture in phoneme acquisition are inconclusive. Moreover, they focus only on part of the acquisition process. The previous experiments measured perceptions by non-native speakers of native speech productions. Native speakers did not have to acquire new sounds. So, previous work did not focus on whether including gesture in language trainings helps people in speech production. Moreover, it is unknown whether any changes in this speech production, due to training with gestures, actually can help people sound less accented. This is where the present study continues: to investigate whether pronunciation by non-native speakers actually can be improved by including gesture in language trainings. To investigate this aspect recordings of non-native speakers will be evaluated by native speakers. More specifically, recordings of Dutch L2 speakers of Spanish will be judged on *comprehensibility* and *accentedness* by native Spanish listeners. To the best of our knowledge this has not been investigated before, filling a gap in scientific research and contributing to the existing knowledge of the relation between L2 acquisition and gestures.

2. Theoretical framework

2.1. Second language acquisition

A second language (L2) can be acquired through a variety of ways, at any given moment in life, and for different purposes. Klein (1986) established six dimensions of second language acquisition that need to be satisfied if a L2 learner wants to be successful in learning a second language. Three of these dimensions characterize the process of acquiring the language: *propensity*, *language faculty*, and *access*. *Propensity* is described as some kind of urge that a learner needs to feel to acquire a new language. It does not have to be clear what causes this propensity, but there must be some kind of inner force that urges the learner to start and continue learning. *Language faculty* encompasses the learner's ability to learn (e.g. intelligence). Also, the learner must have *access* to the language, i.e. to have a tutor, someone who speaks the language, or books to study. Without access to the language the learning capacity of the learner will be severely limited. Each of these dimensions needs to be satisfied if the learner wants to become proficient in a second language.

The other three dimensions of Klein (1986) characterize the process of learning: *structure*, *tempo*, and *end state*. The process of learning a new language will display a certain *structure* in the sense that an L2 learner will follow a certain course. The L2 learner has to become acquainted with all the features of a language and that process will display a certain *structure*. *Tempo* reflects the speed at which the language is acquired and depends on the

propensity, the excellence of language faculty (e.g. some people are good at learning languages, others in mathematics), and the availability (access) to samples of the language. However, there is a certain point at which progression ceases: the L2 learner has reached the *end state*.

In the *structure* phase, an L2 learner is learning all the aspects of a new language: vocabulary, conjugations, irregularities, sentence structures, pronunciation, etc. However, some aspects are more complicated than others. This can be demonstrated by someone who is in the *end state*: this person might significantly surpass the breadth of vocabulary of an average speaker, which is a rather easy aspect of second language acquisition, yet lag significantly in pronunciation, a rather difficult aspect. Pronunciation is a complicated aspect in L2 acquisition because each language has its unique set of sounds. However, obtaining a good pronunciation is important as several studies pointed out that having an accent can have negative consequences for a speaker (Sato, 1991; Clark & Paran, 2007; Hendriks et al., 2016; Lev-Ari & Keysar, 2010). Thus, pronunciation is an aspect of L2 acquisition that should be dealt with adequately to overcome possible negative consequences in L2 communication.

2.2. Gestures

It is generally assumed that L2 learners have at least two goals in their process of acquiring an L2: to communicate successfully and to approximate a native tongue as closely as possible (Van Maastricht et al., 2016). Communication is often assumed to only consist of one aspect: verbal communication. However, a large part of communication is derived from non-verbal cues (Burgoon, Guerrero, & Floyd, 2016). This study focuses on one particular aspect of non-verbal communication: gestures.

Abner, Cooperrider and Goldin-Meadow (2015) describe gestures as visible actions made with the hands to express an utterance. These visible actions could involve: “points, shrugs and nods; illustration of the size, shape and location of objects; demonstration of how to perform actions; depictions of abstract ideas and relationships; and many other everyday communicative actions of the body” (Abner et al., 2015, p. 437). McNeill (1992) established a continuum in which the different types of gestures are categorized, named Kendon’s continuum:

Gesticulation - Language-like gestures - Pantomimes - Emblems - Sign language

In the present study, gestures that fall under the category ‘gesticulation’ are being investigated. These are speech-associated gestures that have no standard or conventionalized form and are produced spontaneously and unwittingly during speech. All gesticulations have at least three phases: the preparation phase (the hands get placed in position), a stroke (the most meaningful and effortful part of the gesture), and the retraction (return or recovery) phase (Gullberg, 2006). The category of gesticulation contains a broad spectrum of gestures which McNeill (1992) categorized in four dimensions: *iconic*, *metaphoric*, *beat*, and *deictic* gestures. These are called ‘dimensions’ because one gesture type could contain aspects of several gesture types (McNeill, 2005; 2006).

Firstly, *iconic* gestures represent a concrete object or event. They are usually used to illustrate or clarify an utterance and, therefore, form a close formal relationship to semantic aspects of speech (McNeill, 1992). Secondly, *metaphoric* gestures are similar to iconic gestures, but instead of representing a concrete object or event, metaphoric gestures represent something abstract, like a concept or idea (McNeill, 1992). Metaphoric gestures are then used to make this abstract concept or idea more concrete. Thirdly, *beat* gestures are produced by simple hand movements (up and down) according to the rhythm of speech. Beat gestures are often used to indicate one part or several parts of an utterance that are particularly important or relevant (Krahmer & Swerts, 2007). Lastly, McNeill (1992) defines *deictic* gestures as a dimension within gesticulation. Deictic gestures are pointing gestures, which can refer to abstract as well as concrete objects. It is used at any time when someone wants to locate something. Deictic gestures are usually executed with arms or hands, but in some cultures other body parts may be used, such as the head or even the lips (Enfield, 2001).

