BEAT GESTURES AND PROSODIC PROMINENCE: IMPACT ON LEARNING

Olga Kushch

TESI DOCTORAL UPF / 2018

DIRECTORS DE LA TESI
Dra. Pilar Prieto Vives
Dra. Carmen Pérez Vidal

DEPARTAMENT DE TRADUCCIÓ I CIÈNCIES DEL LLENGUATGE

Universitat Pompeu Fabra Barcelona

A mi familia

Verbal and nonverbal activity is a unified whole, and theory and methodology should be organized or created to treat it as such.

Kenneth L. Pike

Acknowledgements

First of all, I owe my deepest gratitude to my supervisor Pilar Prieto. I am very grateful to you, Pilar. First, for giving me the opportunity to be a member of the GrEP Research Group; it has been an amazing and very enriching experience for me to be part of GrEP, both from the academic and also the personal point of view. I have learned from you so many things related not only to academics, but also about collaborative working, and personal self-improvement. Second, I want to thank you for your patience, comprehension, and support! So many thanks for everything, Pilar.

I would like to thank my supervisor Carmen Pérez-Vidal Thanks for always providing wise insight, and for always participating with great precision when asked. I learned a lot from you, not only in the field of the research but also teching methods and personal attitude. I really feel very fortunate to have been able to count on the experience and knowledge of my two supervisors. Thanks a lot to both of you for emotional and scientific support during all these years. I really hope we will continue with further research together.

I am also very thankful to Professor Marianne Gullberg from Lund University. I had such a great opportunity to do a research stay at the beginning of my PhD in the Humanities Lab of Lund University directed by Marianne Gullberg. Professor Gullberg provided me a great amount of inspiration and a great amount of knowledge in the field of gestures that will be also reflected in the pending projects.

I am very thankful to Carmen López from the Translation and Interpretation Department of Pompeu Fabra University. My master thesis was elaborated under the supervision of Dra. López. Thank you to opening the doors of this university and giving the inspiration to continue the research!

I would like to express my special gratitude to my family. Without their support nothing would be possible. My heartfelt thanks to my parents Larisa Kushch and Ivan Kushch for having supported my decision of starting a PhD dissertation and being always next to me in spite of the geographical distance. Also, I feel endlessly thankful to my grandparents Elisaveta Suhorukova and Yurii Suhorukov who no longer belong to this world, for always believing in me and giving me the inspiration to do a research. And I am truly thankful to all other members of my family. I am very thankful to my husband Robert Abelló Guimet for being next to me during the last year of my PhD thesis. He always believed in me and encouraged me. Also, many thanks to Abelló-Guimet family for their support during the final stages of PhD writing. Thank you for looking out for me. Love you all very much!

My heartfelt thanks to my friends and colleagues that have been next to me all throughout the years of PhD writing. Eden Fairweather, Marta Garcia Balsas, Tamara Vorobyeva, Maria Jose Valero, Juliana Gómez Medina, Marta Hernandez, Alicia Castello. Thanks to all my friends. I love you. A special thanks to all members of the Prosodic Studies Group! Spetial thanks to Joan

Borràs Comes for helping with statistics. Thank you Alfonso Igualada, Santiago González Fuente, Iris Hübscher, Evi Kiagia, Ingrid Vilà Giménez, Cristina Sanchez Conde and Florence Bails, Daria Gluhareva, Irene Cañada and my other PhD colleagues at the Universitat Pompeu Fabra. Thank you very much Patrick Rohrer for proofreading my work! A special thanks to Anna Denisenko and Judith Llanes Coromina for stimuli recordings. Thanks to all of you for the great times.

Additional thanks are owed to the Department of Translation and Language Sciences at the Universitat Pompeu Fabra for their support during my pre-doctoral stage, as well as for funding my three-month research stay at the Lund University, Sweden and for the financial support that allowed me to go to many conferences. A special thanks go to Àlex Alsina, Cristina Gelpí, Toni Badia, Aurora Bel. Also many thanks to the members of the Secretary of the department, especially to Susi Bolós Ginés and Nuria Abad. This Ph.D. was funded by a FI AGAUR grant 2015-2018 awarded by Generalitat de Catalunya (2015 FI_B 00094).

Many thanks to everyone who provided feedback during the conferences and meetings. A special thanks to Maria Graziano and Sandra Debrislioska from Lund University.

Thanks to all the participants who volunteered to take part in the studies. I am also thankful to the technical support team of at the

Universitat Pompeu Fabra, who provided the technical assistance during the stimuli recording for the studies.

Many thanks to anyone else with a connection to this thesis or who feels close to this work. Thank you very much!

Abstract

Previous research has shown that gestures are beneficial for language learning. This doctoral thesis centers on the effects of beat gestures– i.e., hand and arm gestures that are typically associated with prosodically prominent positions in speech - on such processes. Little is known about how the two central properties of beat gestures, namely how they mark both information focus and rhythmic positions in speech, can be beneficial for learning either a first or a second language. The main goal of this PhD thesis is to investigate how prosodic and gestural prominence expressed together in beat gestures benefit (a) information memorization in one's native language (Study 1); (b) vocabulary learning in a second language (Study 2); and (c) pronunciation improvement in a second language (Study 3).

This dissertation includes three independent studies which all explore the above mentioned theme. The first study aims at investigating the role of prosodic prominence (pitch accents) and gesture prominence (beat gestures) on the recall of contrastive information in a piece of discourse. Results reveal that the presence of gestural prominence adds a further beneficial effect on information recall in comparison with the conditions where beat gestures are absent. The second study investigates the same phenomena and their effect this time (the stimuli contain the combination of the presence or absence of prosodic prominence in speech and the presence or absence of gestural prominence) on L2

novel words memorization. The results show that the strongest effect corresponds to target words presented with gestural prominence together with prosodic prominence. The third study aims to investigate also the same phenomenon, that is, whether the production of beat gestures on the part of the participants, as highlighters of rhythmic information, has additional beneficial effects on second language pronunciation improvement in comparison to only observing beat gestures. The results indicate that producing beat gestures leads to higher gains in accent improvement in comparison to only observing beat gestures.

Overall, the three studies show evidence that beat gestures produced naturally (that is, accompanied by prosodic prominence) favor information memorization in one´s native language (Study 1) and novel words learning in a second language (Study 2). In addition to those findings, they also reveal that producing beat gestures is shown to have a beneficial effect for pronunciation improvement (Study 3). The findings of the studies presented in this thesis support the theories of embodied cognition with new evidence that adding visuospatial information expressed by beat gestures to prosodic prominence boosts learning. These results have implications for second language instruction practices and theories of embodied cognition.

Resumen

Investigaciones anteriores han demostrado que los gestos son beneficiosos para el aprendizaje de idiomas. Esta tesis doctoral se centra en los efectos de los gestos rítmicos, es decir, los movimientos de manos y brazos que acompañan las partes prosódicamente prominentes en el habla. Hay poca investigación sobre cómo las dos propiedades centrales de los gestos rítmicos - el foco de información y las posiciones rítmicas en el habla - pueden ser beneficiosas para aprender una primera o segunda lengua. El objetivo principal de esta tesis doctoral es investigar cómo la prominencia prosódica y gestual, expresada conjuntamente en los gestos rítmicos benefician (a) la memorización de la información en una primera lengua (Estudio 1); (b) el aprendizaje de vocabulario en una segunda lengua (Estudio 2); y (c) mejora de la pronunciación una segunda lengua (Estudio 3).

Esta tesis doctoral incluye tres estudios independientes que exploran el tema mencionado. El primer estudio tiene como objetivo investigar el papel de la prominencia prosódica (acentos tonales) y la prominencia del gesto (gestos rítmicos) en la memorización de la información en un discurso. Los resultados revelan que la presencia de prominencia gestual agrega efectos beneficiosos en la memorización de información en comparación con las condiciones en las que los gestos rítmicos están ausentes. El segundo estudio investiga los efectos de los gestos rítmicos y la prominencia prosódica (a través de estímulos que contienen combinaciones de

presencia o ausencia de prominencia prosódica en el habla y presencia o ausencia de prominencia gestual) en la memorización de palabras nuevas en una segunda lengua. Los resultados muestran que el efecto más fuerte corresponde a las palabras presentadas con prominencia gestual junto con la prominencia prosódica. El tercer estudio tiene como objetivo investigar si la producción de gestos rítmicos, como marcadores de información rítmica, tiene efectos beneficiosos adicionales en la mejora de la pronunciación de un segundo idioma en comparación con solo la observación de gestos rítmicos. Los resultados indican que producir gestos rítmicos conduce a un mayor beneficio en la mejora de acento en comparación con la simple observación de gestos rítmicos.

En general, los tres estudios muestran evidencia de que los gestos rítmicos producidos de forma natural (es decir, acompañados de prominencia prosódica) favorecen la memorización de información en el idioma nativo (Estudio 1) y el aprendizaje de palabras nuevas en un segundo idioma (Estudio 2). Además, los resultados muestran que la producción de gestos rítmicos muestran un efecto beneficioso para la mejora de la pronunciación (Estudio 3). Los resultados de los estudios presentados en esta tesis apoyan las teorías de la cognición corporeizada (embodied cognition) con nuevas evidencias de que añadir información visuoespacial, expresada mediante gestos rítmicos, a la prominencia prosódica resulta beneficiosa para el aprendizaje. Los resultados de este tesis tienen implicaciones para las prácticas de enseñanza de un segundo idioma y las teorías de la cognición orientada a la acción o cognición corporeizada .

Resum

Investigacions anteriors han demostrat que els gestos són beneficiosos per a l'aprenentatge de llengües. Aquesta tesi doctoral se centra en els efectes dels gestos rítmics, és a dir, els moviments de mans i braços que acompanyen les parts prosòdicament prominents de la parla. Hi ha poca investigació sobre com les dues propietats centrals dels gestos rítmics - el focus informatiu y les posicions rítmiques de la parla - poden ser beneficioses per aprendre un primer o segon idioma. L'objectiu principal d'aquesta tesi doctoral és investigar com la prominència prosòdica i gestual, expressada conjuntament en els gestos rítmics, beneficien (a) la memorització de la informació en una primera llengua(Estudi 1); (b) l'aprenentatge del vocabulari en una segona llengua (Estudi 2); i (c) millora de la pronunciació una segona llengua (Estudi 3).

Aquesta tesi doctoral inclou tres estudis independents que exploren el tema esmentat. El primer estudi té com a objectiu investigar el paper de la prominència prosòdica (accents tonals) i la prominència del gest (gestos rítmics) en la memorització de la informació en un discurs. Els resultats revelen que la presència de prominència gestual afegeix efectes beneficiosos en la memorització d'informació, en comparació amb les condicions en què els gestos rítmics estan absents. El segon estudi investiga els efectes dels gestos rítmics i la prominència prosòdica (a través d' estímuls que contenen combinacions de presència o absència de prominència prosòdica en la parla i presència o absència de prominència gestual) en la memorització de paraules noves en una segona llengua. Els

resultats mostren que l'efecte més fort correspon a les paraules presentades amb prominència gestual juntament amb la prominència prosòdica. El tercer estudi té com a objectiu investigar si la producció de gestos rítmics, com a marcadors d'informació rítmica, té efectes beneficiosos en la millora de la pronunciació d'una segona lengua,. Els resultats indiquen que produir gestos rítmics condueix a uns majors guanys en la millora de l'accent en comparació amb la simple observació de gestos rítmics.

En general, els tres estudis mostren evidència que els gestos rítmics produïts de forma natural (és a dir, acompanyats de prominència prosòdica) afavoreixen la memorització d'informació en una primera llengua (Estudi 1) i l'aprenentatge de noves paraules en un segon idioma (Estudi 2). A més, els resultats mostren que la producció de gestos rítmics mostren un efecte beneficiós per a la millora de la pronunciació (Estudi 3). Els resultats dels estudis presentats en aquesta tesi recolzen les teories de la cognició corporeïtzada (embodied cognition) amb noves evidències que afegir informació visuoespacial, expressada mitjançant els gestos rítmics, a la prominència prosòdica resulta beneficiós per a l'aprenentatge. Els resultats d'aquesta tesi tenen implicacions per a les pràctiques en l'ensenyament d’una segona llengua i les teories de la cognició orientada a l'acció o cognició corporeïtzada.

List of original publications

CHAPTER 2

Kushch, O., & Prieto, P. (submitted). Beat gestures increase the effects of prosodic prominence on information memorization in a first language. *Journal of Phonetics*.

CHAPTER 3

Kushch, O., Igualada, A., & Prieto, P. (under revision) Prominence in speech and gesture favor second Language novel word learning. *Journal of Language, Cognition and Neuroscience.*

CHAPTER 4

Kushch, O., Gluhareva, D. Borràs-Comes, J, Pérez, P.,& Prieto, P. (submitted). The role of beat gesture production in L2 pronunciation training. *Bilingualism: Language and Cognition.*

Table of contents

1. INTRODUCTION

This thesis seeks to investigate the role of a specific type of gesture aligned with prosodic prominence of speech, e.g. the so-called beat gestures, on word memorization in one's native language and on word learning and pronunciation improvement in a second language. In this Introduction we provide an overview of research on gestures and their role in language learning. We first discuss gestures as an integrated system (1.1.). We then provide an overview of gesture typology, paying specific attention to beat gestures and their properties, as they are the main topic in our research (1.2.). Then, we go on to discuss the beneficial effects of gestures on learning, and, more specifically we offer an overview of the studies which show that gestures aid first and second language acquisition (1.3.). In order to narrow down the scope of our review, next we proceed to discuss the previous literature focusing on the issue of information memorization in one's native language (1.4.); word learning in one's second language (1.5.); and, finally, on second language pronunciation improvement (1.6.). We conclude the section by presenting the general objectives of the thesis, the research questions and the hypotheses (1.7).

1.1. Gesture and speech as an integrated system

In the same manner as we express our thoughts through speech, we express our thoughts through gestures. Recently, researchers have claimed that gestures may reflect the cognitive processes of people talking as well as ideas that are not directly reproduced in speech. Kendon (1980) was the first to state that gestures cannot be separated from speech, but rather form one integrated system. Several studies support the view that speech and gesture form a single system (Goldin-Meadow, 2003). For example, when we talk on the phone and are perfectly aware that no one can see us, we keep gesturing, sometimes without even being conscious of it (Goldin-Meadow, 2003). Iverson and Goldin-Meadow (1998) analyzed the spontaneous conversations of 12 blind children and adolescents aged 9;1 to 18;10 and those of 12 subjects with normal vision, all of them of approximately the same age. Participants were asked to take part in a reasoning task designed to elicit gesturing in sighted children. The results showed that blind children did not show significant differences in gesturing in comparison to the sighted group.

Another argument to support the gesture-speech integrated view was put forth by McNeill (1992), who detected that 90% of communicative gestures were produced during speech. According to McNeill (1992), gesture and speech are synchronous at three levels: semantic (co-occurring gesture and speech express the same idea), pragmatic (in both modalities speech and gesture serve for the same intentional function), and phonological (prominences of gesture and speech are temporally aligned). There is clear evidence of strict synchronization between speech and gesture. The peak of gestural movements, the so-called strokes or apexes, tend to temporally co-occur with prominence in speech (e.g., Goldin-Meadow, 2003). For example, in order to investigate the anchoring regions in speech that align with pointing gestures, Esteve-Gibert and Prieto (2013) asked fifteen Catalan speakers to participate in a pointing-naming task. Participants were asked to point at a screen while pronouncing a variety of target words with different metrical patterns. 720 instances of pointing- speech combinations were obtained. The results of the study showed that (a) intonation peaks, gesture strokes and apexes varied depending on the distance of the accented syllable to the upcoming phrase boundary; and (b) intonation peaks and gesture apexes were synchronized and followed the same timing patterns.

Kelly et al.'s (2010) study proposed the integrated system hypothesis. The study investigated the nature of the interaction between gesture and the speech systems. They carried out two priming experiments. In the first experiment, they showed videos of performed actions (e.g. “cut”) in 3 different gesture-speech conditions: congruent, weakly incongruent (e.g. gesturing cut while saying “chop”), and strongly incongruent (e.g., gesturing twist while saying “chop”). The results of Experiment 1 showed that when gesture and speech carried the same information these combinations were fast and easy to process and induced fewer errors than when they carried incongruent information. Also, combinations involving strong incongruities produced more errors than weaker ones. In their second experiment, the authors tested the same materials using the same procedure with the only difference being the instructions. Participants were instructed to watch action primes that were followed by information conveyed in speech and gesture, but the task of the participants consisted in saying whether speech content was the same or different from the primes. Their prediction was that if speech and gesture were obligatorily integrated, participants would be slower or less accurate to relate speech targets with action primes, as gesture and speech become increasingly incongruent. A novel result that they encountered was that this interaction was obligatory, in the sense that participants cannot ignore one modality (gesture) while processing the other (speech) (Experiment 2). Therefore, both experiments provided clear evidence for the integrated systems between speech and gesture.

All in all, these findings underscore that gesture and speech form a unique system. In general, evidence shows that to ignore gesture is to ignore an important part of the communicative system (Goldin-Meadow, 2003). In this thesis, we investigate the effects of using one type of hand gesture that typically associates with prosodic prominence in speech (the so-called beat gestures, see section 1.2 below) on L1 word memorization (Study 1) and L2 word learning processes (Study 2), as well as L2 pronunciation learning (Study 3). In order to do that, we now turn to a presentation of the existing views on the categorization of gestures.

1.2. Gesture Categories

1.2.1. Gesture typology

Not all bodily movements can be considered gestures that are part of the human communication system. In an overview of gesture typology, Gullberg (1998) states that beginning with Wundt (1921) there have been different gesture classifications. For example, Ekman and Friesen (1969) classified nonverbal behavior into five categories: affect displays, regulators, adaptors emblems and illustrators. The first two types in this classification do not necessarily involve hand movements.

- Affect displays are primarily centered in an individual's facial expression. They are defined as the mimic muscle movements associated with primary emotions as disgust, anger, sadness, etc.
- Regulators are hand movements or slight changes in the body mage to facilitate a turn taking. They aim to maintain the back and forth nature of speaking and listening between interlocutors. These types of gestures show the speaker to continue, repeat, elaborate, hurry up, talk, etc. A typical regulator is a head nod, which is equivalent to the verbal "mm-hmm."
- Adaptors are hand movements that are maintained by habits, for example, smoothing hair, pushing the glasses up on the nose, holding one's chin. Normally, they are

produced with no intent of communication and a speaker has limited awareness of gesturing.

• Contrary to adaptors, people are always aware of producing emblems. Emblems, however, have no tight alignment to speech, they can be easily produced without speech, as they are widely produced at places when the verbal communication is prevented, for example, by loud music, distance or agreement between speakers. They are nonverbal acts which have a direct verbal translation or dictionary definition. It can be a word or two or even a phrase. The definition of an emblem is known by a group, class or culture, because emblems are established forms that transmit their meaning even produced without any utterance. As emblems we can consider 'thumbs up' or 'OK' sign.

• Finally, illustrators are defined as movements, directly tied to speech, that serve to underscore what is communicated verbally. Illustrators are intimately interrelated with the concomitant verbal behavior on a moment-to-moment basis and they are directly tied to content, inflection and loudness. This type of gesture repeats, substitutes, contradicts or argues for the information that has been provided verbally. Illustrators are informative like emblems, but the person who produces illustrators is less aware of them in comparison to producing emblems.

Goldin-Meadow (2003) presents a more restrictive binary criteria for identifying gestures: (1) gestures are produced as part of an intentional communicative act, in contrast to adaptors; (2) gestures are constructed at a moment of speech, in contrast to emblems. The author states that we cannot consider twisting off the lid of the jelly jar a gesture, even while asking to pass on the peanut butter. The jar-twisting action cannot be considered a gesture, despite the fact that this gesture accompanies speech. The jar twisting is a functional act on an object that should be distinguished as a non-gesture.

In this thesis we will follow McNeill's (1992) basic typology. McNeill (1992) divides all visible movements into gestures and non-gestures. Within non-gestures he includes self-touching (for example stroking the hair) and object manipulation.

McNeill describes four types (dimensions) of gestures: representational, deictic, conventional and beat gestures, the central type for the current research. However, it is important to notice that the gesture categories described in McNeill (1992) are not mutually exclusive and that, in fact, any gesture may involve more than one category and should be considered as a continuum.

1. Representational gestures, which include iconic and metaphoric gestures represent objects and actions in space. Here the form of gesture bears a close relation to the semantic content of speech. Iconic gestures represent tangible information, like for example the shape of an object (e.g., a circle shaped with hands may represent a ball). Metaphoric gestures represent an abstract idea rather than a

concrete object, such as knowledge, language, and genre of narration. These gestures substitute the image of something invisible - an image of the abstraction. For example, touching the head with two fingers may represent the verb “to think”.

2. Deictic gestures are pointing movements, usually performed with the index finger. These gestures are used to indicate objects, people and locations in present physical environment or in abstract space. Most pointing gestures during narration or conversation are of an abstract kind. For example, in a conversation between two previously unacquainted students, one of them asks the other, “Where are you coming from?” and points to space between self and interlocutor. The space, at which this student points, is not a place whereas speaker and hearer currently find themselves, but an abstract space (McNeill, 1992: 18)

3. Conventional gestures include culturally shared symbols, with an arbitrary form and meaning within a given community. A common example would be a “hi” hand gesture.

4. Beat gestures are rhythmic hand and arm movements that serve as visual highlighters of information, and “mark the word or phrase they accompany as being significant (…) for its discourse pragmatic content” (McNeill, 1992:15). Beats, together with pitch accentuation, have been associated with focus marking and discourse structure marking functions in speech (e.g., Loehr, 2012; Shattuck-Hufnagel, Ren, Mathew, Yen & Demuth, 2016). Beat

gestures are a type of rhythmic hand and arm movement that are typically associated with prominent prosodic positions in speech; their function is non-referential and they are generally used in language to signal informational focus (e.g., McNeill, 1992; Shattuck-Hufnagel et al. 2016).

In the following section we describe the properties of beat gestures that are crucial for our research.

1.2.2. Beat gestures as highlighters of prosodic prominence and informational focus

As has been discussed in the previous section, speakers integrate their gestures and speech sounds at a temporal level and this temporal synchrony between speech and gesture provides evidence of an integrated spoken language and gesture communication system (e.g., see Wagner et al. 2014 for a review; Rusiewicz, Shaiman, Iverson, & Szuminsky, 2013; Iverson & Thelen, 1999; McNeill, 1992; Rusiewicz & Esteve-Gibert, in press). In this PhD thesis we focus our attention on beat gestures, which have received relatively little attention in comparison to other gesture types (Kelly, Manning, & Rodak, 2008).

According to its traditional definition, a typical beat is considered to be a short and quick flick of the hand, either up and down, or back and forth (McNeill, 1992). Beat gestures were initially called 'batons' (Efron, 1941; Ekman & Friesen, 1969), because of the similarity to the gestures of an orchestra conductor when beating music time. According to Leonard and Cummins (2011), a typical beat gesture consists of two movement phases: an extension phase and a retraction phase, and they can be thought of as a "rhythmical pulse". Beats place emphasis on the particular word or phrase they accompany with a baton movement, which is often metronomic (e.g., up/down, left/right) (Leonard and Cummins, 2011: 4).

In contrast to other gesture subtypes, beat gestures have been argued to lack abstract semantic content (e.g., Andric and Small, 2012; Alibali et al., 2001; Krauss et al., 1996; McNeill, 1992; Leonard & Cummins, 2011). That does not mean that beat gestures lack communicative value, even when compared to representational gestures. Beats display less meaningful content, it is also clear that they have a pragmatic function in speech as highlighters of information structure. According to McNeill (1992:15), ''the semiotic value of a beat lies in the fact that it indexes the word or phrase it accompanies as being significant (...) for its discourse pragmatic content.'' A beat thus provides extra prominence for a word, for instance, because it conveys new information (McNeill 1992 : 169–170). As Alibali et al. (2001: 84) pointed out, there is a "need for further study of beat gestures and their role in speech production and communication.'

In speech, the emphasis is signaled by prominent prosody, i.e., via pitch accents (e.g., Cruttenden, 1997; Ladd, 1996; Swerts, Krahmer, & Avesani, 2002 among many others). There is clear evidence that beat gestures are temporally synchronized with prosodic markers of prominence (i.e., pitch accents) (McNeill, 1992; Yasinnik, Renwick, & Shattuck- Hufnagel, 2004; Jannedy, & Mendoza-Denton, 2005; Loehr, 2012; Shattuck-Hufnagel, et al., 2016). Prominent parts of gestures and speech occur in tight synchrony. (e.g., De Ruiter, 2000; Esteve-Gibert & Prieto, 2013; Loehr, 2012; Rochet-Capellan, Laboissière, Galván, & Schwartz, 2008; Rusiewicz, 2010; Yasinnik et al., 2004; see Wagner et al., 2014 for a review). Loehr's (2012) analysis of adult narrations showed that prominent accentuations at the intonation phrase level (i.e. pitch accents) were systematically coordinated at temporal level with the stroke of gestures. Yasinnik et al. (2004) showed that during a narration more than 90% of instances of the gesture apexes in English occurred together with a pitch-accented syllable (see also Jannedy & Mendoza-Denton, 2005 for a review). Figure 1.1 illustrates the temporal alignment between the stroke of a beat gesture and the pitch accent in speech.

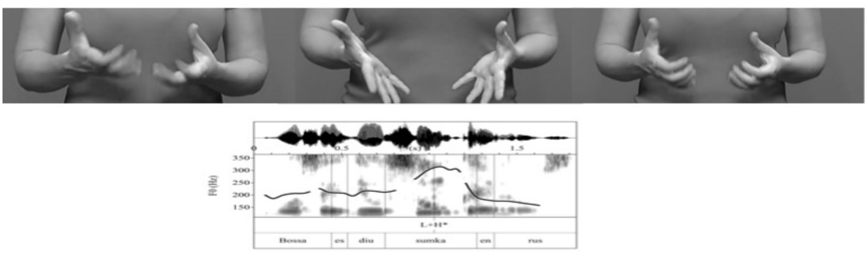


Figure 1.1. Example of temporal alignment between a pitch accent L+H*and the stroke of a beat gesture.

Leonard and Cummins' (2011) study tested whether the timing of beats can influence their identification and interpretation. In their perception study, adult participants were exposed to temporally desynchronized combinations of gestures and speech. Adults detected 200 ms desynchronized late gestures but had problems detecting early desynchronized productions. Their results show that listeners are sensitive to the asymmetries in the relative timing between beats gestures and speech

Recent research has also shown that prosodic prominence is typically correlated with more prominent gestural and articulatory features (e.g., beat gestures, head nods, eyebrow movements, exaggerated articulation, etc.; see, for example, Swerts & Krahmer, 2008; Prieto et al., 2015, see Esteve-Gibert & Prieto, 2013 for a review). Krahmer and Swerts (2007) found that prosodic prominence and visual information work in a complementary fashion and that adding visual beats to prosodic prominence lead to stronger production and perception of prosodic prominence. In this experiment, participants were exposed to the sentence "Amanda goes to Malta" in which the two target words were associated with prosodic and/or gestural prominence. Beat gestures presented together with focused words increased the perceived prominence of these words and decreased the perceived prominence of other target words.

Several studies assessing neurological activations during observations of beat gestures support the idea that beat gestures increase activations of language-related brain areas (e.g., Biau & Soto-Faraco, 2013; Biau, Fromony & Soto-Faraco, 2017; Holle et al., 2012). A functional neuroimaging study by Biau and Soto-

Faraco (2015) showed that beat gestures activated different brain areas in comparison to other non-related movements. Depending on whether speech was synchronized with beat gestures or with other non-gestural stimuli (discs/dots moving on a screen) different brain areas were activated. Beat gestures activated language-related areas of the brain, while non-gesture stimuli activated visual perception areas.

A neurophysiological study conducted by Hubbard et al. (2009) found that adding gestural prominence in the form of beat gestures to prosodic prominence causes greater activity in bilateral nonprimary auditory cortex, suggesting a common neural substrate for processing speech and gesture. This study investigated whether the presence of beat gestures impacted speech perception at the neural level, controlling for the presence of prosodic prominence. Thirteen adult subjects underwent Functional Magnetic resonance imaging (fMRI) while being exposed to videos with spontaneously-produced speech accompanied by beat gestures, nonsense hand movements, or no movement. The bilateral non-primary auditory cortex showed greater activity when speech was accompanied by beat gestures than when speech was presented alone. Additionally, the left superior temporal gyrus/sulcus revealed stronger activity when speech was presented together with beat gestures in comparison to the speech plus nonsense hand movement stimuli. The right planum temporale was identified as a multisensory integration site for beat gesture and speech. Hubbard et al. (2009) found that beat gestures, and not nonsense movements or still images, enhanced auditory processing of speech. These studies support the idea that beat gestures can be distinguished from other potential visual highlighters because of their direct integration in the language system.

All in all, little is known about how two main properties of beat gestures (e.g., marking of information focus and marking of prominent prosodic positions in speech) can be beneficial for learning a first and second language. The main goal of this PhD thesis is to investigate how prosodic and gestural prominence expressed together in beat gestures benefit (a) information memorization in one's native language (Study 1); (b) vocabulary learning in a second language (Study 2); and (c) accent improvement in a second language (Study 3). We believe that two main properties of beat gestures (e.g., marking of prominent prosodic positions in speech and marking of information focus) should be beneficial for learning a first and second language.

1.3. Gestures and learning

In this section we summarize previous research dealing with how gestures impact general learning in the case of children (subsection 1.3.1), first language development (subsection 1.3.2), and second/foreign language development respectively (subsection 1.3.3).

1.3.1. Gestures and learning processes

This section presents an overview of key studies investigating the impact of gesture on learners' cognitive processes while speaking in their native language. Over the last decades, several studies have shown the positive effects of gesturing on learning processes (e.g., Alibali & DiRusso, 1999; Broaders, Cook, Mitchell & Goldin-Meadow, 2008). The study by Alibali and DiRusso (1999) investigated the effect of learner gestures on the process of learning to solve a math problem (namely counting a set of objects). Twenty preschoolers were asked to count chips under the following three conditions while children were counting aloud: (a) gesture prohibited, (b) active gesture, (c) puppets gesturing. It turned out that children were better at counting under the last two conditions than under the non-gesturing condition. However, the number of errors differed when children counted themselves compared with when they were observing the puppet. Active gestures helped children to keep track and coordinate tagging the items and saying the number aloud. Thus, from these findings it is possible to conclude that gesture-promotion may positively affect language learning.

A decade later, the study by Broaders et al. (2007) aimed at determining whether forcing speakers to gesture encouraged them to learn. 106 children were asked to solve six mathematical problems designed in the following way: 6+3+7=__+7. Children were divided into gesture, non-gesture, and control groups. The results revealed that asking children to gesture encouraged them to convey previously unexpected but correct information, proving a positive effect of gestures on the learning process.

Cook, Mitchell and Goldin-Meadow (2008) conducted a math learning task, with 84 third- and fourth- grade children, consisting in solving equations presented in the following manner: 4 + 9 + 3 = 4 + __ . First, there was an instruction phase, where children were exposed to three conditions: (a) Speech condition, which consisted in the instruction "I want to make one side equal to the other side"; (b) Gesture condition, in which the instructor moved her hand under the left side then under the right side and did not speak; and (c) Gesture + Speech condition, in which the instructor said "I want to make one side equal to the other" and accompanied it with gestures. In the second phase, the instructor taught the child how to use the equalizer strategy to solve six more mathematical problems of the following type 4+9+3=4+__. The instructor said "I want to make one side (in this moment the instructor was sweeping the left hand under the left side of the equation) equal to the other side (gesture sweeping the right hand under the right side); so, four plus nine plus three equals sixteen, and four plus twelve equals sixteen; one side (gestures under the left side) is equal to the other side (gestures under the right side)". After the instruction phase, children were given a problem of the same type to solve and were asked to produce one of the previously known strategies (Speech, Gesture, Gesture+Speech). The results showed that asking children to gesture while learning the new concept helped them to retain the knowledge that they had received from the instructor. Children who were taught to gesture while solving the math problem presented better results in the post- test.

Consequently, current research findings seem to point in the direction that starting from an early age, our cognitive processes are interconnected with gesturing, and by promoting the use of gestures we can promote general cognitive processes. In the following subsections we will discuss more studies specifically centered on first language and second language learning.

1.3.2. Gestures and first language learning

This section reviews the relevant literature that deals with how gesturing promotes L1 linguistic development; how and at what stages of development do infants start producing a gesture-speech integrated system; and how such stages seem to serve as a forerunner of incoming changes in speech, at different ages, and with both representational gestures and beat gestures.

We come across gestures from the very beginning of our lives. Gesturing is a prime way of communication between an infant and his/her caregiver. As gesturing occurs naturally when we speak, mothers also gesture. Mothers align their speech and action while communicating with their infants (Meyer et al., 2011). So how does a child benefit from gesturing while acquiring his/her mother tongue? As Goodwyn et al. (2000) state, gestures are a critical component of caregiver's communication and play a crucial role in early language development. Various studies suggest that adult gestures produced in accordance with an object they are naming help children to perceive the connection between words and their lexical referent (e.g., Yoshida & Smith, 2007; Yu, Ballard & Aslin, 2005). Matatyaho-Bullaro et al. (2014) claimed that shaking and looming motions help increase learning of word-object relations. The authors conducted a study with 60, 8-month old infants, exposing them to two words presented by a toy under different motion conditions – shaking, looming, upward and sideway. The result of the experiment showed that the type of gesture used by the adult affected the word learning process of the preverbal infants tested. The study concluded that they learned better with shaking and looming motions produced by their caregivers.

In order to investigate the relationship between maternal verbal and gestural labeling which accompanies a set of common nouns, Zammit and Schafer (2010) conducted a study in which they observed ten mothers interacting with their 10-month old infants. The authors categorized verbal labeling events produced by mothers into three categories: label only, label plus deictic/pointing gesture, label plus iconic gesture. The results of the study showed that maternal use of iconic gestures predicted the timing of acquisition of nouns in comprehension. Zammit and Schafer (2010) detected a correlation between the number of maternal iconic gestures and the number of words that children produce. Particularly, children whose mothers produced iconic gestures while labeling comprehended and thus acquired the vocabulary used in the conversation faster than children whose mothers produced uncoordinated gestures. The results of the study are not surprising, as previous studies had shown that the use of iconic gestures is crucial for comprehension. The study by Namy et al. (2000) involving eighty 15-month aged infants and their parents demonstrated that parents regularly produce gestural labels in their input to infants and that their iconic gestures play an important role in comprehension, as they support the child's recall of object labels.

The next point of interest in research in L1 acquisition has been how and when infants start producing a gesture-speech integrated system. As pointed out above, several authors have argued that gesture and speech form part of an integrated system (e.g., Bernardis & Gentilucci, 2006; Esteve-Gibert & Prieto 2014; Goldin- Meadow, 2003; Kendon, 2004; McNeill, 1992, 2005). The question that arises is when does this integration start to occur: Is there a moment in L1 linguistic development when gesture is used without speech? When do children start producing gestures integrated with speech? Several studies have investigated the synchronization between speech and gesture at early L1 production stages (Butcher & Goldin-Meadow, 2000; Goldin-Meadow & Butcher, 2003; Esteve-Gibert & Prieto 2014). Butcher and Goldin-Meadow (2000) carried out a longitudinal study with six children between the ages of 12-27 months. They found that gesturing (e.g., pointing gestures) occurs in isolation at the beginning of language production and becomes integrated with speech before children begin producing words in combination with other words, i.e. between 14-23 months. Therefore, only when reaching two years of age, does gesture and speech become unified into a single system, which is characterized by semantic and temporal coherence (Goldin- Meadow, 2003). Yet in another study, Esteve-Gibert and Prieto (2013) detected that 11-month old infants already produce synchronous gesture-speech combinations, while isolated pointing is still more frequent at this age. In the longitudinal sample of the study it was possible to detect a significant increase of gesture-speech synchronous productions by 15 months of age.

In general, the period in which a child is able to combine words with gestures can be considered as a transition stage toward the production of two-word combinations (Capirci, Iverson, Pizzuto & Volterra, 1996; Butcher & Goldin-Meadow, 2000). As Goldin-Meadow and Butcher (2003) state, the occurrence of gesture-speech combinations, with gestures conveying supplementary meanings, predicts the later occurrence of two-word utterances. Thus, when a child says "mommy" and points at a cup, he/she shows his/her ability to convey sentence-like meaning across gesture and speech. It predicts that he or she will be able to convey the same meaning entirely within speech and with proto-syntax (Özçaliskan & Goldin-Meadow, 2005). The study by Igualada, Bosch and Prieto (2016) investigated the influence of infants' use of synchronous gesture-speech combinations on later language development. The study was conducted with 19 12-month old infants who were exposed to a declarative pointing task under three different conditions: Available condition, when the adult was visually attending the infant but not the object of reference the adult was pointing to; unavailable condition, when the adult was not visually attending neither the infant nor the object; and a baseline condition, when the adult was engaged with the infant's object of reference. Results revealed a significant interaction between social condition and communicative productions. This combination of synchronous gesture – speech pointing positively correlated with vocabulary and grammatical development at the age of 18 month. The authors conclude that the ability of an infant to use a multimodal communicative strategy at 12 months reveals itself as an effective way to communicate with the adult, and this is related to later vocabulary outcomes on the part of the child.