The dimensions described above refer to the left-hand side of Kendon’s continuum, gesticulation, also called co-speech-, speech-accompanying gestures, or gestures for short. Moving to the right on Kendon’s continuum, spoken language becomes less important as the movements with arms and hands start to have more linguistic properties, which means that more standardised and conventionalised gestures for communication have been established. The category next to gesticulation is *language-like gesture*. These are gestures that replace words in utterances (McNeill, 2006). For example, when a gesture is produced instead of the word ‘hit’ in the sentence “and he [...] a homerun”. The following category on the continuum is *pantomime*. These are gestures that communicate a meaning or even an entire story without the use of speech. *Emblems* are gestures that are culturally specific with a fixed form and meaning (Wagner, Malisz, & Kopp, 2014). For example, the Dutch emblem to indicate that something tastes good, by waving one’s (left) hand next to one’s (left) ear. The last category on Kendon’s

continuum is sign language, which is a language used mainly by deaf people. In this category speech is completely replaced by standardised and conventionalised gestures for communication.

The present study focuses on the left side (gesticulations) of Kendon's continuum, which, paradoxically, seems to have the tightest relation with spoken language of all categories on the continuum. Paradoxically because this is the least standardised and conventionalised category.

2.3. Relation between language and gesture

Previous research in the field of language and speech-associated gesture evidently shows that there is a strong relation between the two. Gullberg (2006), for example, explains that people usually speak when they produce gestures, not when they are silent. Moreover, in several contexts language requires gesture to be (more) meaningful (e.g. asking for a drink in a restaurant only by making an imaginary cup, lifting it to your lips and in some cases raise your eyebrows). Speech-associated gestures also portray a close semantic and temporal relationship with speech (Kendon, 1974): gestures that are being produced tend to express the same (or closely related) meaning as the produced speech, at the same time.

Tomasello et al. (2007) argue that gestures serve as a stepping stone to the production of lexical items and sentences (learning language) for young children. Young children will produce pointing gestures for objects they have not yet learned the word for, and combining a pointing gesture together with words gives them the opportunity to express a sentence-like meaning before they are capable of expressing those meanings entirely in speech. Goldin-Meadow (2007) argues the same point as Tomasello et al. (2007) and adds that gestures do not only serve as a stepping stone for learning language, but also set the stage for creating language. Young deaf children, whose hearing impairment is so severe that they cannot learn to speak, use gestures to communicate even before they have been exposed to a usable model for language, and thus create their own language. These gestures are structured in such language-like ways, that this type of gesture has been labelled as "homesigns".

2.3.1. Gesture and L2 acquisition

The tight relation between speech and gesture gives reason to believe that gestures can help in L2 acquisition. Recently, researchers started investigating the role of gestures in L2 acquisition. Most of these studies focused on semantics (e.g. vocabulary). Tellier (2008) examined the

impact of gestures on the capacity to reproduce English words by young French children. In her experiment, French children learned eight English words. One group of children (N=10) were taught with pictures and the other group (N=10) were taught with pictures accompanied by iconic gestures. Results demonstrated that the group with gestures significantly improved the memorization of the lexical items relative to the group without gestures. In an experiment by Allen (1995) on the effect of gesture on the development and access of mental representations, French students had to recall French expressions. Three groups of French students participated in a pre and post-test. One group learned gestures and used them to recall expressions in the post-test, one group did not see any gestures, the third group was a comparison group that only learned gestures in the post-test. Results demonstrated that groups who were exposed to gesture forgot significantly fewer expressions than the group who was not exposed to gesture. Moreover, in an experiment by Kelly et al. (2009), adults from the US received a brief training session where they would learn Japanese verbs either with or without gesture. Three sets of memory tests (at five minutes, two days, and one week) showed that the greatest amount of verbs were remembered when the verbs were learned with gestures that conveyed iconic information.

The studies by Allen (1995), Tellier (2008) and Kelly et al. (2009) demonstrate that for children, young adults, and adults gestures improved the ability to recall items better than being taught without gestures; they evidently serve as a beneficial tool in L2 acquisition. One might ask whether gestures can also help for pronunciation of these words, i.e. pronouncing novel sounds.

2.3.2. Gesture and L2 phoneme acquisition

Most research that studied the relation of L2 acquisition and gestures focused on semantics (vocabulary). Recently, some studies have investigated other aspects of L2 acquisition and gestures (mainly perceptions of intonation and vowel length contrasts). Amongst some others, most research on this topic has been conducted by Kelly, Hirata, and colleagues (Hirata & Kelly, 2010; Hirata, Kelly, Huang, & Manansala, 2014; Kelly et al., 2014; Kelly et al., 2017). Their results, however, are rather inconclusive.

In Hirata and Kelly (2010), the authors explored whether multimodal information, such as lip movements and gestures, helps to improve the ability to perceive Japanese vowel length contrasts. It is important to distinguish different vowel lengths in Japanese, as different vowel lengths can change the meaning of a word completely. In their experiment, six native English speakers received a specific training, either: audio-only, audio-mouth, audio-hands, or audio-

mouth-hands. Before and after training, the participants conducted phoneme perception tests to measure their ability to identify short and long Japanese vowel lengths. The findings of the study suggested that seeing lip movements significantly helped in identifying Japanese vowel length contrasts. However, seeing gestures did not help.

Hirata et al. (2014) investigated the influence of gesture in auditory learning of an L2 at the segmental phonology level. They researched whether vowel length contrasts in Japanese words (between word-initial and word-final and between slow and fast speaking rates) can be identified by students from the US. In their experiment 88 native English-speaking participants took an auditory test before and after certain trainings involving gestures. There were four different trainings in which participants would either observe an instructor in a video speaking Japanese words while making syllabic-rhythm gestures, or participants would observe an instructor in a video speaking Japanese words while making moraic-rhythm gestures (a 'mora' is the Japanese equivalent of a 'syllable', it distinguishes vowel lengths in Japanese words). The other two trainings consisted of not only observing, but also producing either syllabic-rhythm or moraic-rhythm gestures, together with the instructor in the video. All of the training types yielded a similar improvement in identifying Japanese vowel length contrasts. However, the effects were rather small; the researchers concluded that the overall effect of gesture on learning segmental phonology is limited.

Kelly et al. (2014) explored the role of gestures in learning novel Japanese phoneme contrasts and vocabulary. In their experiment 88 undergraduate students from a university in the US were taught two components of the Japanese language (phoneme contrasts and vocabulary items) under different gesture conditions. Half of the sample only observed the gestures, the other half also produced the gestures that were observed along with the speech. The main finding revealed no major differences across all condition, suggesting that gestures may not be well suited for learning novel phonetic distinctions.

The main findings in Hirata and Kelly (2010), Hirata et al. (2014), and Kelly et al. (2014) all suggested that gestures may not be well suited for learning novel phonetic distinctions. Therefore, in Kelly et al. (2014) they say that gesture-speech integration may break down at the lowest level of language processing, the phonetic level. However, in Kelly et al. (2017) contrasting results were found.

In the study by Kelly et al. (2017), English speaking adults from the US listened to Japanese vowel length contrasts and sentence-end intonation distinctions. After a tutorial about vowel lengths and sentence-end intonations in the Japanese language in the form of PowerPoint slides, participants were asked to watch videos accompanied by either congruent, incongruent

or no gestures. In total there were 24 videos for each participant. For intonational contrasts, identification of the intonations was more accurate when accompanied by congruent gestures than incongruent or no gestures. For vowel length contrasts no clear pattern was detected. Kelly et al. (2017) conclude that metaphoric gestures are beneficial, to some extent, in recognizing novel sounds.

In Kelly et al. (2017), evidence was found that gestures do help, to some extent, in perceiving novel sounds. However, it remains unknown whether gestures can help in L2 phoneme production, and whether possible effects on phoneme production can also be perceived by native speakers.

2.4. Evaluating L2 phoneme acquisition

To better understand the effects of accents in L2's various studies have been conducted. In these studies constructs such as *likeability*, *comprehensibility*, *nativeness*, *accentedness*, and *intelligibility* are often measured (Derwing & Munro, 2005; Hendriks, van Meurs & Hogervorst, 2016; Van Maastricht, Kraemer, & Swerts, 2016). In this study it is also attempted to explore the effects of non-native speakers producing sounds in a foreign language. Moreover, in this study it is investigated whether a gesture training has an effect on pronunciations by L2 speakers. To investigate this, recordings of individual words will be evaluated. This will be done by using some of the constructs that are frequently used in perception studies.

Van Maastricht et al. (2016) stated that L2 learners have at least two goals: to communicate successfully in the L2 and to approximate a native tongue as closely as possible, i.e. to be able to speak clearly and be understood by someone in the L2. Derwing and Munro (2005) refer to *comprehensibility* as “the listener’s perception of the degree of difficulty encountered when trying to understand an utterance”. This refers to the first goal an L2 learner has while learning an L2: successful communication is achieved when a listener faces no difficulties in understanding an utterance by an L2 speaker. Therefore in the present study *comprehensibility* will be measured. Moreover, the second goal an L2 learner has, is to approximate a native tongue as closely as possible. This is also an important aspect in L2 acquisition as various studies pointed out that having an accent in a foreign language can have negative consequences for an L2 speaker (Sato, 1991; Clark & Paran, 2007; Lev-Ari & Keysar, 2010; Hendriks, van Meurs, & Hogervorst, 2016). Therefore, it will be interesting to investigate if gesture trainings can help L2 speakers sound less accented. For this reason *accentedness* is the other construct that will be measured in the present study. Derwin and Munro (2005) describe *accentedness* as “how much an L2 accent differs from the variety of speech commonly

spoken in the community”. One might argue that accents are difficult to evaluate based on solely a single word (which will be done in the experiment of the present study). However, Munro, Derwing and Burgess (2003) state that accents are very easily detected by natives, even if a single word is presented backwards.

The other constructs that are often used in perception studies (*likeability, intelligibility, nativeness*) will not be taken into account in the present studies as these will be difficult to measure considering the experiment of the present study (*likeability, intelligibility*) or measure more or less the same thing (*nativeness*). Therefore, in the present study the constructs *comprehensibility* and *accentedness* will be used as measurements to investigate whether gestures have an effect on the pronunciation of L2 phoneme productions.

2.6. The present study

A considerable number of studies have been published on language and co-speech gesture. They conclude that language and co-speech gestures are strongly related and form an integrated system of expressing meaning (McNeill, 1992); language requires gestures to become (more) meaningful, and in some contexts gesture can even completely replace language (Gullberg, 2006). This made researchers think about the actual value gestures could present to us, i.e. if gesture is so interconnected with speech, it might help in L2 acquisition. Several studies pointed out that gestures do, in fact, help people to memorize expressions or words better (Allen, 1995; Tellier, 2008; Kelly et al., 2009). Moreover, gestures also help people to identify phonemic distinctions in vowel lengths and intonations (Kelly et al., 2017).

To summarize, co-speech gesture is very closely linked to speech, it increases the ability to reproduce foreign language at a semantic level, and it also increases the ability to identify small differences in foreign languages at a phonetic level. However, an aspect that has not yet been studied is whether gestures can help in the *production* of phonetic aspects of foreign languages, i.e. improve L2 pronunciation.

The present study aims to explore whether non-native speakers can improve their pronunciation with the help of gestures. As this has not been studied before, it will fill a gap in scientific research. Besides this, it will contribute to existing knowledge of the relation between L2 acquisition and gestures. Should gestures significantly improve pronunciation, multiple practical implications might apply, e.g. teaching second languages at schools.

To investigate whether non-native speakers can improve their pronunciation by means of including gesture in their pronunciation training, native speakers will evaluate recordings of non-native speakers. More specifically, recordings of Dutch people speaking Spanish will be

judged on *comprehensibility* and *accentedness* by native speakers of Castilian Spanish. The following research question has been established:

RQ: “To what extent do native speakers of Castilian Spanish respond differently to recordings of Dutch speakers of Spanish who were trained on phoneme production under the condition with/without gestures in terms of:

1A. the degree of comprehensibility;

1B. the degree of accentedness?”

3. Method

3.1. Materials

In the present study native speakers of Spanish judged sound fragments produced by Dutch L2 speakers of Spanish. The sound fragments consisted of Dutch speakers producing a Spanish word containing either the phoneme /θ/ or /u/, whose graphemes (z and u) have a different pronunciation in Spanish than in Dutch. The Dutch L2 speakers of Spanish who produced the sound fragments did not have any knowledge about the Spanish language.