In their study, Özçaliskan and Goldin-Meadow (2005) tried to find an answer to whether the production of supplementary gesture-speech combinations by children of 14, 18 and 22 months presaged oncoming changes in their speech and served as a forerunner of linguistic advances. Forty children were videotaped while they were interacting with their caregivers; all the meaningful sounds and communicative gestures were transcribed. The results of the study showed, not surprisingly, that speech and gesture change with age. What is, however, more important is that it also showed that the types of gesture-speech combinations also change over time and indicate changes in speech. Children who produced one word at a time used gesture to supplement their speech, thus turning a word into an utterance with gestures and presenting a sentence-like meaning (eat + point at cookie). Children did not produce utterances containing verbs with two arguments, as for example “mommy the cookie” or with an argument and predicate “me touch” until they were 22 months old, but many children revealed their readiness to compose this construction by gesture-speech combinations: indeed, children produced such constructions as gesture-speech combinations, as for example “mommy” + point at couch or “you”+ hit gesture. During the study it also turned out that few children produced utterances with two predicates in speech, as, for example, “help me find”, even at 22 months, and they produced them in gesture- speech combinations “I like it” + eat gesture. Importantly, it was only after a short period that children began producing these combinations in speech. Thus, the study by Özçaliskan and Goldin-Meadow (2005) confirms the fact that gesture can precede and signal changes in speech.

In the field of first language acquisition, most of the existing studies have centered on the role of representational gestures – in McNeill’s terms – gestures that represent the semantic context of speech, in L1 learning (e.g., Cook, Mitchell, & Goldin-Meadow, 2008; Goldin-Meadow, Kim, & Singer, 1999; Tellier, 2005). However, there is also research that investigates the role of beat gestures for discourse comprehension by children, unfortunately with contradictory results thus far. For example, Macoun and Sweller's (2016) study assessed the effects of beat gestures on narrative comprehension by children, and obtained negative results. Macoun and Sweller (2016) investigated the effects of four gesture conditions (iconic gestures, deictic gestures, beat gestures, and no gestures) on preschoolers' (age range 3.25–5.58) narrative comprehension and recall of information with a between-subjects experimental design. The results showed that whereas iconic and deictic gestures provided benefits for comprehending and recalling information in narratives, beat gestures yielded no beneficial effects in comparison to the condition without gestures. However, recent experiments conducted by Llanes et al, (under revision) and Vilà-Giménez, Igualada, and Prieto (under revision) showing that beat gestures, which act as markers of both focus and discourse structure information, have positive effects on both the recall and comprehension of information in narratives (Llanes et al under revision) and on narrative performance skills (Vilà-Giménez et al. under revision).

Finally, there is evidence on the positive role of beat gestures on information recall by children. For example, Austin and Sweller's (2014) study tested whether beat gestures would help 3- to 4-year-old children and adults recall spatial directions. In a between-participants design, children and adults were given verbal route directions to follow a path through a small-scale spatial array, with the speaker using no gestures, beat gestures, or representational gestures. While no significant results were found for adults, children recalled the directions better when they were accompanied by either a beat or a representational gesture than when there was no gesture. Also, Igualada et al. (2017) confirmed that preschool children remember more words when they were accompanied by beat gestures in pragmatically relevant contexts (for more information, see Section 1.5. below)

In sum, on the basis of the abovementioned evidence it seems clear that representational gestures favor learning in a first language, less is known about the effects of beat gestures. Further research is needed to provide more evidence on the beneficial effects of beat gestures on first, and for that matter, second language learning; a topic to which we now turn.

1.3.3. Gesture and second language learning

This section deals with the effect of gesturing in L2 acquisition. Gullberg (2010) stresses the crucial role of gestures in second and bilingual language acquisition, as they provide evidence on such issues such as the nature of representations and knowledge at a given moment in time (see also Gullberg, 2006a; Nicoladis, 2007; Gullberg, de Bot and Volterra, 2008). More specifically, studies have tested the use of gestures with different levels of proficiency in the target language and different degrees of fluency, at different ages, with different types of content to be retold (i.e. narratives, speech related to space). Often lexical access has been the focus of these studies. The typology of gesture we apply for this review is McNeill's which identifies representational, deictic, conventional and beat gestures (see 1.2.1.). From a psycholinguistic perspective, gestures can be seen as an integral part of the input to which learners are exposed and need to process, when learning a target language (Pérez-Vidal, 2014).

As we have discussed already, people convey information through speech and also through gesture. We talk and we gesture to provide information that is similar or supplementary to that of speech. Gesturing helps us to acquire our mother tongue, but does it help to acquire a second language?

With respect to second language acquisition, Quinn-Allen (1995) conducted a study assessing the impact of gestures on memory for verbal information in a foreign language. The authors hypothesized that students would learn better if expressions were presented simultaneously with gestures, resulting in better retention and more persistent recall over time. The experimenter presented 10 French expressions accompanied with gestures to 112 English-speaking first-semester students learning French. An example of the sentence and accompanying gesture "C'est un type qui sait louvoyer" ("That's a

sly one") - move hand in a snake-like fashion in front of you. The participants were divided into three groups. The first group was to learn the expressions simultaneously with gestures and to use them to recall the learned expressions; the second group never saw gestures when learning new expressions; the third group did not have to learn gestures but saw them while they had to remember the words. The results showed a significant positive effect of learning expressions simultaneously with gestures on short- term and long-term recall. That is, the students who saw gestures recalled more expressions that the group that did not see the gestures.

The idea that language proficiency influences gesture production was investigated in the late seventies by Marcos (1979). He asked sixteen Spanish – English bilinguals to speak about love or friendship in their non-dominant languages. Some of the bilinguals were Spanish-dominant, while others were English-dominant. The results showed that the less proficient a speaker was in the L2, the more gestures occurred in his or her speech. Gullberg (1998) conducted a production study that examined the strategic use of gestures by native speakers and language learners. The subjects were five Swedish students with an intermediate level of proficiency in French and six French students with the same level of proficiency in Swedish. Participants saw and memorized a printed cartoon that contained no text or words, and had to retell it both in their native and in their respective second languages to a native speaker. Native speakers evaluated the recordings. The results showed that strategic gestures do not replace, but complement speech and can be used to solve lexical problems. Also, the influence of gestures on proficiency assessment was observed: the subjects produced more gestures in their non-dominant second language. The study by Gullberg (1998) is in accordance with Marcos' (1979) results. It can therefore be considered as a confirmation that there is a positive correlation between proficiency and gesturing in that gesturing increases alongside with linguistic difficulties.

Graziano and Gullberg (2013) studied the compensatory role of gestures in speech production by fluent and non-fluent speakers. They analyzed three multimodal sets of narrative discourse. They tested three categories of subjects: The first category contained 33 Italian children divided in 3 equal groups (ages 4-5, 6-7, 8- 10). Children were asked to retell a cartoon in their L1 (Italian) to an adult they knew before. The adult, in their turn, was not to interrupt. The second group of participants was comprised of 11 Italian adults who were also asked to retell the cartoon to his or her interlocutor, also in their native language. The third group consisted of 11 Dutch adult learners of French. This group studied French for a minimum of 4 years and had different levels of proficiency. None of the participants has ever lived in a French- speaking country. Participants were asked to narrate a cartoon to a native speaker. It turned out that L2 learners were more likely to gesture than the other two groups. Results showed an effect of proficiency of language in the sense that less proficient learners produced a greater amount of gesturing during spontaneous speech production. On the other hand, the study also showed that when speech stops, gesture also stops- something that once again confirms the integration of speech and gesture into one system.

Nicoladis, Pika, Yin and Marentette's (2007) study investigated whether the relation between gesture use and language proficiency is mediated by task complexity. The authors tested the idea that task complexity would lead toward the use of more gestures in the less dominant language. The experiment was conducted with 16 adults having Chinese as their first language and English as their second language. Subjects were asked to watch 6 min of a Pink Panther cartoon that consisted of two different stories. Subsequently, they had to retell the story to an interlocutor who had never seen the cartoon before, under two conditions: (a) to retell the story in English to a native speaker of English, and (b) to narrate the story in Chinese to a native speaker of Mandarin Chinese. Participants produced longer stories in their native language and used more iconic and non-iconic gestures while retelling the story in their L2, indicating that gesture production depends on task complexity. Crucially, the authors assumed that retelling the story in an acquired second language is a more difficult task than telling a story in a native and/or dominant language. To conclude, the more difficult the task is for the speaker, the more gestures will be produced.

There is an ongoing discussion as to why gesturing appears so often in L2 speech production. One of the hypotheses put forward by Rauscher, Krauss and Chen (1996) was that representational gestures play a role in lexical access. To test this hypothesis, Rauscher, Krauss and Chen (1996) conducted an experiment with 41 undergraduates who were fluent and native speakers of English (only one was not native). Participants were asked to watch six videotaped excerpts from a Warner Brothers' cartoon, averaging 2 min 45 sec in length. Subjects were videotaped while describing the cartoon in the following experimental conditions: no-gesture condition (they were instructed to keep their hands unmoved);

gesture condition, crossed with three speech-conditions: normal-speech condition, obscure speech condition (subjects were to use as many obscure words as possible), constrained-speech condition (subjects were to avoid using words that contained a specific letter). The results revealed that the speakers who were asked not to gesture during their speech produced the same number of speech errors as the speakers who had an obstacle (i.e., obscure words or constrained words) to their speech. Thus, preventing speakers from gesturing adds an extra difficulty to their verbal expression. Rauscher et al. (1996) predicted that gesturing would become more frequent when the conceptual content of the speech is spatial. It turned out that when the speech contained spatial content, the participants spoke more slowly when they could not gesture, but when the content was non-spatial, they accelerated the speech without gesturing.

Beattie and Caughlan (1999) investigated the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. In their experiment, Beattie and Caughlan (1999) presented 25 definitions of rare words to 60 adult participants who were asked to find the word corresponding to the definition. The speakers were divided into two groups, namely those who were allowed to gesture and those who were not. The hypothesis was that iconic gestures have a functional role in lexical access, in other words, those participants who were free to gesture were expected to have less trouble in accessing the words. The results of the study indicated that gestures were associated with lexical search, however not only iconic gestures were detected. Additionally, there was no evidence that the presence of iconic gestures helped the speaker to resolve the tip-of-the-tongue states.

Ravizza (2003) also examined whether the presence of non-iconic gestures (such as beat gestures) influences word retrieval. The author conducted an experiment with 20 undergraduates at the University of California. The participants were supposed to find the word corresponding to each respective item in a set of a hundred definitions. In the first experiment, 20 students were divided in the two groups according to the motionless and tapping condition. Motionless conditions referred to the impediment to gesture; during the tapping condition participants were asked to produce up/down movements with their left hands. The results of the study confirmed the hypothesis that simple movements without semantic content (i.e., beat gestures) can have a positive effect on word retrieval. Yet, a second experiment conducted within the same study (Ravizza, 2003) concluded the opposite. 38 participants were given one minute to retrieve as many words as possible starting with the letter Q. No instructions were given as to whether they could or could not gesture at this stage. Next, one more minute was given to recall additional words, and during this task participants were asked to either move or keep still. The results showed the benefit of not moving on fast word retrieval. Having considered these experiments, Ravizza (2003) concluded that beat movements may be beneficial in cases when lexical items have already been selected but need additional activation.

All in all, the studies show that gestures are beneficial for L2 language acquisition (e.g., Kelly et al., 2009; Macedonia et al., 2011; Tellier, 2008; Quinn-Allen, 1995), however there are contradictory results concerning the type of gestures (beat gestures specifically) which are related to acquisition (see Section 1.6 for more detailed information), which reveals that more research is needed to investigate the role of beat gestures in the field of second language learning.

We now turn to the more narrow-focused themes in relation to the studies in this PhD, starting with beat gestures and memorization.

In order to narrow down the focus of our review, next we proceed to discuss the previous literature regarding the issue of information memorization in one's native language (1.4.); word learning in one's second language (1.5.); and, finally, on second language pronunciation improvement (1.6.). We conclude the section by discussing the general objectives of the thesis, the research questions and the hypotheses (1.7).

1.4. The benefits of beat gestures on information memorization

Starting with Cohen & Otterbein (1992), there is a growing body of research on how gestures influence information memorization in both adults and in children. In their research, Cohen and Otterbein demonstrated that adult subjects who were exposed to sentences accompanied by representational gestures remembered more sentences than subjects who were exposed to speech only, and two subjects who where exposed to sentences accompanied by non-meaningful gestures, like shaking hands. In the experiment participants watched a video with sentences in their native language and then had to write down as many sentences as they were able to remember in a free recall task. In a similar experimental setup, Feyereisen (1998) asked adult participants to remember sentences in three conditions: without gestures, with iconic gestures and with iconic gestures that did not match the context of the sentences. The results of the experiment showed that the sentences that were accompanied by meaningful iconic gestures were recalled better.

Research also shows the beneficial effects of representational gestures on information memorization in children. In the study by Tellier (2005) 4-5year-old children were divided into three groups. Children watched three videos that contained a list of 10 words in their native language. One group watched videos where the words were pronounced by the experimenter. Another group watched videos in which words were pronounced by the experimenter and accompanied by pictures. The third group watched videos where words were accompanied by representational gestures. The results showed that children remembered words significantly better when pictures and gestures accompanied them. There was no significant difference between the effect of the pictures and the gestures, since gestures affected only the visual modality, (i.e., they were only looked at and not produced). A second study by Tellier (2007) examined whether producing gestures in L1 had an additional beneficial effect on children´s memory in comparison to only looking at iconic gestures. 5-6 year-old children were divided into three groups. One group listened to the words and repeated them. The second group listened to the words accompanied by gestures but only repeated the words. The third group listened to the words accompanied by gestures and repeated the words and the gestures. The results of a free recall task conducted afterwards showed that the group that repeated the words together with gestures remembered significantly more words compared to the other two groups.

The results of the above mentioned studies confirmed that representational gestures have a beneficial effect on memory both in adults and in children, in the case of the participants tested. However, with respect to the mnemonic effect of beat gestures, there are still some contradictory results in the literature. Feyereisen (2006) argued that beat gestures might not enhance memory recall. In his experiment, he examined the mnemonic effect of three types of gestures defined as: meaningful gestures (i.e. representational or iconic gestures), non-meaningful (i.e., gestures in which the referent was hard to identify), and beat gestures, and detected no effect of non-meaningful gestures on memory. However, non-meaningful gestures and beat gestures were both grouped together as non-representational gestures and it was therefore impossible to analyze their effects separately. A more recent study by So et al. (2012)

aimed to find out by means of two experiments whether beat gestures would improve word recall in adults and in children. In the first experiment, 30 adults were exposed to three different lists of verbs in three conditions (iconic gestures, beat gestures, and no gesture). They were exposed to 10 words in total. In each condition a participant watched a video on a computer screen and then had to recall as many words as possible. The results of the experiment showed that both iconic and beat gestures enhanced word recall in adults. In the second experiment, the same procedure was applied to children aged 4-5 years. The number of words was reduced from 10 to 5 to accommodate the mnemonic span of the children. The results showed that iconic gestures enhanced memory recall in children, but no effect of beat gestures was detected. All in all, the study by So et al. (2012) proves that beat gestures also enhance memory recall in adults, but not in children. However, we must note that So et al.'s (2012) lists of words accompanied by beat gestures were presented in isolation and without a discourse context; thus, the crucial property of beat gestures to highlight the most prominent part of the discourse was lost, as each item in a row was emphasized. Consequently, it may be stated that the absence of a pragmatically relevant context for the experimental stimuli might explain the non-beneficial effects found for beat gestures in these two studies. Importantly, prosodic prominence was not controlled in any these studies.

By contrast, other studies seem to confirm the opposite, e.g., that beat gestures do support word recall in a first language by children.

Austin and Sweller (2014) tested whether beat gestures would help 3- to 4-year-old children and adults recall of spatial directions. In the experiment children and adults were given verbal route directions to follow a path through a small-scale spatial array, with the speaker using no gestures, beat gestures, or representational gestures. Children recalled the directions better when they were accompanied by either a beat or a representational gesture than when there was no gesture. A more recent study by Igualada et al. (2017) confirmed that 3 to 5 year old preschool children recalled more words in a pragmatically relevant discourse context when they were accompanied by beat gestures in comparison to words that were not accompanied by gestures. Similarly, a recent study by Llanes et al. (under review) has also shown that preschool children remember and comprehend information better in a set of target discourses when they are accompanied by prominence in both speech and gesture. Importantly, in all the above-mentioned experiments, the beat gestures in the experimental materials were associated with prosodically prominent words in the discourse.

The specific influence that beat gestures have on the encoding of information by listeners is still largely unknown. Moreover, given that beat gestures are typically associated with prominent positions, previous experimental studies have not teased apart whether the effects of beat gestures are mainly due to the presence of concomitant prosodic prominence or not.

Independent evidence has shown that there is a positive effect of prosodic prominence on both the ability to recall novel information and on the comprehension of information (e.g., Birch and Clifton, 1995, 2002; Bock and Mazzella, 1983; Dahan, Tannenhaus & Chambers, 2002; Fraundorf et al., 2010). For example, during the investigation of the effects of pitch accent on memory, Fraundorf et al. (2010) asked participants to listen to short recorded stories that included two contrasted items, specified in the continuation of the story. In the continuation, the pitch accent of the target item was manipulated so that it ranged from non-contrastive (H*) to contrastive pitch accentuation (L+H*). The results of a recognition memory test conducted after the training session showed that the contrastive L+H* pitch accentuation triggered a facilitation effect on the recall of contrastive focus.

In sum, while representational gestures encode semantic information that can help learners memorize novel words, it is not clear whether beat gestures, and also their concomitant prosodic prominence features, can also be of help in the vocabulary memorization process. Thus, the aim of Study 1 in this thesis will be to investigate the effects of prosodic prominence (i.e., pitch accentuation) and visual prominence (i.e., beat gestures) on behavioral tasks involving word memorization by adults.

1.5. The benefits of beat gestures on L2 word learning

The organization of vocabulary knowledge and its influence on L2 proficiency has received recognition in the last few decades and has become a significant part of the study of L2 acquisition (e.g., Adams, 1980; De Jong, Steinel, Florijn, Schoonen & Hulstijn, 2012; Ellis, 1994; Milton, 2009; Meara & Miralpeix, 2017; Miralpeix & Meara, 2014; Nation, 1990, 2001). In this context, several authors have addressed issues related to vocabulary learning as well as teaching strategies (e.g., Muñoz, 2013). For example, starting from 1990s there is a growing body of research that centers on methods of vocabulary instruction and the effect of different practice activities (such as bottom-up, top-down, structured, unstructured; e.g., Sanaoui, 1995)

With regard to the role of gestures in L2 word learning, research has thus far centered largely on the role of representational gestures rather than beat gestures, as they convey semantic content of speech. A number of studies have shown that items accompanied by meaning-related representational gestures can be learned faster in an L2, as they facilitate word-meaning associations (e.g., Kelly et al., 2009; Macedonia et al., 2011; Tellier, 2008; Quinn-Allen, 1995).