The sound fragments were gathered from a previous production study. This production study was set up as follows. The recordings were taken at two different times. Once at the start of the experiment and once after a short training. Dutch participants were asked to read aloud 16 Spanish sentences containing words with the target sounds. Figure 1 shows examples of sentences that were presented in the production study (from left to right: “*the fox is nice*”, “*the route is long*”). After reading the sentences, participants would receive a short training, after which they were asked to read the sentences again. Trainings consisted of an explanation about how to produce the phonemes /θ/ and /u/, given with examples, the type of which depended on the condition that the participant was in. Participants were randomly assigned to one of four different training conditions (between participants design): audio-only condition, audio-visual condition, audio-pointing gesture condition, audio-iconic gesture condition. The audio-only condition only presented examples of the target sounds auditorily, while in the audio-visual condition a native speaker of Spanish was visible on the screen who produced the target sounds. The audio-pointing gesture condition portrayed the speaker executing a pointing gesture to her lips during the production of the relevant phoneme. Lastly, in the audio-iconic gesture condition the speaker produced an iconic gesture about the target sound, while producing the relevant phoneme. This iconic gesture represented the articulatory gesture needed for correct phoneme

production, as described in the training. The recordings that were used in the present study consisted of the data collected in some of these conditions, namely: pre-training, audio-visual condition, audio-pointing gesture condition, and audio-iconic gesture condition. The condition ‘audio-only’ was left out, because it does not mimic a multimodal classroom setting, which the other three conditions do have. The pre-training is included as a baseline, to investigate whether there is an effect of training at all.

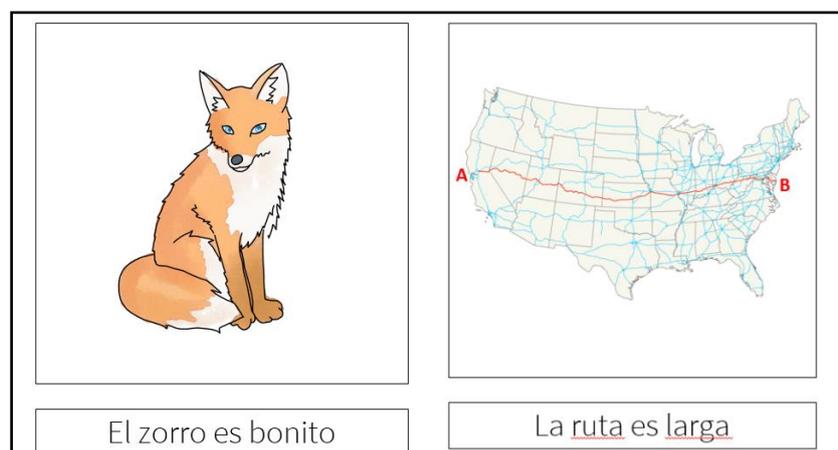


Figure 1. Examples of stimuli from the production study

In the current experiment, a selection of the recordings from the production study was used. In the production study 16 sentences were read aloud by the participants, only eight of those sentences contained target words (four with the target phoneme /θ/ and four with target phoneme /u/, with the phonemes in the same position of the word), the rest were filler items. The selection of words in the present study consisted of only the target words, which have the target phoneme at the same position in the words: *zorro*, *zona*, *zueco*, *zeta*, *ruta*, *suma*, *muro*, and *nube*. The data (the recordings) were annotated using the program PRAAT. Single words containing the target sounds were cut out (using PRAAT) of the sound files and presented in an online survey in Qualtrics. In the online survey participants judged one word after another with regard to *comprehensibility* and *accentedness*. The items included data from 21 speakers (seven speakers for each type of training condition: audio-visual condition, audio-pointing gesture condition, and audio-iconic gesture condition) and eight items per speaker (four target words - two with /θ/ and two with /u/ - from the pre-training and the same four words from the post-training), resulting in 168 items. Not all eight target words were used per speaker as this would make the experiment too long.

3.2. Subjects

A total number of 46 subjects participated in this study. Almost all of them were from Spain (45), there was one Mexican (this person immigrated at a very young age to Spain, therefore she was not excluded from the data analyses). The subjects were not all from the same region in Spain, but rather scattered over the country. There were subjects from Catalonia, Basque country, Andalucía, Madrid area and Valencia. In these areas exist dialects which differ from Castilian Spanish. However, Castilian Spanish is the main language of the country and all subjects have this language in their vocabulary. There were 25 subjects who judged the sound fragments on *comprehensibility*, the other 21 subjects judged the sound fragments on *accentedness*. The average age of the subjects was 30.8 years with a range between 19 and 70, more than half of the subjects were male (58.7%). The most frequent educational level was ‘Educación Superior: Universidad - Grado’, the second highest attainable educational level in Spain.

3.3. Design

The experiment had a 4 (type of training: pre-training, audio-visual condition, audio-pointing condition, audio-iconic condition) x 2 (target sound: /u/ or /θ/) within-subjects design. All subjects were exposed to all of types of training and target sounds. The subjects were asked to perform one task which measured the dependent variable, subjects either judged:

1. the comprehensibility of the audio fragment, or;
2. the accentedness of the audio fragment.

3.4. Instruments

Listeners participated in an online experiment in which they judged speakers’ pronunciation on either *comprehensibility* or *accentedness*.

Comprehensibility was measured with the statement “Es fácil entender a esta persona” (“*I find this speaker easy to understand*”) followed by a 7-point Likert scale anchored by “absolutamente en desacuerdo - totalmente de acuerdo” (“*totally disagree - totally agree*”) (based on Derwin and Munro, 1997).

Accentedness was measured with the statement “Esta persona habla ...” (“*This speaker speaks ...*”), followed by a 7-point semantic differential anchored by “sin acento extranjero – con un acento extranjero fuerte” (“*without a foreign accent - with a strong foreign accent*”) (based on Jesney, 2004).