The study by Macedonia et al. (2011) compared the effect of iconic vs. meaningless gestures (e.g., toughing the head, stretching the leg) produced by instructors on noun recall in students of a foreign language. In this study, 33 German-speaking participants were trained to remember 92 nouns from a corpus of artificial words that was created according to Italian phonotactic rules. In the training session, words were accompanied by either iconic or meaningless gestures. Participants were trained over four days and then performed a recall test on the fifth and sixth days. The results showed significantly better recall of words accompanied by iconic gestures in both short- and long-term memory tests. Also, Macedonia and Knösche (2011) conducted a study that investigated the impact of gestures on abstract word learning. The authors elaborated 32 sentences, each consisting of 4 grammatical elements (subject noun, verb, adverb and object noun). 16 sentences were to be memorized audio-visually without gestures and 16 were enacted, i.e., accompanied by a symbolic gesture (for example for the word meaning ‘theory‘, the actress in the video simulated the opening of a book in front of her). The experiment lasted for 6 days, with a daily assessment of memory performance. The enacted option enhanced memorization not only for concrete but also for abstract nouns. Participants were also asked to produce new sentences with the words they had learned, and it turned out that enacted items were recruited more often than audio-visually learned vocabulary.

Further support of the importance of iconic gestures for L2 word learning was presented by Kelly et al. (2009). In their study, 12 Japanese verbs with common everyday meanings were presented to adults with no previous knowledge of Japanese. The words were presented in blocks of three in the following four conditions: (a) speech, (b) speech + congruent gesture (for example, showing the gesture of drinking while saying the verb “drink”), (c) speech + incongruent gesture (showing the gesture of washing one´s face while saying the verb “drink”), and (d) repeated speech. The results showed that the most positive effect on word learning was achieved when the items were presented in the speech + congruent gesture condition.

Tellier (2008) conducted a study to test the impact of gesture on second language word learning. Twenty French children learning English of an average age of 5.5 years, were presented with 8 common English words, four of them were accompanied by gestures and four of them by pictures. In the gesture condition, children had to reproduce gestures while repeating words. It turned out that enacted items were memorized better than those presented with the pictures, thus confirming the impact of gesture production on word memorization.

Another line of research has explored the beneficial use of a specific type of visuospatial gestures, the so-called pitch gestures that mimic or represent the melody of speech, on L2 lexical tone discrimination and word learning in a target tonal language. For example, in the study by Morett and Chang (2015), 57 English monolingual participants were asked to learn a total of 20 Mandarin words that were accompanied by either hand gestures illustrating the shape of the tone pitch (pitch gesture – defined by Morett and Cang (2015)), semantic (representational) gesture conveying the meaning of the word, or no gesture. The results showed that pitch gestures helped subjects distinguish between the meanings of Mandarin words that varied only in tone. These findings provide evidence that the visuospatial features of such pitch gestures might facilitate the discrimination between Mandarin words differing in their lexical tones and thus indirectly enhance L2 word learning. However, Morett and Chang's study did not observe the potential effects of pitch gestures on pitch discrimination abilities. In contrast, a recent study by Baills et al. (under revision), has confirmed that observing and producing pitch gestures favors both tone discrimination and lexical word identification and recognition by non-tonal learners of Chinese.

To our knowledge, no research has specifically focused on the issue of whether beat gestures (gestures that convey prosodic rather than semantic information) can also enhance L2 word learning. Moreover, previous experimental studies have not teased apart whether the effects of beat gestures are mainly due to the presence of concomitant prosodic prominence (as beat gestures are typically associated with prominent prosodic positions) or not. Study 2 in this PhD thesis has the goal of assessing the effects of the presence or absence of visual prominence (i.e., beat gestures) combined with prosodic prominence (i.e., pitch accents) on L2 novel vocabulary learning.

1.6. The benefits of beat gestures on L2 pronunciation learning

In the field of second language acquisition, pronunciation instruction is typically centered on teaching segmental aspects of speech, that is, on teaching the pronunciation of individual phonemes (see Derwing & Munro, 2015, and Mora & Levkina, 2017) for a review). For example, the study by Mora & Levkina (2017), investigated whether a treatment geared toward improving L2 perception of target vowels contrast would be effective for the production of the target vowel contrasts and whether the improvement might be generalized to new items or speakers. Results after the treatment showed that there was an improvement in perceptual sensibility to the vowel contrasts and that there was a generalization to new non-words. Another study by Aliaga-Garcia & Mora (2009) investigated the effects of phonetic training on the perceptual and productive competence of learners of English as an L2, focusing on four L2 sound contrasts that Catalan and Spanish bilinguals have reported to be difficult (Cebrian, 2002, 2006; Mora, 2007; Mora & Fullana, 2007). Results confirmed that learners produced and perceived the target pairs of sounds significantly more accurately after the training, i.e., the training sessions were successful for the experimental group in the improvement of pronunciation accuracy.

Importantly, there is evidence that suprasegmental deviances have a larger influence on L2 accentedness, comprehensibility and intelligibility ratings than segmental deviances (e.g., Anderson-Hseih, Johnson, & Koehler, 1992; Edmunds, 2010; Field, 2005; Kang, 2010; Ulbrich, 2013; Munro & Derwing 1995, 1999; Derwing & Munro, 1997, van Maastricht, Krahmer and Swerts 2015; White and Mattys 2007). The study by van Maastricht, Krahmer and Swerts (2015) showed that L1 speakers of Dutch were able to distinguish recordings of native speakers from non-native ones based on prosodic cues alone. A series of well-known studies by Munro and Derwing (Munro & Derwing 1995, 1999; Derwing & Munro, 1997) that analyzed the relative contribution of segmental and suprasegmental features to native English speakers' judgements of accentedness, intelligibility, and comprehensibility. The results of these studies showed that prosodic errors of second language speakers strongly affected the native language speakers' ratings of accentedness and comprehensibility. Kang's (2010) experiment analyzed the speech of 11 international teaching assistants using native speaker ratings and confirmed that suprasegmental features are responsible for a good amount of variability in accentedness ratings of foreign language speech by native speakers. Specifically, 41% of variance in accentedness ratings was due to pitch range, word stress, and mean length of pauses. Only 35% of variance was due to speech rate alone. Following this line of research, White and Mattys (2007) also found that rhythm had a significant influence on native speaker's ratings of foreign L2 speech. In their work, they found that in the case of three previously-identified rhythm metrics (VarcoV, nPVI-V, and %V), all were significantly correlated with native English speakers' ratings of foreign accent in L2 speech. One of these metrics in particular, VarcoV, was found to be an effective predictor of accent ratings, especially when coupled with speech rate. The rhythm of L2 speech, in other words, was shown to play a significant role in the extent to which a speaker is judged as being native.

A number of experimental classroom studies in second language acquisition have demonstrated that it is important to teach suprasegmental components in order to improve overall fluency and comprehensibility in language learners' speech (e.g., Berhman, 2014; Derwin, Munro and Wiebe, 1998; Derwing & Rossiter, 2003, Derwing et al., 2003; Gordon, Darcy and Ewert, 2013). The ESL classroom study by Derwin et al. (1998) confirmed that global pronunciation instruction (e.g., a type of instruction that addresses speaking rate, intonation, rhythm and stress both at the word and the sentence levels) during 11 weeks had significantly better results in comparison to learners who took only segmental or no pronunciation instruction. A more recent study by Gordon, Darcy and Ewert (2013) also confirmed that overall comprehensibility after 3 weeks of training improved significantly in the group that received suprasegmental instruction (focused on stress, rhythm, linking of sounds and reductions). Behrman (2014) conducted the study that compared the effects of segmental and prosody training on reducing speakers´ foreign accent. Segmental training focused on the articulation of consonants, while prosodic training focused on four prosodic utterance levels: rise-fall pitch in one-word utterances, rising, falling, and rise-fall intonation in three-word utterances, informational and yes/no questions, and prosodic rhythm of longer utterances. The difference in this study approach was that it used individualized instruction, and accuracy was assessed after each session. While the segmental training focused on the articulation of consonants, the prosodic training centered on four prosodic utterance levels (rise-fall pitch in one-word utterance, rising, falling, and rise-fall intonation in three-word utterances). The results of the study showed that a combination of both types of instruction produced the most successful outcomes in English learners.

The studies mentioned above provide evidence of the importance of suprasegmental instruction in L2 pronunciation teaching. However, to our knowledge, there is almost no work testing different training paradigms and there is little evidence on the comparison between one suprasegmental training method over another. Most suprasegmental trainings involve instructions and analysis of prosodic targets, but not concrete, empirically-tested exercises designed to promote pronunciation improvement in L2. Also, there is no agreement on how to implement rhythm instruction in ESL teaching. One of the methods proposed for ESL classroom is Graham's (1978) Jazz chants where learners are asked to finger-tap out the beats of short, poem-like structures in order to follow the target rhythm. Still, there is no assessment that empirically shows the improvement of this method. This finger-tapping method was also used in Derwing, et al. (1998) and Derwing and Rossiter (2003) studies. However, besides this method, very little has been investigated on how gestures may be used as a tool for L2 rhythm and pronunciation training.

Recent studies have assessed the effectiveness of the use of gestures on pronunciation learning, specifically in regards to a set of phonological processes such as tonal and intonation learning, as well as phoneme duration, with controversial results.

On the one hand, one group of studies has explored the beneficial use of the so-called pitch gestures (or metaphoric gestures that mimic or represent the melody of speech) in learning L2 tones and intonation, with positive results (Hannah, Wang, Jongman & Sereno, 2016); Yuan, González-Fuente, Baills, and Prieto (in press), Morett & Chang, 2015). In their study Hannah, Wang, Jongman & Sereno (2016) looked at how pitch gestures affect non-native Mandarin tone perception by testing 25 English speakers on tone identification. Participants in the study listened to two monosyllabic words with the four tones under four conditions: audio-facial/congruent, audio-facial/incongruent, audio-facial-gestural/congruent and audio-facial-gestural/incongruent and then immediately decided if they heard a "level", dropping, "rising" or "falling" tone. The results of the study showed that participants in the audio-facial-gestural/congruent condition obtained significantly better scores at tone identification than participants in the audio-facial/congruent condition. The study by Yuan, González-Fuente, Baills, and Prieto (in press) investigate how pitch gesture observation can help learning difficult Spanish intonation pattern in the case of sixty-four Mandarin Chinese learners with a basic level in Spanish. Half of the participants received intonation training without gestures while the other half received the same training with pitch gestures representing nuclear intonation contours. Results showed that observing pitch gestures during the learning phase improved the learners' production outcomes significantly better than training without gestures.

Another line of research has investigated the effect of pitch gestures on L2 lexical tone discrimination and word learning in a target tonal language (e.g., Morett and Chang, 2015. In the study by Morett and Chang (2015), 57 English monolingual participants were asked to learn a total of 20 Mandarin words that were accompanied by either hand gestures illustrating the shape of the tone pitch (pitch gesture), semantic (representational) gesture conveying the meaning of the word, or no gesture. The results showed that producing pitch gestures helped subjects distinguish between the meanings of Mandarin words that varied only in tone. These findings provide evidence that the visuospatial features of such pitch gestures might be facilitating the discrimination between Mandarin words differing in the lexical tones and thus indirectly enhance L2 word learning. However, in Morett and Chang's study the availability of pitch gestures failed to improve the participants' performance in a pitch identification task. By contrast, a recent study by Baills et al (under revision) has confirmed that both observing and producing pitch gestures favored both tone discrimination and lexical word identification and recognition by non-tonal learners of Chinese. In the study, Catalan native participants with no previous knowledge of Chinese were asked to observe (Experiment 1) and produce pitch gestures (Experiment 2) during a short multimodal training session on Chinese tones and words. Participants were tested on tone identification and word learning after the training sessions. Results of the study showed the positive effects of pitch gesture observation and pitch gesture production in comparison to no gesture condition.

Another group of studies (Hirata & Kelly, 2010; Hirata et al., 2014, Kelly et al., 2017) investigated the role of some types of rhythmic gestures that metaphorically map the duration of vowel sounds and no effect of these gestures was found on the perception of mora´s length in Japanese. The study by Hirata and Kelly (2010) investigated the role of co-speech gesture perception in auditory learning of Japanese vowel length contrasts. In the study participants were exposed to videos in which Japanese speakers produced Japanese short and long vowels with and without hand gestures that were associated with the rhythm of those vowels. A short vertical chopping movement was used for marking short vowels and a long horizontal sweeping movement was used for marking long vowels. The results of the experiment showed that there was no difference when participants learned vowel length viewing the videos with gestures in comparison to the condition without gestures. The study by Kelly, Hirata, Manansala and Huang (2014) and Hirata, Kelly, Huang, and Manansala (2014) explored whether hand gestures influence auditory learning of an L2 at a segmental level. For this purpose, they carried out an experiment in which English speakers were trained to learn Japanese words by either observing or producing gestures that coincided with syllable or mora, as half of the gestures metaphorically represented the information on syllable structure and half of the gestures represented the information on Japanese mora structure. The main finding of the study was that participants across four conditions (Syllable Observe, Syllable Produce, Mora Observe, and Mora Produce) performed similarly in the auditory identification and vocabulary test performed after the training sessions. Thus, the authors of the previously mentioned studies concluded that there is a limited effect of hand gestures of learning durational contrasts in a second language, in the case of the participants tested.

There might be a set of reasons that can explain the discrepancy between the results of the above mentioned studies. First, as Kelly, Bailey and Hirata (2017) note, while pitch gestures have been found to have a systematic positive effect on learning L2 pitch differences, this is not the case with length/duration gestures. Indeed, Kelly, Bailey and Hirata (2017) explored the role that metaphoric gestures play in perceiving foreign language speech sounds that differ according to length and intonation. English-speaking adult participants were exposed to videos with a trainer producing Japanese length contrasts and sentence final intonation distinctions accompanied by congruent metaphoric, incongruent and no gestures. The results showed that for intonation contrasts, congruent metaphoric gestures had a positive effect, as identification was more accurate in comparison to other conditions. For the length contrast identification, however, these results were not obtained, and no clear and consistent pattern emerged. In fact, congruent metaphoric gestures made length contrasts identification more difficult.

We also think that the difference in the type of gestures used in Kelly, Hirata and colleagues studies (Hirata & Kelly, 2010; Hirata et al., 2014, Kelly & Lee, 2012; Kelly et al., 2017) might have had an influence on the results obtained in their study. Even though the authors interpreted them as suggesting a lower limit of speech-gesture integration, they also pointed out that there could have been more effective types of gestures (Kelly et al., 2014). Specifically, the mora gestures used in the studies by Kelly, Hirata and colleagues (e.g., the short vertical chopping movements) may be considered as "non-intuitive" to English speakers and act as an incongruent combination of speech and gesture that might impede learning durational information in the second language (see Kelly et al., 2009).

Against such a backdrop it seems evident that more research is needed to investigate whether another type of rhythmic gesture (e.g., beat gestures integrated with prosodic prominence) might help pronunciation improvement in a second language. McCafferty (2006) pointed out the strong relationship between beat gestures and emerging second language prosody, however there is little empirical evidence on the role of these types of gestures on the acquisition of suprasegmental elements, in other words, teaching a rhythm in a foreign language. Even though this kind of gesture appears naturally in spontaneous speech, there is still little evidence on the effects of beat gestures on the acquisition of suprasegmental elements. To our knowledge, the only study investigating this issue is Gluhareva and Prieto's (2017) article, which showed that beat gestures used to mark the rhythm of speech are an effective aid for L2 acquisition of pronunciation. The study consisted of a brief within-subjects training, during which participants were asked to watch an English instructor produce a set of target sentences in English framed in a discourse situation. Some discourses were accompanied by rhythmic beat gestures and some were produced without gestures. Twenty Catalan participants improved their accentedness significantly on the most difficult trained items when the discourse was accompanied by rhythmic beat gestures.

Study 3 in this thesis investigates whether participants show higher gains in accent improvement if they are instructed to imitate the experimenter and produce beat gestures themselves in comparison to only observing beat gestures. First, we believe that the presence of visible and natural beat gestures working together with prosody can represent a further benefit for boosting rhythmic information in an L2 and thus favoring pronunciation learning. Second, there is large evidence that producing gestures in comparison to only observing them results in more beneficial effects for general learning processes (e.g., Beilock and Goldin-Meadow, 2010; Goldin-Meadow, 2014; Goldin-Meadow, Cook & Mitchell, 2009). Studies within the embodied cognition perspectives have demonstrated that physically producing actions leads to better learning and memory than just observing them (e.g., Saltz and Donnenwerth-Nolan, 1981; Cohen, 1989). As mentioned in section 1.3.3, studies in the field of second language research have also provided evidence that gesture and speech interact during L2 speech production (e,g, Gullberg, 2006; Gullberg et al., 2008) and producing gestures has a facilitative effect on word learning (e.g., Tellier, 2008; Macedonia et al., 2011). Still, to our knowledge no previous research has addressed whether methodologies involving embodied cognition practices are also more effective for pronunciation learning

1.7. General objectives, research questions and hypothesis

This dissertation aims to investigate the role of a specific type of gesture aligned with prosodic prominence of speech, called beat gestures. Specifically, we are interested in examining whether there is a potential beneficial effect of beat gestures on word memorization in one's native language, on word learning in one's second language and on pronunciation improvement in one's second language.

The following three research questions will be addressed, each one in a separate chapter:

1) Do beat gestures have additional beneficial effects on adult's word recall in contrastive discourse compared to prosodic prominence without beat gestures? (Study 1)

2) Do adult participants learn more words in a second language when they are presented with prosodic and gestural prominence (beat gestures)? (Study 2)

3) Does beat gesture production improve second language pronunciation more in comparison to beat gesture observation? (Study 3)

Our general hypothesis is that beat gestures aligned with prosodic prominence have a beneficial effect on learning. Our specific hypotheses are the following. In the first study (Chapter 2), adults will benefit from beat gestures presented together with prosodic prominence in a word recall task in a first language. In the second study (Chapter 3), beat gestures and prosodic prominence will have a beneficial impact on second language word learning. In the third study (Chapter 4), the production of beat gestures used to mark the rhythm of second language will be beneficial for second language pronunciation improvement.

Study 1 (Chapter 2) investigates the role of prosodic prominence (pitch accents) and gesture prominence (beat gestures) on the recall of information. Twenty Catalan-dominant native speakers were asked to watch 48 short videotaped discourses each containing two contrast sets with two items (e.g., the fish shop and the grocery shop and snow and ice) placed at the beginning and at the end of the discourse. In the sequence the critical word from one of the contrast set was presented under two experimental conditions: 1) accompanied by prosodic prominence (L+H* pitch accent); and 2) accompanied by prosodic prominence and gestural prominence (L+H* pitch accent + beat). The item from the second contrast set was unaccented (L* pitch accent). The results of the recall task revealed that the presence of prosodic prominence favored the recall of contrastive information in comparison to unaccented information. Furthermore, beat gestures associated with prosodic prominence added a beneficial effect to information recall in comparison to the condition without beat gestures.

Study 2 (Chapter 3) investigates the effects of prosodic prominence (e.g., focal pitch accent) and visual prominence (e.g., beat gesture) on L2 novel vocabulary acquisition. Ninety-six Catalan-dominant native speakers were asked to learn 16 Russian words that were either accompanied by prosodic prominence in speech (L+H* pitch accent) or no prosodic prominence and either accompanied by visual prominence (beat gesture) or no visual prominence. Thus, participants in a within-subject design were exposed to four conditions that resulted from the combination of two factors +/ - spoken prominence and +/ - gestural prominence. The results of recall and recognition tasks conducted after the training session showed that there was a positive effect of visual prominence only when accompanied by prosodic prominence.

Study 3 (Chapter 4) investigates whether beat gesture production leads to higher gains in accent improvement compared to only observing beat gestures. In a between-subject, pre-post training design, twenty-six Catalan speakers watched a training video in which an instructor gave spontaneous responses to English discourse prompts accompanied by beat gestures. While one group of participants simply repeated the discourses, another group of participants repeated discourses and accompanied their speech with the gestures seen in the videos. The speech of the participants was elicited before and after training and was assessed for accentedness by 6 native speakers. The analysis of the accentedness ratings in the pre-training and post-training speech samples demonstrated that participants in the gesture-produce group received significantly higher scores in accentedness ratings than participants in the gesture-observe group. The results of this study demonstrate that producing beat gestures leads to higher gains in accent improvement in comparison to only observing beat gestures.

Kushch, O., & Prieto, P. (submitted). Beat gestures increase the effects of prosodic prominence on information memorization in a first language: *Journal of Phonetics.*

2. CHAPTER 2: “Beat gestures increase the effects of prosodic prominence on information memorization in a first language”

2.1. Introducción

Human speech is seldom monotonous. When we speak, we naturally attempt to focus the listener’s attention on the most important parts of our discourse, and we can achieve this goal by instantiating prominence patterns in two complementary domains, speech and gesture. Prosodic prominence assignment (or pitch accentuation) is one of the strategies most commonly used by languages to focus information in a discourse (see Cutler, Dahan & van Donselaar 1997, for a review). Recent research in audiovisual prosody has shown that prosodic prominence is typically associated with more prominent gestural and articulatory features (e.g., beat gestures, head nods, eyebrow movements, exaggerated articulation, etc.; see for example Swerts & Krahmer, 2008; Prieto, Puglesi, Borràs-Comes, Arroyo & Blat, 2015). Beat gestures are rhythmic hand and arm movements that serve as visual highlighters of information, and “mark the word or phrase they accompany as being significant (…) for its discourse pragmatic content” (McNeill, 1992:15). Beats, together with pitch accentuation, have been associated with focus marking and discourse structure marking functions in speech (e.g., Loehr, 2012; Shattuck-Hufnagel, Ren, Mathew, Yen & Demuth, 2016).

Independent evidence has shown that the appropriate use of prosodic prominence has a positive effect on both the ability to memorize novel information (Fraundorf, Watson & Benjamin, 2010) and the comprehension of information by adults (Bock & Mazzella, 1983; Birch & Clifton, 1995, 2002; Dahan, Tannenhaus & Chambers, 2002). In their investigation of the effects of pitch accent on memory, Fraundorf et al. (2010) asked participants to listen to short recorded narratives that included two contrasting items *(*e.g., *British scientists and French scientists).* In the second part of the narrative, the pitch accent of the target item was manipulated so that it ranged from non-contrastive (H*) to contrastive (L+H*). The results of a subsequent recognition memory test showed that the contrastive L+H* pitch accent triggered a facilitation effect on the memorization of novel information.

Similarly, Bock and Mazzella (1983) investigated the role of pitch accentuation in auditory sentence comprehension. Forty native speakers of English were exposed to 20 sets of four sentence pairs. The target sentence pairs consisted of a simple negative declarative sentence followed by a target positive sentence (e.g., *Arnold didn't fix the radio. Doris fixed the radio.*). The target sentence pairs in the four sets were identical except for the location of the pitch accent within the sentence. The sentence pairs were presented under four experimental conditions: appropriate target accent (the content word in the context sentence that was prominently stressed had the same syntactic role as the changed and accented word in the target sentence, e.g., *ARNOLD didn't fix the radio. DORIS fixed the radio*), inappropriate target accent (stress fell on a word in the context sentence that had a different syntactic role than the stressed word in the target sentence, e.g., *Arnold didn't FIX the radio. DORIS fixed the radio*), no context accent (none of the words in the context sentence received special emphasis, e.g., *Arnold didn't fix the radio. DORIS fixed the radio*), and control (no accent in either the context or the target sentences, e.g., *Arnold didn't fix the radio. Doris fixed the radio*). The results showed that sentences in the appropriate target accent condition were understood faster than sentences in the other three conditions and thus the presence of an appropriate pitch accent in the context sentences facilitated understanding of subsequent targets. Moreover, comprehension time was faster when the accent fell on the information focus and not on another part of the sentence.

In a related study, Birch and Clifton (1995) aimed to investigate whether the relationship between discourse structure and appropriateness of pitch accent location affected comprehension processes. In their experiment, participants were asked to perform two tasks. First participants listened to question-answer pairs and had to decide as quickly as possible whether they made sense as a unit of conversation. The question in the question-answer pairs consisted of a broad focus question, such as *Isn't Ketty pretty smart?*, which was followed by either a single-accented response (e.g., *She teaches MATH)* or a double-accented utterance (e.g., *She TEACHES MATH*). Participants accepted both singl-accented and double-accented responses, though the ratings for the double-accented utterances were higher than for the singl-accented utterances. These results constitute clear evidence that pitch accent marking of focus can facilitate comprehension processes. The findings of this study were later extended in Birch and Clifton (2002). In six new experimental studies they explored the influence of syntactic argument structure and pitch accent patterns - in particular the accentuation of phrasal adjuncts - on the acceptability of spoken utterances. In the experiments listeners were asked to provide judgments about the prosodic appropriateness and comprehensibility of a set of dialogs in which focus on the adjunct information was manipulated. All six experiments showed that adjuncts can not project prosodic focus. All in all, the results of these studies show that an appropriate use of prosodic prominence positively influences both the recall and the comprehensibility of information.

On the other hand, there is also empirical evidence that beat gestures are beneficial for the recall of information. Studies have confirmed the positive effect of beat gestures for the recall of lexical information by both adults (So, Sim Chen-Hui & Low Wei-Shan, 2012) and children (Igualada, Esteve-Gibert & Prieto, 2017; Austin & Sweller, 2014; Llanes, Kushch, Borràs-Comes, Prieto,). The study by So et al. (2012) aimed to find out by means of two experiments whether beat gestures would improve word recall in adults and children. In the first experiment, 30 adults were exposed to three different lists of verbs in three conditions (accompanied by iconic gestures, beat gestures, or no gestures). In each condition a participant watched a video on a computer screen and then had to recall as many words as possible. The results of the experiment showed that both iconic and beat gestures enhanced word recall in adults. In the second experiment, the same procedure was applied to children aged 4-5 years. The number of words was reduced from 10 to 5 to accommodate the shorter mnemonic span of the children. The results showed that while iconic gestures enhanced memory recall in the children, no effect of beat gestures was detected. In short, the study proved that beat gestures enhanced memory recall in adults but not in children. By contrast, Austin and Sweller (2014) proved that beat gestures can be of help in the recall of spatial directions in 3- to 4-year-old children. In their experiment, children recalled information about spatial directions better when the spatial directions were accompanied by beat gestures. A more recent study by Igualada et al. (2017) likewise showed that 3- to 5-year-old children were better at recalling words when they were presented with a beat gesture. Taken together, these results support the idea that beat gestures do help children to recall prominent words presented in a relevant discourse. Similarly, a recent study by Llanes et al. (under review) has also shown that preschool children remember and comprehend information better in a set of target discourses when they are accompanied by prominence in both speech and gesture. Importantly, in all the abovementioned experiments, the beat gestures in the experimental materials were associated with prosodically prominent words in the discourse.

Importantly, however, to our knowledge no studies have attempted to properly disentangle the role of these markers of prominence on information memorization. First, we know that there is a strong temporal connection between beat gestures and pitch accentuation and that they tend to appear together in speech. Many authors have pointed out that the most prominent part of co-speech gestures (the gesture stroke or apex) is temporally aligned with prominent parts of speech (i.e., accented syllables, intonation peaks) (e.g. McNeill, 1992; see Esteve-Gibert & Prieto, 2013 for a review). For example, when they analyzed a corpus of videotaped academic lectures, Yasinnik, Renwick, and Shattuck-Hufnagel (2004) found that in 90% of instances the gesture apexes in beat gestures coincided with a pitch-accented syllable.

Similarly, there is evidence that visual and prosodic cues mutually influence each other in the perception of prominence. Krahmer and Swerts (2007) found that prosodic prominence and visual information work in a complementary fashion and that adding visual beats to prosodic prominence leads to a stronger perception of prosodic prominence. In this experiment, participants were exposed to the sentence "*Amanda goes to Malta*" in which the two target words were associated with combinations of the presence or absence of prosodic and gestural prominence. When one of the target words was associated with a manual beat gesture, regardless of the presence or absence of prosodic prominence, this increased its perceived prominence and at the same time decreased the prominence of another target word. Nonetheless, despite these few studies, much remains to be learned about the relative contribution of prosodic prominence and beat gestures to memory or comprehension processes.

The goal of the present study is therefore to investigate the effects of prosodic prominence (i.e., pitch accentuation) and visual prominence (i.e., beat gestures) on the recall of contrastive information by adults. Borrowing the word recall paradigm from Fraundorf et al. (2010), we exposed participants to 48 video-recorded discourses with contrastive items that were presented under the following experimental conditions: 1) prominence in both speech and gesture (L+H* pitch accent + beat gesture), 2) prominence in speech only (L+H* pitch accent), and 3) prominence in neither speech nor gesture. We hypothesized that (1) words associated with prosodic prominence in the form of pitch accentuation should lead to stronger memory recognition scores than words with no prosodic prominence; and (2) adding gestural prominence in the form of beat gestures to prosodic prominence should lead to stronger mnemonic effects.

Importantly, we did not include a condition combining beat gestural prominence without prosodic prominence, since this combination is rarely present in natural speech. Research has demonstrated that naturally produced beat gestures are almost invariably linked to prosodic prominence in speech (Shattuck-Hufnagel et al., 2016; Yasinnik et al., 2004; see Wagner, Malisz, & Kopp, 2014 and Jannedy & Mendoza-Denton, 2005 for a review). As noted above, Yasinnik et al. (2004) showed that 90% of the beat gestures that appeared in a corpus of academic lectures were associated with pitch accents. We think that this specific combination (gestural prominence with prosodic prominence) could trigger a perception of unnaturalness and thus increase processing integration costs. In fact, a recent ERP study conducted by Dimitrova, Chu, Wang, Ozyurek, and Hagoort (2016) showed evidence of the increased computation costs of combining beat gestures with non-focused (e.g., non-prominent) words. In this experiment, words focused by prosodic prominence and words focused by beat gesture were processed more attentively than non-focused words, as they elicited a N1 and P300 component. Importantly, the combination of non-focused words accompanied by beat gestures gave rise to a late positivity 600-900 ms relative to target word onset. The authors attribute these results to the increased processing costs of this combination, and point out that "the late positivity in our study may thus reflect increased computation costs needed to arrive at a coherent interpretation of the message when beat gesture emphasizes non-focused information, which should not have been highlighted" (p. 23).

However, despite this evidence at the neural processing level, to our knowledge no previous studies have been conducted on the expected positive effects of gestural beat prominence on behavioral cognitive measures. Our study will run a memory task controlling for two levels of multimodal focusing of critical words, namely prominence in speech and prominence in speech and gesture simultaneously. We expect that visual prominence will strengthen the memory results obtained by participants. In other words, the combination of prosodic prominence with visual prominence will have the strongest beneficial effect on the recall of highlighted information.

Our hypothesis on the role of beat gesture observation in memory recall in relation to prosodic prominence has a bearing on the predictions of the grounded (or embodied) cognition paradigm, which claims that sensorimotor experiences are involved in cognitive and language processing (e.g., Barsalou, Simmons, Barbey & Wilson, 2003; Barsalou, 2008; Lakoff & Johnson, 1980; Smith, 2005). The behavior of co-speech gestures as important forms of embodiment in language has been usually put forward as an argument in favor of this theory, as such gestures are closely linked to memory and comprehension (Barsalou, 2008; see references above). Gesture is considered an important form of embodiment in language, and it is closely linked to memory (Barsalou, 2008). However, research within the embodied research paradigm has primarily focused on gestures that are associated with referential and meaningful information (e.g., representational gestures), and less is known about the role of beat gestures, which do not have a referential meaning in themselves but merely accompany prosodically focused words

Interestingly, there are studies confirming that pitch gestures - hand gestures describing spatially the melodic curves of tones - have a beneficial effect on learning. For example, the study by Baills, Suárez-González, González-Fuente, and Prieto (submitted) explored the effect on Catalan-speaking participants of observing and producing pitch gestures in their learning of Chinese tones. The results of the study demonstrated that participants were better at discriminating tones and learning tonal words when they observed pitch gestures during training than when they were not shown pitch gestures. Also, when they were told to imitate pitch gestures while learning this enhanced their tone discrimination and tonal word learning more than when they merely observed pitch gestures and repeated words aloud. Another study by Yuan, González-Fuente, Baills, and Prieto (in press) also looked at how pitch gestures can help Mandarin-speakers learn Spanish intonation. In a between-subjects experimental design, half of the participants were exposed to intonation training without gestures while the other half received training that included observing pitch gestures representing nuclear intonation contours. Results showed that observing pitch gestures during the learning phase improved pronunciation skills significantly more than being trained without observing such gestures.

Thus while representational and pitch gestures encode semantic information that can help learners memorize novel words, it is not clear whether beat gestures, which only encode prominence features, can also be of help in the information memorization process. If, independently of prosody, beat gestures positively affect recognition memory, this will strengthen the arguments in favor of the embodied cognition paradigm because it suggests that recognition and comprehension can be enhanced by either visual or prosodic prominence cues. Therefore embodied cognition theory would support our prediction that adding beat gestures as visuospatial information to prosodic prominence will favor information processing.

Another important point to consider is the potential effect of observing beat gestures on memory recall. While the embodied cognition paradigm has concentrated on gesture production, less is known about gesture perception. There is neurophysiological evidence that self-performing a gesture when learning verbal information can help a learner form sensorimotor networks that represent and store the words in both native (Masumoto, Yamaguchi, Sutani, Tsuneto, Fujita & Tonoike, 2006) and foreign languages (Macedonia, Müller & Friederici, 2011). In addition, there is evidence that not only gesture production but also gesture observation leads to the formation of motor memories in the primary motor cortex (Stefan, Cohen, Duque et al., 2005), which is considered a likely physiological step in learning. The study by Stefan et al. study used transcranial magnetic stimulation (TMS) to show that observing another individual performing simple repetitive thumb movements gave rise to a kinematically specific memory trace of the observed motions in the primary motor cortex. After participants were exposed to an extended period of observation of thumb movements, this caused their own TMS-evoked thumb movements to tend to fall within the observed direction. These findings support a role for the mirror neuron system in memory formation and possibly human motor learning. Following this, our hypothesis will be that memory processes will be positively conditioned by both perceptual and motor modalities (e.g., Borghi & Caruana, 2015). We expect memory to be activated on the basis of external states (perception) together with internal states (proprioception, emotion, and introspection) as well as bodily actions (simulation of the sensorimotor experience with the object or event they refer to.

2.2. Methods

2.2.1. Participants

Twenty Catalan-dominant native speaking students (mean age = 20.5 years, SD = 2.327) from the Universitat Pompeu Fabra participated in the study. Participants were asked to complete a language questionnaire and provided written informed consent to process their data. Participants reported using Catalan for an average of 75.6% (SD = 8.7) of their daily communication needs. Participants were each financially remunerated with 5 euros.

2.2.2. Materials

The materials used in the experiment were an adaptation of the materials used by Fraundorf et al. (2010). They consisted of video-recordings of 48 short narratives in Catalan containing two parts. The first part was a context passage, an example of which is given in (1) below, which established two contrast sets containing **two items**. The context passage was followed by a second part of the discourse, that is, a continuation pattern, as illustrated in (2), which mentioned only **one item from each contrast set**. The contrast sets of items differed in their grammatical and thematic roles across narratives.

(1) Context passage: *Ahir l'Esmeralda havia planejat anar a la **peixateria i a la fruiteria**. No obstant això, quan va sortir de casa va veure que la tempesta havia deixat molts carrers coberts de **neu i gel**.*

Translation: *Yesterday Esmeralda was planning to go to the fish and grocery shops. However, when she left her house, she saw a storm had left many of the streets covered with snow and ice.*

(2) Continuation passage: *L'Esmeralda va aconseguir anar a (**la peixateria/la fruiteria**), però com que l'altre lloc estava cobert (**de neu/de gel**), va decidir tornar-hi un altre dia.*

Translation: *Esmeralda managed to get to the (fish/grocery) shop, but since the other shop was covered with (snow/ice) she decided to go back another day.*

The two paired items (e.g., *peixateria/fruiteria* and *neu/gel*) were independently randomized across subjects. An item from either the first or second contrastive pair was considered the target and presented either with prominence in both speech and gesture (Condition 1) or with prominence in speech only (Condition 2). The item from the other contrastive pair was presented without prominence in either speech or gesture (Condition 3). This led to a total of four possible combinations, with target items counterbalanced according to whether they were the first or second item within each pair and whether they were the first or second pair to appear in the sentence. Thus each of the 48 narratives were video-recorded in the following four versions:

1) Prominence in both speech and gesture (Condition 1) in the first position and prominence in neither speech nor gesture (Condition 3) in the second position.

2) Prominence in speech only (Condition 2) in the first position and prominence in neither speech nor gesture (Condition 3) in the second position.

3) Prominence in neither speech nor gesture (Condition 3) in the first position of the continuation passage, and prominence in both speech and gesture (Condition 1) in the second position.