3.5. Procedure

In a period of a week (November - December, 2018) subjects were asked to take part in the online experiment. Subjects were gathered via social media (WhatsApp, Facebook, e-mail). Subjects were motivated to take part in this experiment by the chance of winning a lottery ticket for 'El Gordo', the Christmas lottery in Spain from 'La lotería Nacional'. Once arrived at the welcome page of the online experiment, subjects would receive a brief overview of what they were going to do. It was mentioned that only native speakers of Castilian Spanish could participate, and that they were going to judge pronunciations on either *comprehensibility* or *accentedness*. As these constructs might be difficult to understand a brief explanation of the two constructs was given. Also, it was mentioned that it was crucial to use headphones, because the subjects had to judge small differences in pronunciations in single words. The experiment was conducted on an individual basis and on average took about 15 minutes to complete.

3.6. Statistical treatment

In this experiment subjects judged sound fragments which contained three different conditions (audio-visual, audio-pointing, audio-iconic) with pre-training as baseline and two target sound (/θ/ and /u/). The subjects were exposed to all types of training conditions and target sounds. To compare all the means of all conditions a factorial repeated-measures ANOVA analysis was conducted to investigate whether there was an effect of (type of) training and target sound on *comprehensibility* and *accentedness*.

4. Results

4.1. Comprehensibility

Mauchly's test of sphericity indicated that the assumption of sphericity had been violated for the main effect of 'type of training' ($\chi^2(5) = 14.51, p = .013$). Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .68$ for the main effect of type of training). A repeated measures analysis for comprehensibility with 'type of training' and 'target sound' as within-subject factors showed a significant main effect for 'type of training' ($F(2.04, 40.72) = 10.26, p < .001, \eta^2 = .34$), but no significant main effect for 'target sound' ($F(1, 20) < 1, p = .749$). Contrasts revealed that ratings of words produced in pointing gesture trainings ($F(1, 20) = 18.76, p < .001, \eta^2 = .48$) and iconic gesture trainings ($F(1, 20) = 5.61, p = .028, \eta^2 = .22$) were significantly higher than words produced in the pre-training, with regard to ratings on comprehensibility. Means and standard deviations can be seen in Table 1.

The repeated measures analysis for comprehensibility also showed a significant interaction effect between the 'type of training' and 'target sound' ($F(3, 60) = 10.74, p < .001, \eta^2 = .35$). This indicates that the target sounds had different effects on people's ratings on comprehensibility, depending on which type of training was used. To break down this interaction, contrasts were performed comparing all trainings to their baseline (pre-training) and both target sounds. This revealed a significant interaction when comparing the target phoneme /θ/ to the target phoneme /u/ for iconic gesture training compared to pre-training ($F(1, 20) = 19.01, p < .001, \eta^2 = .49$). Looking at Figure 1, this effect reflects that the target phoneme /θ/ (compared to target phoneme /u/) lowered scores significantly more in iconic gesture training than it did for pre-training. The remaining contrasts revealed no significant interaction terms when comparing visual gesture training to pre-training ($F(1, 20) < 1, p = .413, \eta^2 = .03$), and pointing gesture training to pre-training ($F(1, 20) < 1, p = .527, \eta^2 = .02$).

Table 1. Means and standard deviation (between brackets) for comprehensibility (1 = not comprehensible at all, 7 = very comprehensible) of target sounds under different trainings

Type of training:		Pre-training <i>M (SD)</i>	Audio-visual training <i>M (SD)</i>	Pointing gesture training <i>M (SD)</i>	Iconic gesture training <i>M (SD)</i>
Target sounds:	/u/	4.35 (.89)	4.33 (.77)	4.76 (.98)	4.79 (.70)
	/θ/	4.45 (.89)	4.55 (.92)	4.79 (.96)	4.31 (1.01)
Total:		4.40	4.44	4.77	4.55

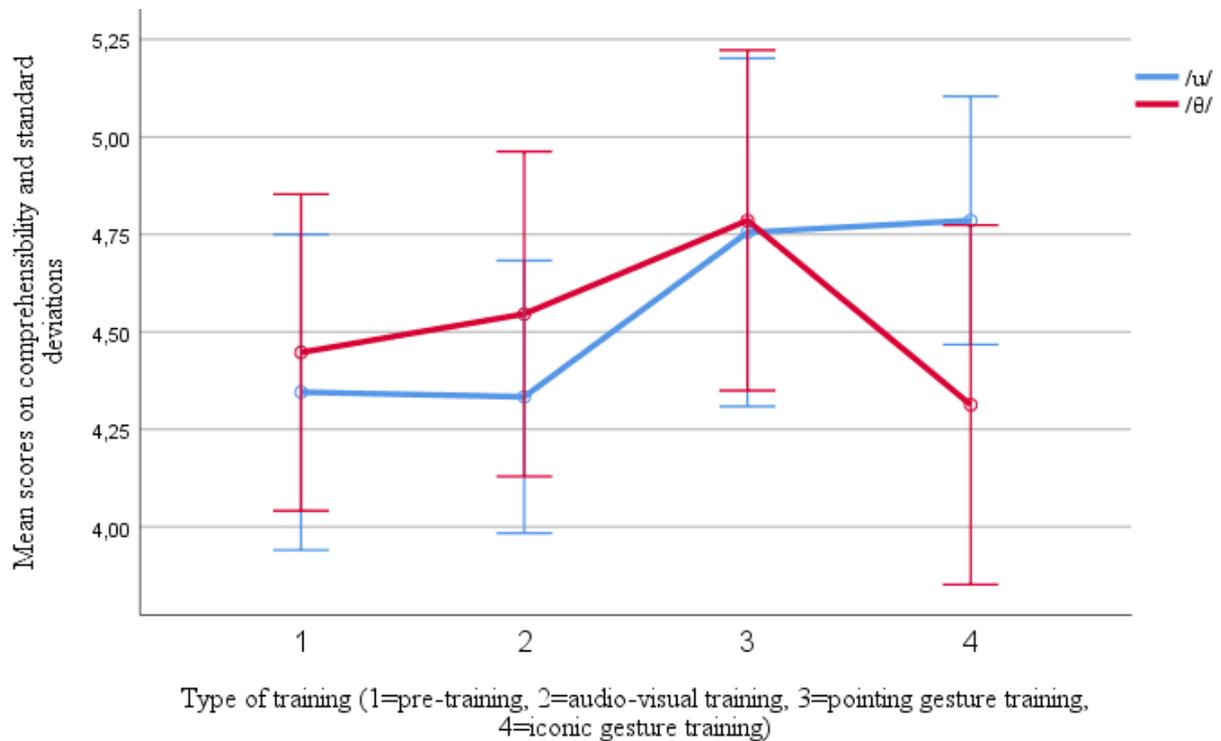


Figure 2. Line graph of mean scores for comprehensibility (1=not comprehensive at all, 7=very comprehensive) on target sounds /u/ and /θ/ produced under different training types

4.2. *Accentedness*

A repeated measures analysis for accentedness with ‘type of training’ and ‘target sound’ as within-subject factors showed a significant main effect for ‘type of training’ ($F(3, 72) = 16.17$, $p < .001$, $\eta^2 = .40$), but no main effect for ‘target sound’ ($F(1, 24) < 1$, $p = .606$). Contrasts revealed that ratings of words produced in pointing gesture trainings ($F(1, 24) = 27.07$, $p < .001$, $\eta^2 = .50$) and iconic gesture trainings ($F(1, 24) = 11.25$, $p = .003$, $\eta^2 = .32$) were significantly lower than words produced in the pre-training, with regard to ratings on accentedness. Means and standard deviations can be seen in Table 2.

The repeated measures analysis for accentedness also showed a significant interaction effect between the ‘type of training’ and ‘target sound’ ($F(3, 72) = 12.94$, $p < .001$, $\eta^2 = .35$). This indicates that the target sound had different effects on people’s ratings on accentedness, depending on which type of training was used. To break down this interaction, contrasts were performed comparing all trainings to their baseline (pre-training) and both target sounds. These revealed significant interactions when comparing the target phoneme /θ/ to the target phoneme /u/ both for visual gesture training compared to pre-training ($F(1, 24) = 11.68$, $p = .002$, $\eta^2 = .33$) and iconic gesture training compared to pre-training ($F(1, 24) = 19.60$, $p < .001$, $\eta^2 = .45$). Looking at Figure 3, these effects reflect that the target sound /θ/ (compared to target sound /u/) lowered scores significantly more in visual gesture training than it did for pre-training, and increased scores significantly more for iconic gesture training than it did for pre-training. The remaining contrast revealed no significant interaction term when comparing pointing gesture training to pre-training ($F(1, 24) < 1$, $p = .746$).

Table 2. Means and standard deviation (between brackets) for accentedness (1 = without a foreign accent, 7 = very strong foreign accent) of target sounds under different trainings

Type of training:	Pre-training <i>M (SD)</i>	Audio-visual training <i>M (SD)</i>	Pointing gesture training <i>M (SD)</i>	Iconic gesture training <i>M (SD)</i>
Target sounds: /u/	4.56 (.84)	4.73 (.82)	4.10 (1.07)	4.11 (.91)
Target sounds: /θ/	4.48 (.77)	4.29 (.83)	3.96 (.90)	4.55 (.80)
Total:	4.52	4.51	4.02	4.33

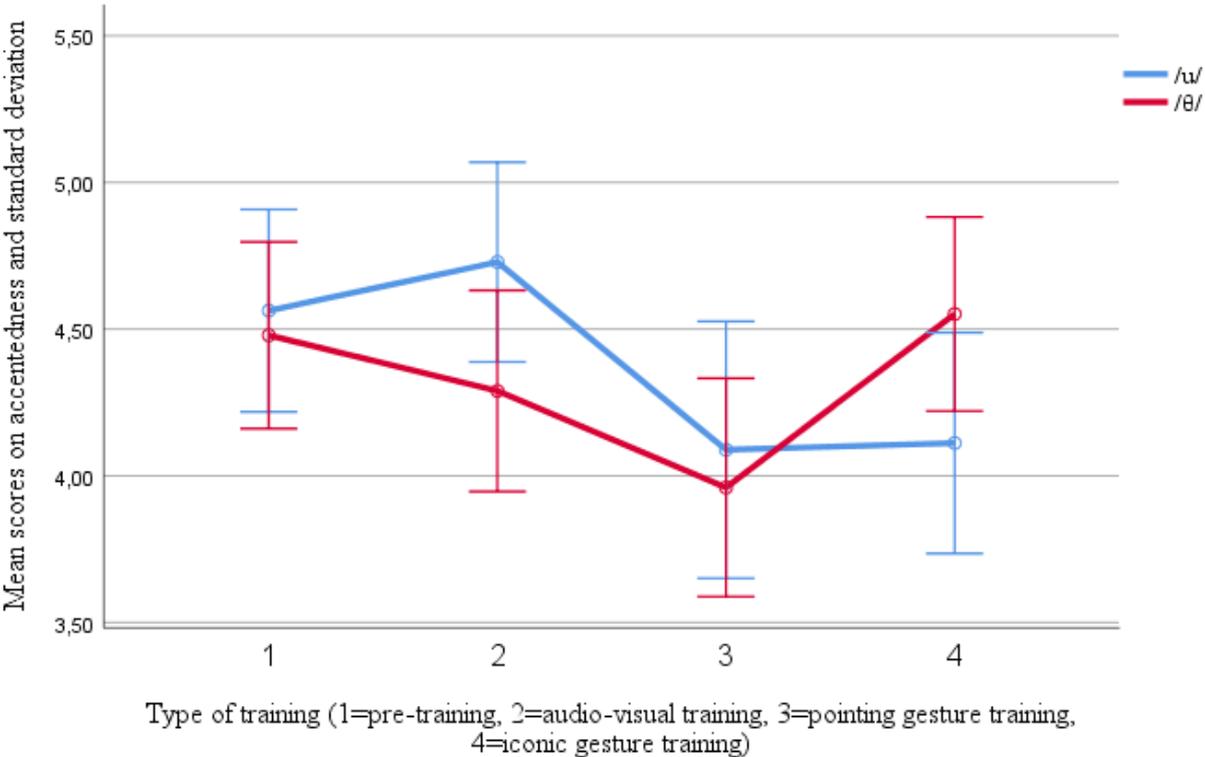


Figure 3. Line graph of mean scores for accentedness (1=without a foreign accent, 7=very strong foreign accent) on target sounds /u/ and /θ/ produced under different training types

5. Conclusion

The purpose of this study was to investigate whether different types of pronunciation trainings (audio-visual, audio-pointing gesture, audio-iconic gesture) have an effect on evaluations by native speakers on the pronunciation of L2 speakers with regard to *comprehensibility* and *accentedness*. The first point of interest was to investigate whether different types of training conditions affect judgements by native speakers on comprehensibility and accentedness. The second point of interest was to investigate whether there is a difference in judgements in terms of comprehensibility and accentedness for the target sounds /θ/ and /u/.