4) Prominence in neither speech nor gesture (Condition 3) in the first position and prominence in speech only (Condition 2) in the second position of the continuation passage.

Thus a total of 192 video clips (84 narratives × 4 versions) were produced. The narratives were read off a teleprompter by a Catalan native speaker, who had been specially trained to produce the narratives in the abovementioned prosodic and gestural conditions. To facilitate this, target words in Conditions 1 and 2 were highlighted in capital letters on the teleprompter text that the speaker was reading.

The target words in Conditions 1 and 2 (prominence in both speech and gesture, and prominence in speech alone, respectively) were produced with a L+H* pitch accent. In Condition 1, beat gestures produced with a bimanual open-hand movement (see Figure 1, left panel) were timed to coincide with the pitch-accented syllable of the target item. Target words in Condition 3 were produced with a non-focal L* pitch accent. During the recordings of the stimuli special attention was paid to the speaker's control over prosodic prominence in Condition 1 (Prominence in both speech and gesture) and Condition 2 (Prominence in speech only). The speaker was trained to produce the discourse and target items with identical prosody in Conditions 1 and 2. The three panels in Figure 2.1 show sample stimuli of the target word *neu* 'snow' under the three experimental conditions (Condition 1 is presented in the left panel, Condition 2 in the central panel, and Condition 3 in the right panel).

All recordings were carried out in an experimental laboratory at the Universitat Pompeu Fabra with a PMD660 Marantz professional portable digital video recorder, a Rode NTG2 condenser microphone, and a Tecco SP 15 II teleprompter. Videos were later edited with the Adobe Premiere Pro CC video-editing program.

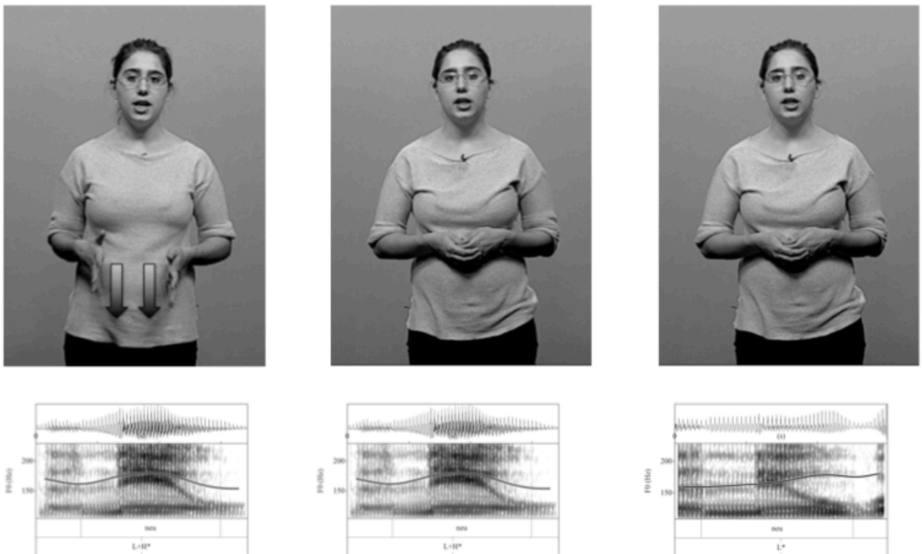


Figure 2.1: Sample video and audio stimuli of the target word *neu* 'snow' in the context passage presented in the three experimental conditions. Condition 1 (both gestural and

After the recordings, we acoustically labeled a sample of 20% of the recordings to ensure that prosodic measures in Condition 1 and Condition 2 did not differ across conditions. Table 2.1. shows the mean duration of the target words and target accented syllables as well as the mean pitch range of the target pitch accent in Conditions 1 and 2. T-tests confirmed that there were no significant differences between the duration and pitch range measures in the target items across Conditions 1 and 2.

Measure	Condition 1 (Prominence in both speech and gesture)		Condition 2 (Prominence in speech only)		F	Significance
	Mean	SD	Mean	SD		
Target word duration (ms)	0.545	0.134	0.614	0.170	1.721	0.195
Accented syllable duration (ms)	0.236	0.124	0.227	0.990	1.185	0.281
Pitch range of the target pitch accent (Hz)	36.721	15.511	39.857	20.699	0.669	0.417

Table 2.1: Mean acoustic measures by accent type for 20% of the experimental materials.

2.2.3. Procedure

Each participant was instructed to watch all of the 48 video-recorded narratives on a laptop computer by him/herself in a quiet classroom and told that he or she would afterwards be tested on their recall of the narratives. The format of the subsequent memory test was not specified prior to the listening session.

During the listening phase, participants listened to all 48 narratives, each story containing two target items. One of the target items was assigned to either Condition 1 or Condition 2, half of them chosen from the first contrastive pair and half of them chosen from the second. The second target item was assigned to Condition 3 Since this was a within-subject experimental design, the order of the narratives was randomized within subjects.

In the test phase, participants performed a memory recognition test. They were given written transcripts of the 48 narratives they had previously heard, in the same order that they had heard them. However, in the written version, the target and unaccented words in the continuation passage were replaced by blanks, as illustrated in (3).

> *(3) Ahir l'Esmeralda havia planejat anar a la peixateria i a la fruiteria. No obstant això, quan va sortir de casa va veure que la tempesta havia deixat molts carrers coberts de neu i gel. L'Esmeralda va aconseguir anar a __________, però com que l'altre lloc estava cobert de __________, va decidir tornar-hi un altre dia.*
>
> Translation*: Yesterday Esmeralda was planning to go to the fish and grocery shops. However, when she left her house, she saw a storm had left many of the streets covered with snow and ice. Esmeralda managed to get to the __________, but since the other shop was covered with ____________ she decided to go back another day.*

Participants were asked to fill in the blanks according to what they remembered hearing. The total time of the listening and testing phases was approximately 30 minutes for each participant.

2.3. Results

Responses by subjects for the recognition memory test were first coded as '0' or '1', with 1 indicating that they had recalled the correct word from the contrastive discourse and '0' that they had not. All such responses obtained (960 trials in total: 20 subjects × 48 narratives) were submitted to a Generalized Linear Mixed Model (GLMM), using IBM SPSS Statistics 23. The dependent variable was Response (1–right reply; 0–wrong reply). The fixed factors were CONDITION (three levels: prominence in both speech and gesture; prominence in speech alone; no prominence non-prominent), ITEMPOSITION (two levels: first position and second position), as well as their interactions. PARTICIPANT, ITEM, and DISCOURSE were set as random factors. The GLMM results revealed a significant main effect of CONDITION ($F(1,956) = 25,136$), $p < .001$). Follow-up paired comparisons showed a significant difference between the three experimental conditions: Condition 1 – Condition 2 ($p < .05$), Condition 1– Condition 3 ($p < .001$), and Condition 2– Condition 3 ($p < .006$), confirming that (a) participants performed better in the recognition task when the critical word was accompanied by prosodic prominence than when the critical word was given neither gestural nor prosodic prominence; and (b) participants performed better when the critical word was accompanied by prominence in both speech and gesture than when the critical word was not accompanied by visual prominence. Interestingly, a main effect of ITEMPOSITION was found ($F(1,956) = 14,625$, $p < .001$), showing that the first position induced a higher recall level than the second position. There was no significant interaction between ITEMPOSITION and CONDITION. Figure 2.2 shows the mean proportion of recalled words across the two item positions (first and second) separated by the three prominence conditions.

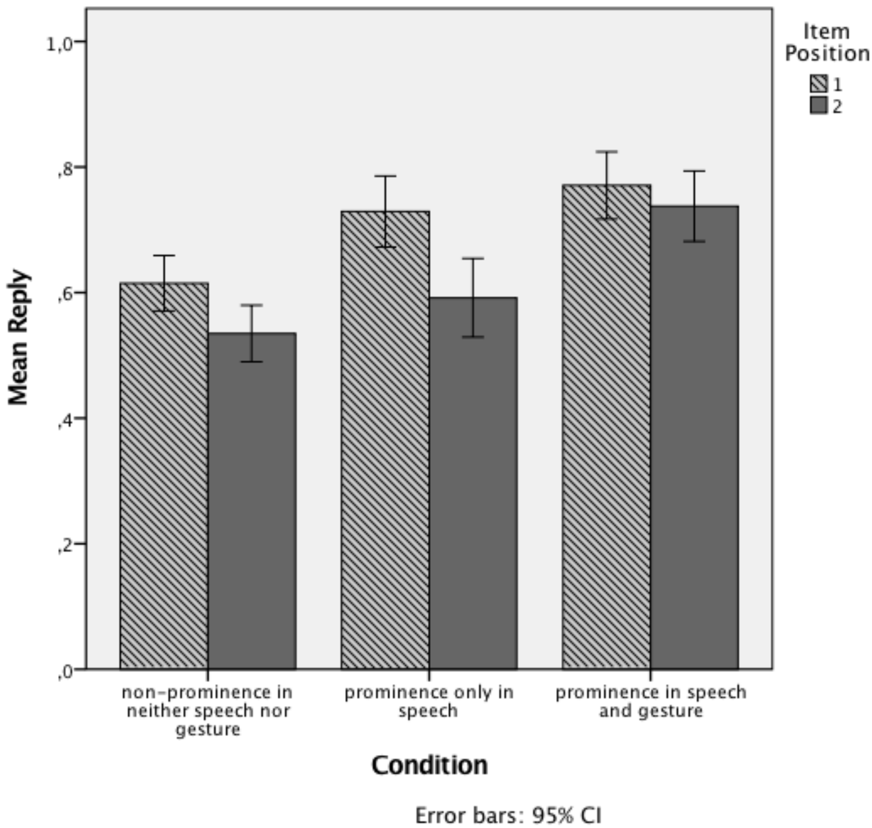


Figure 2.2: Mean proportion of recalled words across the two item positions (first and second) separated by the three prominence conditions: prominence in neither speech nor gesture (Condition 3), prominence in speech only (Condition 2) and prominence in both speech and gesture (Condition 1).

2.4. Discussion and Conclusions

The aim of our study was to investigate the potential positive effects of gestural prominence encoded in beat gestures on information memorization in a contrastive discourse, controlling for the effects of prosodic prominence (i.e., pitch accentuation). We specifically intended to test whether adding a beat gesture to a prosodically accented item would enhance its memorization. While previous research in the field of gesture has centered on representational gestures and their role in information memorization (e.g., Goldin-Meadow, 2003; Quinn-Allen, 1995), little is known about the potential effects of beat gestures. Previous studies investigating the effects of beat gestures on information recall (e.g., So et al., 2012, Igualada et al., 2017, Austin & Sweller, 2014) did not control for the potential effects of prosodic prominence. To our knowledge, this study is the first one to test whether naturally produced beat gestures accompanying prosodic prominence are beneficial in the context of information recall in one's native language as compared to prosodic prominence with no beat gestures. Based on the previous research on the effects of prosodic prominence (e.g., Fraundorf et al., 2010; Bock & Mazzella, 1983), we hypothesized that (a) words marked with prosodic prominence would be recalled better than words accompanied by no prominence cues at all; and (b) if beat gestures were accompanied prosodic prominence cues, this would strengthen the effect of prosodic prominence and induce better mnemonic effects. The results of the study confirmed our two hypotheses.

First, the results of the study clearly revealed that words in focused position presented with prosodic prominence (in the form of pitch accentuation) were recalled better than words presented with no prosodic prominence. These results confirm previous results by Fraundorf et al. (2010) for English. Second, the results also revealed that words produced with both beat gestures and prosodic prominence were recalled better than words produced with prosodic prominence alone or no prominence of any sort. These results are in line with Krahmer and Swerts's (2007) study, which found that adding visual beats to pitch accentuation led to a stronger perception of prominence. In our data, the beat gestures might have led to a stronger perception of prominence, which in turn induced higher recall rates.

Given these findings, one is tempted to speculate whether the presence of any sort of other non-representational movement (such as grooming movements) might have similar effects on memory recall. However, a number of studies have already uncovered neural evidence that ~~of~~ the activation of language-related areas during beat gesture observation is not seen during the observation of nonsense hand movements (e.g., Hubbard, Wilson, Callan & Dapretto, 2009; Biau & Soto-Faraco, 2013; Dimitrova et al., 2016; Wang & Chu, 2013). First, the results by Dimitrova et al. (2016) showed that beat gestures induced a different reaction than other hand movements such as grooming gestures, which did not interact with focus processing. In the fMRI study conducted by Hubbard et al. (2009) thirteen adult subjects were exposed to videos with spontaneously-produced speech accompanied by either beat gestures, nonsense hand movements, or no movements of any kind. The bilateral non-primary auditory cortex showed greater activity when speech was accompanied by beat gestures than when speech was not accompanied by any kind of movement. Additionally, the left superior temporal gyrus/sulcus revealed stronger activity when speech was presented together with beat gestures in comparison to nonsense hand movements. Interestingly, the findings of the study suggested a mutual substrate for processing speech and gesture, the left posterior temporal lobe, which is known to also be sensitive to speech prosody (Humphries, Love, Swinney & Hickok, 2005).

In general, the positive effects of beat gestures found in our data, in comparison with prosodic prominence alone, are consistent with and strengthen the view defended by the embodied cognition paradigm, which underlines the relevance of body movements and multimodal supporting channels in cognition and in favoring memory traces (Barsalou et al., 2003, Barsalou, 2008; see also Paivio, 1990). Thus far, the evidence that co-speech gestures reinforce memory and comprehension processes have come mainly from studies involving representational gestures. In our experiment, the addition to verbal information of gestural information that was not semantically loaded seems to have created memory traces that are even more multimodal, thus increasing their strength. We believe that not only representational gestures but also non-referential beat gestures bear a close relation to cognitive processing. By adding a sensory modality to prosodic prominence, beat gestures fulfill a unique focusing function in highlighting discourse information, which in turn benefits the subsequent recall of this information. Our results support the idea that beat gestures serve as a visual attention-getting strategy in online interaction that helps the listener to direct his or her focus of attention. A recent study by Biau and Soto-Faraco (2013) investigating the time course of beat-speech integration in speech perception showed that beat gestures are integrated with speech early on in time and that they modulate word-evoked potentials in early stages of speech processing. In this study, participants were asked to watch a discourse by a Spanish politician with a wide range of beat gestures, in two conditions, audio-visual and audio only. By measuring participants' ERPs, the researchers found a positive shift in ERPs at an early sensory stage in the audiovisual condition. By contrast, there was no ERP difference when the same words were heard unaccompanied by any visual content. All in all, the results suggest that prosodic and visual prominences can be used together in language to activate the attentional system of the listener, which in turn can lead to more efficient language learning processes.

Kushch,O., Igualada, A., & Prieto, P. (under revision) Prominence in speech and gesture favor second Language novel word learning. *Journal of Language, Cognition and Neuroscience.*

3. CHAPTER 3: “Prominence in speech and gesture favor second language novel word learning”

3.1. Introduction

There is broad consensus in the literature that humans use multimodal cues in their communication. We express information and convey our cognitive processes through both speech and gesture modalities (e.g., manual and facial gestures, body posture). Research over the past decades has shown that speech and manual co-speech gestures are tightly integrated at both the phonological (i.e., temporal) and semantic-pragmatic levels (e.g., Kendon, 1980; Levinson & Holler, 2014; McNeill, 1992, 2005; Poggi, 2007) and form part of a single communicative system (e.g., Bernardis & Gentilucci, 2006; Goldin-Meadow, 2003; Kendon, 2004; McNeill, 1992). Co-speech gestures are produced as part of an intentional communicative act, are constructed within speech and are not a functional act on an object or person (McNeill, 1992).

A good number of studies have demonstrated the positive role of representational gestures (which have also been referred to as *iconic* gestures, as the form of the gesture bears a close relation to the semantic content of speech; McNeill, 1992) in enhancing word memory recall and thus facilitating lexical access in first languages (Cohen, 1981; Cohen & Bean, 1983; Cohen & Stewart, 1982; Nilsson & Craik, 1990; Woodal & Folger, 1985, So et al., 2012).

However, much less is known about the cognitive effects of beat gestures. Beat gestures are a type of rhythmic hand and arm movement that are typically associated with prominent prosodic positions in speech; their function is non-referential and they are generally used in language to signal informational focus (e.g., McNeill, 1992; Shattuck-Hufnagel et al. 2016).

In human communication, the focusing of information is commonly achieved through prosodic prominence. Recent research in audiovisual prosody has shown that prosodic prominence in speech is typically produced simultaneously with prominence expressed with gestural features (e.g., such as head nods, eyebrow movements, hand beat gestures, or exaggerated articulation, see, for example, Prieto et al., 2015; Dohen, 2009; Ekman, 1979; Shattuck-Hufnagel et al., 2016; Swerts & Krahmer, 2008). There is a strong temporal connection between the presence of prosodic prominence (or pitch accentuation) and beat gestures. Typically, the most prominent part of co-speech gestures (the gesture stroke or apex) is temporally aligned with prominent parts of speech (i.e., accented syllables) (e.g. McNeill, 1992; see Esteve-Gibert & Prieto, 2013 for a review). Yasinnik et al. (2004) showed that during a narration more than 90% of instances of the gesture apexes in English occurred together with a pitch-accented syllable (see also Jannedy & Mendoza-Denton, 2005 for a review).

3.1.1. Effects of beat gestures on memory in L1

With respect to the mnemonic effect of beat gestures, there are contradictory results in the literature. First, Feyereisen (2006) argued that beat gestures might not enhance memory recall. In his experiment, he examined the mnemonic effect of producing meaningful gestures (i.e., representational gestures) vs. nonmeaningful gestures (such as beats), and detected no effect of nonmeaningful gestures on memory. However, indefinite gestures (i.e., gestures in which the referent was hard to identify) and beat gestures were both grouped together as “nonmeaningful” gestures and it was therefore impossible to analyze their effects separately. More recent results obtained by So et al. (2012) revealed that while iconic gestures enhanced memory recall in adults and children, beat gestures played a positive role only for adults. However, we must note that in So et al.’s study words accompanied by beat gestures were presented as a sequence of isolated words and not within a discourse context, whereas one of the crucial functions of beat gestures is precisely to highlight the most prominent part of a discourse (e.g., McNeill, 1992; Shattuck-Hufnagel et al., 2016). By contrast, a more recent study by Igualada et al. (2017) carried out with 3- to 5-year-old children showed that children were better at recalling words when they were presented with a beat gesture than when they were not. The results support the idea that beat gestures do help children to recall prominent words presented in a relevant discourse. Similarly, Austin and Sweller (2014) showed that beat gestures can be of help in the recall of spatial directions in 3- to 4-

year-old children. In their experiment, children recalled information about spatial directions better when the spatial directions were accompanied by beat gestures. A recent study by Llanes-Coromina et al. (under revision) has also shown that preschool children remember and comprehend information better in a set of target discourses when it is accompanied by prominence in both speech and gesture.

However, all the empirical evidence suggesting that beat gestures can bootstrap mnemonic processes has so far come from studies where participants were asked to recall information in their native language, and rather less is known about the effects of beat gestures on learning new words in a second language. In the following subsection we summarize the research that has explored the beneficial effects of gestures on second language word memorization.

3.1.2. Effects of gestures on vocabulary learning in L2

Research on the role of gestures in L2 word learning[1] has thus far centered largely on the role of representational gestures rather than beat gestures. A number of studies have shown that items accompanied by meaning-related representational gestures can be learned faster in an L2, as they facilitate word-meaning associations

[1] In this article, the term 'second language learning' is used as a cover term that refers to the process of learning another language after the native or dominant one. This is a common strategy in the field, which uses this term to refer to the learning of a third or a fourth language (Gass, 2013).

(e.g., Kelly et al., 2009; Macedonia et al., 2011; Quinn-Allen, 1995; Tellier, 2008). The study by Macedonia et al. (2011) compared the effect of iconic vs. meaningless gestures produced by instructors on noun recall in students of a foreign language. In this study, 33 German-speaking participants were trained to remember 92 nouns from a corpus of artificial words that was created according to Italian phonotactic rules. In the training session, words were accompanied by either iconic or meaningless gestures. Participants were trained over four days and then performed a recall test on the fifth and sixth days. The results showed significantly better recall of words accompanied by iconic gestures in both short- and long-term memory tests. Tellier (2008) conducted a study to test the impact of iconic gestures on L2 word learning by children. Twenty French-speaking children were presented with 8 common English words, half of them accompanied by iconic gestures and the other half by pictures. When asked to recall the words, the children performed significantly better in the gesture condition. Further support for the beneficial effect of iconic gestures on L2 word learning was presented by Kelly et al. (2009). In their study, 12 Japanese verbs with common everyday meanings were presented to adults with no previous knowledge of Japanese. The words were presented in blocks of three in the following four conditions: (a) speech, (b) speech + congruent gesture (for example, showing the gesture of drinking while saying the verb "drink"), (c) speech + incongruent gesture (showing the gesture of washing one´s face while saying the verb "drink"), and (d) repeated speech. The results showed that the most positive effect on word learning was achieved when the items were presented in the speech + congruent gesture condition.

Another line of research has explored the beneficial use of pitch gestures, or gestures that mimic or represent the melody of speech, on L2 lexical tone discrimination and word learning in a target tonal language. For example, in the study by Morett and Chang (2015), 57 English monolingual participants were asked to learn a total of 20 Mandarin words that were accompanied by either hand gestures illustrating the shape of the tone pitch (pitch gesture), semantic (representational) gestures conveying the meaning of the word, or no gesture. The results showed that pitch gestures helped subjects distinguish between the meanings of Mandarin words that varied only in tone. These findings provide evidence that the visuospatial features of such pitch gestures might be facilitating the discrimination between Mandarin words differing in the lexical tones and thus indirectly enhance L2 word learning. However, in Morett and Chang's study the availability of pitch gestures failed to improve the participants' performance in a pitch discrimination task. By contrast, a recent study by Baills et al. (under revision) has confirmed that observing and producing pitch gestures favors both tone discrimination and lexical word identification and recognition by non-tonal learners of Chinese.

3.1.3. The current study and theoretical implications

To our knowledge, none of the research carried out thus far has specifically focused on the issue of whether beat gestures (gestures that convey prosodic rather than semantic information) can also enhance L2 word learning (but see Hirata et al., 2014 and Gluhareva & Prieto, 2017 for the use of beat gestures to teach L2 pronunciation). Moreover, given that beat gestures are typically associated with prominent prosodic positions, previous experimental studies have not teased apart whether the effects of beat gestures are mainly due to the presence of concomitant prosodic prominence or not. In sum, while iconic gestures encode semantic information that can help learners memorize novel words, it is not clear whether beat gestures, and also their concomitant speech prominence features, can also be of help in the vocabulary learning process.

The present study has the goal of assessing the effects of the presence or absence of visual prominence (i.e., beat gestures) combined with prosodic prominence (i.e., focal pitch accents) on L2 novel vocabulary learning. To this end, we presented participants with novel L2 vocabulary in which the target words were shown in one of the following four multimodal conditions: no prominence either in speech or in gesture; prominence in both speech and gesture; prominence in speech but not in gesture; and prominence in gesture but not in speech. Our hypothesis is that vocabulary presented with redundant visual prominence together with prosodic prominence would be the most beneficial condition, followed by prosodic prominence without visual prominence. We expect the prosodic prominence without visual prominence condition to show positive results as it is comparable to a natural production of speech prominence (e.g., Bock & Mazzella, 1983; Fraundorf et al., 2010). By contrast, visual prominence without the support of prosodic prominence represents quite an unnatural cross-modal combination, and we therefore hypothesized that it would entail no benefit relative to the control condition, where speech lacked either prosodic or gestural prominence marking.

That said, it is worth pointing out that there exist different theories about the effects of multimodal encoding of information on working memory. A number of studies have explained the memory enhancement triggered by gestures in terms of the depth of encoding. For example, Quinn-Allen (1995) states that gestures provide a context for verbal language that results in deep processing and internalization of the verbal information. Tellier (2008) explains the depth of encoding in terms of multimodality and refers to the support of dual coding theory (Paivio, 1971; Baddeley, 1990). According to the dual coding theory, learning processes can be improved when both auditory and visual modalities work together. Baddeley's (1990) model claims that information is better coded using the two modalities because the combination of different modalities leaves more traces in the memory system. The model posite three independent components: 1) the articulatory loop, which is a speech-sound-based storage system of a limited quantity of phonological items; 2) the visuo-spatial sketchpad, which encodes non-verbal visual and spatial information; and 3) the central executive device, which coordinates the two other components and directs attention to incoming stimuli. Thus, according to Baddeley (1990), working memory consists of separate auditory and visual working memories, and consequently the representation of auditory and visual information occurs in independent systems. By this logic, the addition of redundant visual information (gestural prominence) to speech information (regardless of whether the visual information is semantically related with the associated speech or not) should create multimodal memory traces and learning can be improved when the information is presented visually and auditorily. Also, incorporating an additional redundant visual modality (beat gestures) to auditory modality (prosodic prominence, in our case) should also lead to improvement. On the other hand, the independence between the two systems in this model should make it less likely that cross-modal interactions exist in memory. In our case, it is possible that the multimodal benefit of beat gestures might critically depend on whether redundant/matched information is being encoded or not, e.g. shared prosodic prominence information. In sum, a dual coding theory like the one proposed by Baddeley (1990), which crucially maintains a strict separation between modality-specific subsystems, would predict that adding a redundant modality to an existing one (either audio or visual) in a novel words presentation task should lead to their better memorization. By contrast, it would not predict any cross-modal interactions. In our experiment, Baddeley's (1990) model would expect a benefit of adding visual prominence in both conditions namely the one which includes redundant/matched information regarding prominence in gesture and prominence in speech, and the one including non-matched information regarding prominence in gesture and non-prominence in speech.

By contrast, the grounded or embodied cognition theory (Barsalou 2008) would make a different set of predictions in our experiment. A main principle of grounded or embodied cognition theory (Barsalou, 2008) is that cognition is grounded in multiple ways and that it shares mechanisms with not only perception and introspection but also with action. A set of studies within this paradigm highlight the importance of the body in cognition, as bodily states can cause cognitive states and be the result of them (e.g., Barsalou et al., 2003; Lakoff & Johnson, 1980; Smith, 2005 cited in Barsalou, 2008). And gesture is considered an important form of embodiment in language, and it is closely linked to memory (Barsalou, 2008). Recent work on embodied cognition states that language and body movements are supported by the same neural substrates (e.g., Glenberg and Kaschak, 2002; Pulvermüller et al., 2005). The cognitive system utilizes the environment and the body as external informational structures that support internal representations (Barsalou et al., 2003; Niedenthal et al., 2005). There is neurophysiological evidence that self-performing a gesture when learning verbal information forms sensorimotor networks that represent and store the words in both native (Masumoto, 2006) and foreign languages (Macedonia et al., 2011). In addition to this, there is also evidence that not only gesture production, but also gesture observation leads to the formation of motor memories in the primary motor cortex (Stefan et al., 2005), which is considered a likely physiological step in motor learning. Thus, in contrast with dual coding theories, such as Baddeley's (1990) model, embodied cognition theory would suggest that it is the *integration* of these components into one system that leads to memory improvement (see also Quak, London & Talsma, 2015). Thus, embodied cognition theories would predict that the integration of gesture and speech (that is when beat gestures add naturally congruent/matching visuospatial information to prosodic prominence) can favor information coding. Following this theory we could expect that words presented with beat gestures combined with no prosodic prominence (e.g., a crossmodally incongruent interaction) would not have the same beneficial effect as a congruent audiovisual combination (e.g., beat gestures associated with prosodic prominence).

In summary, our hypothesis is that vocabulary presented with redundant visual prominence together with prosodic prominence would be the most beneficial condition, followed by prosodic prominence without visual prominence. We expect prosodic prominence without visual prominence condition to show positive results as it is comparable to natural productions of speech prominence (e.g., Bock & Mazzella, 1983; Fraundorf et al., 2010). Finally, the condition which presents visual prominence without the support of prosodic prominence represents an unnatural crossmodal combination, because in natural life situations beat gestures are produced together with prosodic prominences (Shattuck-Hufnagel et al., 2016), and function in congruent prosodic contexts (Dimitrova et al., 2016). We thus hypothesize that the multisensory matching (and not mismatching) integration between visual and prosodic prominence will be central for optimal working memory processes.

3.2. Methods

The study has a within-participant design with two factors, Prosodic Prominence (No prominence; Prominence) and Gesture Prominence (No prominence; Prominence), combined in a 2 × 2 Latin square design.

3.2.1. Participants

An initial group of 157 students at the Universitat Pompeu Fabra in Barcelona, Catalonia, volunteered to participate in the experiment. Following the administration of a screening questionnaire to determine whether the subjects were Catalan-dominant or not, 61 had to be eliminated, as they were Spanish-language-dominant (e.g., they reported using Spanish for more than 70% of their daily verbal communication needs). Due to the fact that in the experiment both the instructions and the stimuli themselves were in Catalan, we decided to eliminate Spanish dominant participants (who sometimes have little contact with Catalan) in order to be sure that our participants had a homogeneous linguistic profile. The remaining 96 subjects (mean age = 18.77; SD=1.33; range 18-22; 64 females, 32 males) proceeded to take part in the study. They reported using Catalan for on average 75.2% (SD = 8.9) of their daily communication needs and having no previous formal experience with the Russian language. All subjects provided written informed consent giving permission to process their data.

3.2.2. Materials

3.2.2.1. Audio-visual recordings: Discourse Completion Task

In order to find out which types of beat gestures and intonation are used in contexts of second language teaching, 11 Catalan-speakers recruited from the University Pompeu Fabra participated in the Discourse Completion Task including eight discourse situations[2] (Blum-Kulka, House, & Kasper, 1989; Billmyer & Varghese, 2000; Félix-Brasdefer, 2010). They were asked to imagine that they were English teachers and that they were teaching new words to their students. They were also instructed to strongly emphasize the target English word to help the imagined student remember the word in English. No specific instructions were given as to how they were to produce emphasis on the English words or whether they were to produce gestures. An example of a target sentence is as follows: *Finestra es diu "window" en anglès* ('Window is called [target word] in English'). They were recorded with a PMD660 Marantz

[2] The Discourse Completion Task is an inductive method which has been applied for many years in research on pragmatics and sociolinguistics, and also recently on prosody, with good results (e.g., Prieto and Roseano, 2010).

professional portable digital video recorder and a Rode NTG2 condenser microphone as they carried out this production task.

This resulted in a total of 88 sentences produced by our participants (11 participants x 8 discourse situations In each of the 88 recorded tokens, the target word in English was isolated and then analyzed for gestural and prosodic information. From a prosodic point of view, the most common pitch accent was found to be L+H* on the target word (98.20%). From a gestural point of view, only 52.27% of the utterances were accompanied by gestures. Different kinds of gestures were detected in these sentences, for a total of 46 gestures. Of these, 32.60% were iconic gestures, 8.64% were deictic gestures (i.e. pointing gestures), and 27% were beat gestures. As the aim of the study was to elicit beat gestures, we focused on the 12 beat gestures that were detected during the completion of the task. A type of rhythmic hand and arm movement that is typically associated with prominent prosodic positions was considered as a beat gesture (McNeill, 1992). Beat gestures varied in their form and size, for example, some were produced as energetic flicks of a slightly opened palm, and others were produced by raising the hands up from the elbows, etc. The selection of the beat gesture used in the stimuli was based on the most frequent type of beat gesture, the palm-up bimanual gesture, that appeared in 8 out of 12 beat gestures (see Figure 2).

3.2.2.2. Preparation of the stimuli

The stimulus materials consisted of 64 test items prepared on the basis of 16 Russian words (all nouns[3]) conveying common everyday meanings (16 target words x 4 multimodal conditions; see Appendix A). All words had a disyllabic CVCCV structure with stress on the first syllable and complied with the phonotactic constraints of the Catalan language.

For each Russian word we created four 2-second videos corresponding to the four experimental conditions that result from combining the two factors 1) prominence in speech and 2) prominence in gesture: no prominence in either speech or gesture: (Condition1 - baseline); prominence in both speech and gesture (Condition 2); prominence in speech but not in gesture (Condition 3); and prominence in gesture but not in speech (Condition 4). The stimulus videos consisted of an instructor producing the target Russian word in a standard context sentence, e.g., *Bossa es diu “sumka” en rus* ‘Bag is called “sumka” in Russian’.

A Catalan-Russian bilingual speaker was videotaped producing the 16 stimulus sentences in two conditions: 1) no prominence in either speech or gesture, and 2) prominence in both speech and gesture. To create the first of these conditions, the instructor produced the target word with a non-focal L* pitch accent and kept her hands still (see Figure 3.1). To create the second condition, the instructor

---

[3] The decision to work with nouns was primarily due to the fact that this allowed us to have better control over the number of syllables, syllable types, and stress positions within the target Russian words.

produced the target word with a focal L+H* pitch accent and a beat gesture realized with her two hands held with the palms open (see Figure 3.2). Videos were recorded with a PMD660 Marantz professional portable digital video recorder and a Rode NTG2 condenser microphone. During the stimuli preparation we made sure that in all the conditions the speaker kept the same facial expression and we controlled that no additional markers of prosody (for example head nods or eye-brow movements) were present. The Catalan-Russian bilingual instructor was specifically trained for thi

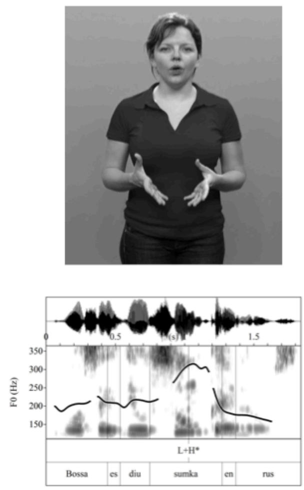


Figure 3.1. Sample stimuli for the word “сумка” /sumka/ “bag” produced with no prominence in either speech or gesture.

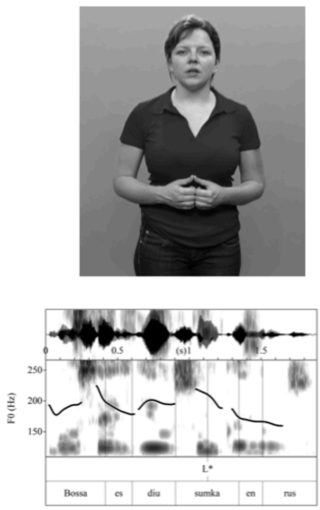


Figure 3.2. Sample stimuli for the word “сумка” /sumka/ “bag” produced with prominence in both speech and gesture.

To generate conditions 3 and 4 (prominence in speech but not gesture, prominence in gesture but not speech), the audio recording of the target word in Condition 1 (no prominence) was replaced by the one in Condition 2 (prominence in speech) and vice versa by using Premiere Pro CS6 software. To check that the resulting audio-visual combinations did not differ from the remaining videos in perceived naturalness (synchrony), we asked 10 native Catalan-speakers to rate the naturalness of the videos from 1 (no synchronization) to 5 (totally synchronized). All videos were rated as highly synchronized (M = 4.69, SD = 0.46). Also, there was no significant difference noticed between the conditions in naturalness rating. The ratings were the following: Condition 1 (M = 4.71, SD = 0.45); Condition 2 (M = 4.65, SD = 0.48); Condition 3 (M = 4.75, SD = 0.46); Condition 4 (M = 4.68, SD = 0.46) ($p > .5$).

3.2.3. Procedure

The experiment consisted of two parts: training and testing sessions. Participants were trained and tested in a single day in four groups of 24.

Training

In the training session, all 96 participants were exposed to words in a within-subjects design (i.e., each participant was exposed to all four conditions). Each subject was exposed to a total of 16 words repeated 4 times ($16 \times 4 = 64$). Each word was assigned to one condition. Importantly, for a given participant this condition remained the same for each word during the 4 repetitions. Across subjects, words were assigned to different conditions to prevent a potential effect of the lexical items, thus creating a total of 4 presentations. For example in Presentation 1 (i.e., participant 1) the word “bag” was assigned to Condition 1, in Presentation 2 (i.e., participant 2) the same word “bag” was assigned to Condition 2, in Presentation 3 the word “bag” was seen under Condition 3, in Presentation 4 the word “bag” was presented in Condition 4, etc. A total of 24 participants were assigned to one of the presentations (e.g., 24 × 4 = 96 participants).

In order to avoid order effects, the following steps were taken. First, the 16 target words were organized in 4 blocks within one repetition. Each block contained one word in each condition (i.e., four words). The order of the words within the block was counterbalanced, and the position of the blocks was also counterbalanced across repetitions. For example, for a given subject the word “bag” (assigned to Condition 1) appeared in position 1 in a given block, in position 2 in the second block, etc., and moreover, these blocks of 4 words appeared in position 1 in the first repetition, in position 2 in the second repetition, etc.

At the beginning of the training session, participants were asked to attentively follow the instructions presented in the first five slides of a PowerPoint presentation, which contained a detailed explanation of the experiment structure. After that, the 16 Russian target words were presented, as follows.