First of all, findings of the present study indicate that, on the whole, the more visual information was presented to a Dutch L2 speaker of Spanish in a training condition, the more comprehensible and less accented he/she was judged by native Spanish listeners. The exception to this was the iconic gesture training condition of the target sound /θ/. The pronunciation of this consonant appeared to be rather difficult in comparison with the vowel /u/ for Dutch L2 speakers of Spanish, apparent by lower scores on comprehensibility and higher scores on accentedness. It may be that the production of an iconic gesture for the target phoneme /θ/ conveyed too much visual information and hereby distracted the Dutch L2 speakers from the already difficult task.

Next, a conclusion of the other findings is given. For the sake of clarity the results of the two constructs will be discussed separately, starting with *comprehensibility*. The research question of this study was:

“To what extent do native speakers of Castilian Spanish respond differently to recordings of Dutch speakers of Spanish who were trained on phoneme production under the condition with/without gestures in terms of:

1A. the degree of comprehensibility;

1B. the degree of accentedness?”

The findings of this study revealed that native speakers of Castilian Spanish do respond differently to recordings of Dutch L2 speakers of Spanish in terms of comprehensibility. In this study the focus was put on two target phonemes, namely: the vowel /u/ and the consonant /θ/. For the vowel /u/, Dutch L2 speakers of Spanish who received the audio-pointing gesture training or the audio-iconic gesture training were judged as most comprehensible by the native Spanish listeners, with a marginal advantage for the speakers from the audio-iconic gesture

training condition. However, for the consonant /θ/, the Dutch L2 speakers of Spanish who received the audio-pointing gesture training were judged as most comprehensible, but the speakers who received the audio-iconic gesture training were judged least comprehensible of all by the native Spanish listeners. This phenomena gets explained by the interaction effect that was found between the gesture training condition and the target sounds. It indicated that there was a significant difference in how the Spanish subjects responded to the target sounds /θ/ and /u/ when these sounds were produced after receiving the audio-iconic gesture training, compared to the baseline (pre-training) condition. For the vowel /u/, Dutch L2 speakers of Spanish were judged significantly more comprehensible by native Spanish listeners than for the consonant /θ/.

Moving on to accentedness, the research question for accentedness was:

“To what extent do native speakers of Castilian Spanish respond differently to recordings of Dutch speakers of Spanish who were trained on phoneme production under the condition with/without gestures in terms of:

1A. the degree of comprehensibility;

1B. the degree of accentedness?”

The findings of this study revealed that native speakers of Castilian Spanish, same as for comprehensibility, do respond differently to recordings of Dutch speakers of Spanish in terms of accentedness. The findings were in line with the findings for comprehensibility. For the vowel /u/, Dutch L2 speakers of Spanish who received the audio-pointing gesture training or the audio-iconic gesture training were judged as least accented by the native Spanish listeners, with a marginal advantage for the speakers who received the audio-pointing gesture training. However, for the consonant /θ/, Dutch L2 speakers of Spanish who received the audio-pointing gesture training were judged least accented, but the speakers who received the audio-iconic gesture training were judged as most accented by the native Spanish listeners. For Dutch L2 speakers of Spanish who received the audio-visual training an improvement in their accentedness was detected for the consonant /θ/ relative to the baseline (pre-training) condition. However, for the vowel /u/, the opposite appeared to be true; after receiving the audio-visual training the Dutch L2 speakers of Spanish were judged more accented by the native Spanish listeners relative to the baseline (pre-training) condition. The differences in judgements by the native Spanish listeners between the audio-visual training and the audio-iconic gesture training condition can be explained by the interaction effect that was found. For the vowel /u/, Dutch

L2 speakers of Spanish were judged significantly more accented by native Spanish listeners than for the consonant /θ/ when these phonemes were produced after receiving the audio-visual training, compared to phonemes produced in the baseline (pre-training) condition. For the Dutch L2 speakers of Spanish who received the audio-iconic gesture training, the vowel /u/ was judged to be significantly less accented by native Spanish listeners than the consonant /θ/, compared to the baseline (pre-training) condition.

To conclude, according to the findings of the present study, the words that were produced by the Dutch L2 speakers of Spanish after receiving the audio-pointing gesture training were considered as most comprehensible and less accented. This holds for vowel phoneme (/u/) as well as for the consonant (/θ/). The words containing the target consonants in the experiment tended to be more difficult to pronounce as the words containing the target consonant, overall, were judged as less comprehensible and more accented by the native Spanish listeners relative to the words containing the vowel sound.

6. Discussion

The findings of this study are rather difficult to compare with previous studies, as what has been researched in this study, was not done before in this manner. However, the findings of the present study can be related to previous studies in this field on the relation between second language (L2) acquisition and gestures.

The findings of this study indicate that including gesture in L2 pronunciation trainings can, in fact, help people to improve their pronunciation in an L2. The significant effects that were found show, as in numerous previous studies, the existence of the relation between speech and co-speech gestures (McNeill, 1985; 1992; Allen, 1995; Gullberg, 2006; Tomasello et al, 2007; Tellier, 2008; Kelly et al, 2009; 2017). The relation between speech and co-speech gesture was already well established on a semantic level (Allen, 1995; Tomasello et al, 2007; Tellier, 2008; Kelly et al, 2009). However, very little research has been done on the relation between speech and co-speech gesture on a phonetic level. The studies that have been conducted, are rather inconclusive. Hirata and Kelly (2010), Hirata et al. (2014), and Kelly et al. (2014) did not find a relation between speech and co-speech gesture on a phonetic level, which is in contrast with the findings of the present study. They suggested that gesture-speech integration may break down at the lowest level of language processing. However, Kelly et al. (2017) did find a relation between co-speech gesture and speech at the lowest level of language processing, arguing that, also at the lowest level of language processing, speech and co-speech

gestures are related to each other and exist together in an integrated system of expressing meaning. This is in line with the findings of the present study, where gesture trainings significantly improved *comprehensibility* and *accentedness* of L2 speakers on two target phonemes.

6.1. Limitations and suggestions for future research

The present study investigated whether the pronunciation of two L2 target phonemes (/u/ and /θ/) could be improved by the use of gesture trainings. However, in the experiment that was conducted entire words containing the target sounds were presented and judged by the native Spanish listeners. Therefore, construct validity is at stake; it cannot be said with certainty whether judgements of Spanish listeners on comprehensibility and accentedness were caused only by the actual target sound, or that maybe other phonemes or syllables affected the judgements of the native Spanish listeners. Future research could present the phonemes in isolation instead of within an entire word. However, doing so would jeopardize the ecological validity; it will no longer be representative to a ‘real-life’ situation.

Moreover, only one vowel (/u/) and one consonant (/θ/) were investigated in the present study. It provides us with a start, but there are many more phonemes that can be investigated. This is a suggestion for future research, to investigate whether the effectiveness of gesture training holds for other vowels and consonants as well.

Next, several significant effects were found in the present study. However, the differences in the means between the training conditions were rather small. Future research could focus on a larger sample size to provide more certainty to the means and replicate the results.

Lastly, the present study conducted an experiment with a fairly small group of subjects (N = 46) which were widely spread over Spain. For example, there were subjects who normally speak the Catalan or Basque dialect, others were from Madrid or Malaga. This might have affected the results. To find out whether origin plays a role with regard to judging sound fragments of L2 speakers, it is suggested to gather a large sample size from one specific region or multiple large groups originated from specific regions. However, even with a relatively small number of subjects from a widely scattered demographic area (within Spain), significant effects were found. This indicates that a strong relation exists. A larger sample size would provide more strength to the results.

6.2. Implications

There are a number of implications that come with the present study. Firstly, it contributes to the existing scientific knowledge about the relation between speech and co-speech gestures. Speech and co-speech gesture seem to relate to each other at every level of language processing, from semantics to phonetics. Secondly, the present study filled a gap in scientific literature. The effect of gesture trainings on phoneme production of L2 speakers judged by native speakers is an area that had not been studied before. It is important to widen our knowledge in this field as it could help people in mastering novel sounds in foreign languages more easily. Pronunciation is a rather difficult aspect of L2 acquisition, and more efficient ways to teach this aspect will be beneficial for L2 learners. Therefore, a practical implication of the present study is that teaching methods could be adjusted more efficiently. More research is needed, but findings of this study suggest that the use of gesture can help L2 learners to improve their pronunciation in a second language.

7. References

- Abner, N., Cooperrider, K., & Goldin-Meadow, S. (2015). Gesture for linguists: A handy primer. *Language and Linguistics Compass*, 9(11), 437-451.
- Allen, L. Q. (1995). The effects of emblematic gestures on the development and access of mental representations of French expressions. *The Modern Language Journal*, 79(4), 521-529.
- Burgoon, J. K., Guerrero, L. K., & Floyd, K. (2016). *Nonverbal communication*. London and New York: Routledge.
- Clark, E., & Paran, A. (2007). The employability of non-native-speaker teachers of EFL: A UK survey. *System*, 35(4), 407-430.
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *Tesol Quarterly*, 39(3), 379-397.
- Enfield, N. J. (2001). 'Lip-pointing': A discussion of form and function with reference to data from Laos. *Gesture*, 1(2), 185-211.
- Engwall, O. (2012). Analysis of and feedback on phonetic features in pronunciation training with a virtual teacher. *Computer Assisted Language Learning*, 25(1), 37-64.
- Goldin-Meadow, S. (2007). Pointing sets the stage for learning language—and creating language. *Child Development*, 78(3), 741-745.
- Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon). *IRAL-International Review of Applied Linguistics in Language Teaching*, 44(2), 103-124.
- Hendriks, B., Van Meurs, F., & Hogervorst, N. (2016). Effects of degree of accentedness in lecturers' Dutch-English pronunciation on Dutch students' attitudes and perceptions of comprehensibility. *Dutch Journal of Applied Linguistics*, 5(1), 1-17.
- 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.
- 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.
- 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. DOI: <http://doi.org/10.1525/collabra.76>.
- 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. DOI: <https://doi.org/10.3389/fpsyg.2014.00673>.
- Kelly, S. D., 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(2), 313-334.
- Kendon, A. (1974). Movement coordination in social interaction: some examples described. *Nonverbal Communication*. Oxford, New York.
- Klein, W. (1986). *Second language acquisition*. Cambridge University Press. Cambridge.
- Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57(3), 396-414.
- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093-1096.
- McNeill, D. (1985). So you think gestures are nonverbal?. *Psychological Review*, 92(3), 350.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago, IL: University of Chicago Press.
- McNeill, D. (2006). Gesture and Communication. In K. Brown (Ed.), *The Encyclopedia of Language and Cognitive Processes*, 22(4), 473-500.
- Munro, M. J., Derwing, T. M., & Burgess, C. S. (2003). The detection of foreign accent in backwards speech. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 535-538). Barcelona, Spain: Universitat Autònoma de Barcelona.
- Sato, C. J. (1991). Sociolinguistic variation and language attitudes in Hawaii. *English around the world: Sociolinguistic Perspectives*, 647-663.
- Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, 8(2), 219-235.
- Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, 78(3), 705-722.

- Van Maastricht, L., Krahmer, E., & Swerts, M. (2016). Native speaker perceptions of (non-) native prominence patterns: Effects of deviance in pitch accent distributions on accentedness, comprehensibility, intelligibility, and nativeness. *Speech Communication*, 83, 21-33.
- Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57, 209-232.