For each target word, participants first saw the written Catalan equivalent on the screen for 3 seconds (e.g., *bossa* 'bag'). They were then shown a 2-second video in which the instructor uttered the carrier sentence with the Russian translation embedded in it (e.g., *Bossa es diu 'sumka' en rus* 'Bag is called 'sumka' in Russian'). After all 16 words were presented in this fashion, there was a 2-minute break during which participants were asked to complete a distraction task.[4] Three further word presentations followed, using the same procedure but with the order of blocks and the order of words within each block changed and the same kind of distraction task filling the 2-minute intervals between presentations.

Altogether, the training session lasted approximately 10 minutes. After the final training phase, there was a 5-minute break, which was followed by the testing session.

Testing

The testing session consisted of two memory tasks. In the first task (a free recall test) participants were presented with the audio recording (with no video input) of the Russian words heard during the training sessions. The words were presented twice and in an order, which was different from the order in the training session. The same audio file was used in training and testing sessions, so the condition of the target word heard in the training session (prominent

---

[4] Participants had to memorize 2 rows of 9 numbers shown briefly on a PowerPoint slide and then write them down on a sheet of paper.

vs. non-prominent accent) was the same as participants heard during the testing sessions. After hearing each word twice the participants wrote the Catalan translation.

In the second task (a recognition test) the participants were auditorily presented with the Russian words heard during the training session (words were presented in a different order from the recall test). Participants heard the audio twice. For each word, participants had to choose between four possible Catalan translations of the Russian word, which included the correct translation and the translations of three other words heard during the training session, and circle the right answer from the four available options. Words used as distractors were from different blocks, but they were presented under the same experimental condition as in the training session.

The total time for training and testing was approximately 20 minutes.

3.3. Results

Responses by subjects for the two tests were first marked ‘0’ or ‘1’, with 1 indicating in the recall task that they had recalled the correct Catalan translation of the Russian word they heard and ‘0’ that they had not, and ‘1’ indicating in the recognition task that they had selected the correct translation from the four alternatives while ‘0’ indicated that they had chosen the wrong translation. All such responses obtained (3,072 trials in total) in both tests were submitted to a Generalized Linear Mixed Model (GLMM), using IBM SPSS Statistics 21. The dependent variable was a combination of Response 1 (in a recall test) and Response 2 (in a recognition test), numerical measures (1-right reply; 0–wrong reply). The fixed factors were Prosodic Prominence (prominence in speech vs. no prominence in speech), Gestural Prominence (prominence in gesture vs. no prominence in gesture), Task (Two levels: free recall test and recognition test), as well as two-way and three-way interactions of those factors (e.g., Prosodic Prominence x Gestural Prominence; Prosodic Prominence x Task, Gestural Prominence x Task, Prosodic Prominence x Gestural Prominence x Task). Participant, Item, and Item Position were set as random factors. Figure 3 shows the results obtained from the GLMM analysis.

The GLMM results revealed a significant main effect of Task ($F(1,3064) = 332,697)$, $p < .001$), confirming that participants performed better in the recognition test than in the free recall test. This is not surprising, since presumably it is easier to select a correct answer from a list of four than independently recall a translation from memory. Importantly, a main effect of Prosodic Prominence was found ($F(1,3064) = 30.487$, $p < .001$), showing that items accompanied by prominence in speech were remembered better. There was no main effect for Gestural Prominence ($F(1,3064) = 0.358)$, $p = .55$), but a significant interaction was found between Prosodic and Gestural Prominence ($F(1,3064) =

4.885), $p < .05$), indicating that only beat gestures produced with prosodic prominence had a significant positive effect on L2 word recall. Post-hoc analyses showed that when the gesture was prominent, scores for the condition with prosodic prominence were higher in comparison to the condition without prosodic prominence ($p < .001$), and when the gesture was not prominent, the condition with prosodic prominence scored higher than the condition without prosodic prominence too ($p < .05$). By contrast, when prosody was prominent the condition with gesture prominence scored higher than the condition without gesture prominence ($p < .05$), but crucially when prosody was not prominent the condition with gesture prominence scored similarly to the condition without gesture prominence ($p = .27$). The fact that there were no significant interactions with task (all ps > .40) indicates that these patterns of results regarding the effects of prosodic prominence or gestural prominence are the same in both the free recall and recognition tasks. Figure 3 shows the mean proportion of memorized words across the four training conditions in the two tasks (free recall and recognition), separated by the prosodic and gestural conditions (prominence vs. no prominence).

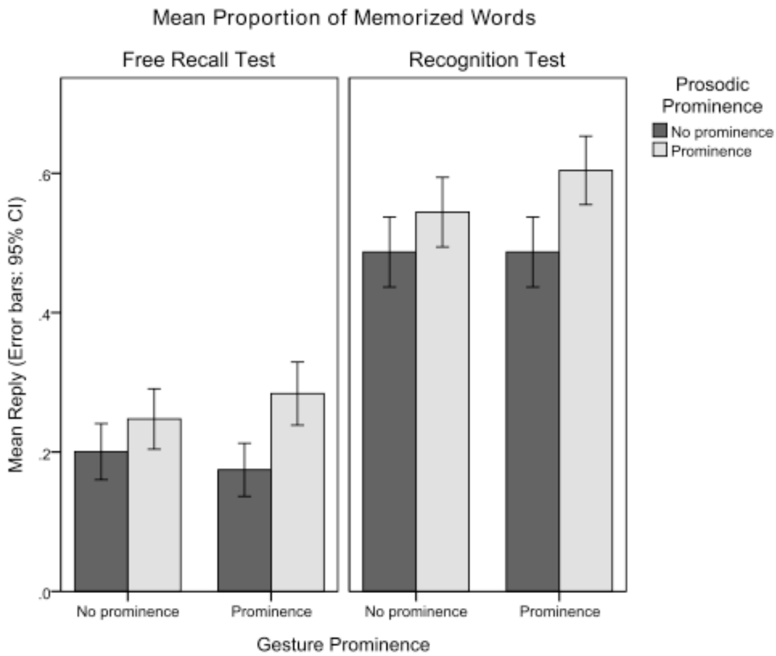


Figure 3.3: Mean proportion of memorized words across the four conditions in the two tasks (Free Recall Test in the left panel and Recognition Test in the right panel).

3.4. Discussion and conclusions

The aim of our study was to investigate the potential positive effects of prosodic and gestural prominence on second language novel word learning. While previous research on lexical learning in an L2 has fully acknowledged the positive role of representational gestures (e.g., Kelly et al., 2009; Macedonia et al., 2011; Tellier, 2008; Quinn-Allen, 1995), little is known about the potential effects of beat gestures. Our study is the first to test whether accompanying the target words with beat gestures (whether accompanied by prominent prosody or not) is beneficial in the context of novel word learning in a second language. Moreover, the experimental design in our study allowed us to independently assess the potential effects of prosodic and gesture prominence on novel word learning.

In this respect, one of the important results that comes out of our experiment is the **asymmetry** between the effects of the combination of prosodic and gestural prominence. First, the presence of prosodic prominence alone had a clear beneficial effect on novel word learning in our data in comparison with the no prominence condition. Participants in the study remembered more words in the focal pitch accent condition (prominence in speech) than in the non-focal condition (no prosodic prominence), both in free recall and recognition tests, demonstrating that prominence in speech helps learners to acquire novel words in a second language. This effect is consistent with previous studies that have reported a positive role of prosodic prominence on information comprehension and memorization in an L1 (e.g., Bock & Mazzella, 1983; Fraundorf et al., 2010; Kushch & Prieto, 2016). By contrast, presenting words with beat gestures but without prosodic prominence did not have beneficial effects on word learning. In the free recall test the mean proportion of successfully memorized words in the no prominence in speech but prominence in gesture condition was even slightly lower than in the baseline no-prominence condition. Importantly, the presence of beat gestures (i.e., gestural prominence) had an (optimal) effect only when it was accompanied by prosodic prominence. Also, when target words were produced with both gestural and prosodic prominence the beneficial effects were strongest in comparison to other conditions.

On the one hand, the optimal effects of the joint association of visual (gestural) and prosodic prominence (as compared with prosodic prominence alone) found in the present study confirm previous research in the L1 field that shows that beat gestures aid word recall in both adults (So et al., 2012) and children (Igualada et al., 2017; Llanes-Coromina et al., under revision). Though the abovementioned studies featured beat productions naturalistically accompanied by prominent prosody, thus not controlling for the possible interaction between the two factors, their results point to the fact that naturally produced beat gestures do have a positive impact on lexical recall. In the last of these studies, (Llanes-Coromina et al., under revision) compared the effects of beat gestures and prosodic prominence on information memorization in contrastive discourse in a first language by adults. Following the design used by Fraundorf et al. (2010), 20 participants were asked to listen to 48 video-recorded stories where the target items were presented in a contrastive focus discourse. Each story contained two pairs of items in the introductory sentence. In the next sentence only one of the items was mentioned. This item was considered the target word and was presented under one of the following two experimental conditions: prominence in both speech and beat gesture, and prominence in speech alone. The results showed that participants performed significantly better in the recall task when the target item was associated with both prosodic and visual gestural (beat gesture) prominence than when it was associated with prosodic prominence alone. All in all, this research provides evidence that not only gestures that contain semantic information (such as representational gestures) but also gestures that help focus the attention on the important part of the discourse (beat gestures) have beneficial effects for word memorization.

An explanation for the enhancing effects of beat gestures is related to the attention processes and also language-related processes triggered by these manual gestures. A neurophysiological study conducted by Hubbard et al. (2009) investigated whether the presence of beat gestures impacted speech perception at the neural level, controlling for the presence of prosodic prominence. Thirteen adult subjects underwent an fMRI while being exposed to videos with spontaneously-produced speech accompanied by either beat gestures, nonsense hand movements, or a still body. Their findings suggested that adding gestural prominence in the form of beat gestures to prosodic prominence (a) causes greater activity in bilateral nonprimary auditory cortex, suggesting a common neural substrate for processing speech and gesture; and (b) causes an increase of activity in the areas responsible for speech intelligibility, namely the left anterior areas of left superior temporal gyrus and sulcus. Thus, Hubbard et al.'s (2009) results suggest that beat gestures, as mentioned above, may help focus the viewer's attention on speech (see also Biau & Soto-Faraco, 2013 and Dimitrova et al., 2016).

Several studies assessing neurological activations during observations of beat gestures support the hypothesis that beat gestures might increase attention processes and activations of language-related brain areas (e.g., Biau & Soto-Faraco, 2013; Holle et al., 2012; see also Hubbard et al. 2009). The functional neuroimaging study by Biau and Soto-Faraco (2015) showed that beat gestures activated different brain areas in comparison to other non-related movements. Depending on whether speech was synchronized with beat gestures or with other non-gestural stimuli (discs/dots moving on a screen) different brain areas were activated. Beat gestures activated language-related areas of the brain, while non-gesture stimuli activated visual perception areas. Hubbard et al. (2009) found that beat gestures, and not nonsense movements or still images, enhanced auditory processing of speech. These studies support the idea that beat gestures can be distinguished from other potential visual highlighters because of their direct integration in the language system. However, if the beneficial effect of beat gestures was exclusively due to attention, we would expect the target items presented with beats be learnt at the expense of others. Interestingly, there is some evidence that points to the contrary. In a recent experiment with preschoolers, Llanes-Coromina et al. (under revision) assessed the memorization of a list of nouns within a child-directed discourse context. While a beneficial effect was seen for the items associated with beat gestures and prosodic prominence, the results also showed that there was no negative effect for the items presented with no beats within the same lists.

These results show that the positive effects of beats are probably not only due to attentional saliency effects.

Our results have clear implications for the models of multimodal cognition and learning reviewed above, namely a classical version of the dual coding theory which maintains a strict separation between modality-specific subsystems (e.g., Baddeley 1990) and the embodied cognition theory which supports the direct integration of the two modalities. Remember that in our results beat gestures without prosodic prominence (a mismatching or incongruent combination) did not have a beneficial effect on memory. As mentioned before, naturally produced beat gestures are almost invariably linked to prosodic prominence in speech (Shattuck-Hufnagel et al., 2016; Yasinnik et al., 2004; see Wagner et al., 2014 and Jannedy & Mendoza-Denton, 2005 for a review). As described in the Introduction, these two conceptions lead to two different predictions regarding the effects of this incongruent audiovisual presentation. While the dual coding theory predicts a positive effect of beat gestures across conditions, regardless of whether they were matched or mismatched with prosodic prominence, this is not the case for embodied cognition theories. Thus, embodied cognition approaches correctly predicted that only matched cross-modal interactions has a reinforcing effect and leaves stronger memory traces. The results of our study thus seem to confirm the predictions based on embodied cognition paradigm, and show that not only gestures that convey semantic meaning have positive effects on learning processes.

However, it is important to also point out at this stage that more recent approaches to working memory (e.g., Baddeley 2000) no longer maintain a strict separation between modality-specific subsystems. Baddeley (2000) proposed an extension to the working memory model presented in Baddeley (1990) by introducing a component called episodic buffer which includes a temporary storage of information that is presented through multimodal code, and which transforms the information into a unitary episodic representation. One of the main focus in Baddeley's new theoretical model is on the process of integrating the modality-specific information, rather than maintaining the isolation of the information in different subsystems. Thus the model proposed by Baddeley (2000) does not sustain the idea of storing the complex images in two separate isolated systems, that are responsible for the maintenance of verbally cued images, but rather suggests the existence of a store that draws information from the slave system and from long-term memory which holds it in the integrated form. Thus more recent versions of the dual coding theory seems to also converge with the grounded or embodied cognition theories. All in all, the results of our experiment point to the relevance of incorporating a multisensory approach to the study of working memory, as recently argued by Quak et al. (2015).

Despite our integration results, two studies seem to suggest that beat gestures and pitch accent affect speech processing independently. First, Krahmer and Swerts (2007) found that beat gestures had a significant effect on perceived prominence independently from prosodic prominence. In this experiment, participants were exposed to the sentence “*Amanda* goes to *Malta*” in which the two target words were associated with prosodic and/or gestural prominence. Beat gestures presented together with focused words increased the perceived prominence of these words and decreased the perceived prominence of other target words. It should be noted that the prominence rating task used by Krahmer and Swerts differs substantially from the word learning task used in the present study. In order to deal with a prominence assignment task, participants might have used a purely perceptual and discrimination ability that differs from that needed for the higher-level task involved in word learning. Still, our results are in line with Krahmer and Swerts’ (2007) study in the sense that adding visual beats to prosodic prominence led to a stronger perception of prosodic prominence, meaning that beat gestures reinforced the effects of prosodic prominence. This interpretation is confirmed by the recent findings reported in Dimitrova et al. (2016). The results of this ERP study showed that words focused by prosodic prominence and words focused by beat gesture were processed more attentively than non-focused words, as they elicited a N1 and P300 component. Importantly, the beat gesture condition and the prosodic prominence interacted in a late time window 600-900 ms relative to target word onset, giving rise to a late positivity when non-focused words were accompanied by beat gestures. The authors attribute these results to the increased computational processing costs of information.

The results of an ERP study by Wang and Chu (2013) seem to highlight the independent role of hand gestures relative to prosodic prominence. In this experiment, participants were asked to watch videos of a person speaking and gesturing. Target videos contained critical words, which were presented under six experimental conditions containing a combination of two factors, namely hand movement and pitch accent (accented and unaccented). The results showed that both beat gestures and pitch accentuation elicited smaller negativities in the N400 time window. These results suggest that prominence in gesture, like prominence in speech, triggers the attentional system separately for semantic processing. Thus Wang and Chu's results seem to contradict the results of our experiment, showing that beat gestures have positive effects only when accompanied by prosodic prominence. Interestingly, aside from differences between the tasks performed in Wang and Chu (2013) and those in the present study, in Wang and Chu's experiments participants could not see the mouth of the speaker. Further research is needed to disentangle this issue, as the authors' decision to hide lip movements might have had a negative effect on the natural integration of prominence coming from speech and both articulatory and hand gestures.

In conclusion, our results are in line with recent work which has suggested that the use of hand movements (e.g., pointing gestures) can substantially benefit cognitive processing and enhance the learning of ideas, whether textual or diagrammatic (Hu et al., 2015). The results of our study have implications for instructional practices in foreign language teaching because they suggest that the teachers' use of prosodic and gestural prominences may help students to acquire novel words in a second language. Vocabulary is a core part of learning a new language. Vocabulary is obviously indispensable for comprehension and producing speech with appropriate meanings as well as generating syntactic, morphological, and phonological structures. In the second language classroom special attention is paid to vocabulary learning and especially to the ways novel vocabulary can be presented and learnt. In the current study we have used a translation procedure to introduce novel vocabulary in a foreign language. The results of the study showed that accompanying target words with prosodic and gestural prominence aids their learning. However, it is necessary to note that learning a novel word is a complex process. Lexical knowledge includes knowing a series of characteristics associated with the word beyond the mere notion of its semantic meaning (e.g., Richards, 1976; Nation, 1990). Thus our study only addresses the initial stage of the vocabulary learning process; the effects of prosodic and gestural prominence on L2 vocabulary learning could also be tested in a longitudinal design and with words embedded in a more natural pragmatic context. Still, the results of our study have direct implications in the foreign language classroom, where it is common to see a fair amount of gesture use by language teachers (e.g., Smotrova & Lantolf, 2013). The use of prosodic and gestural prominence together might constitute a good teaching strategy to cue relevant information in the foreign language classroom and also a promising approach for teaching vocabulary in computer-mediated education environments.

Kushch, O., Gluhareva, D. Borràs-Comes, J, Perez, P.,& Prieto, P. (submitted). The role of beat gesture production in L2 pronunciation training. *Bilingualism: Language and Cognition.*

4. CHAPTER 4: “The role of beat gesture production in L2 pronunciation training”

4.1. Introduction

4.1.1. The role of suprasegmental influence on second language (L2) pronunciation

In the field of second language acquisition (SLA), much has been said about the possible ways to improve learners’ phonological skills in their second language (L2). Pronunciation instruction, when provided, has been typically centered on teaching segmental and phonemic aspects of speech (see Derwing & Munro, 2015 for a review). In the last decades, however, a new line of research has highlighted the importance of suprasegmental instruction for improving learners' overall accentedness and comprehensibility in the second language.

A large body of evidence has demonstrated that suprasegmental deviances have a larger influence on L2 accentedness, comprehensibility and intelligibility ratings than segmental deviances (e.g., Anderson-Hseih, Johnson, & Koehler, 1992; Edmunds, 2010; Field, 2005; Kang, 2010; Ulbrich, 2013, among others). In studies by Munro and Derwing (1995, 1999) and Derwing & Munro (1997), the authors analyzed the relative contribution of segmental and suprasegmental features to native speakers' judgments of accentedness, intelligibility, and comprehensibility. Overall, this work showed that prosodic errors produced by second language speakers affected the native language speakers' ratings of accentedness, intelligibility and comprehensibility to a larger degree than segmental errors. Kang (2010) established a similar pattern, demonstrating that 41% of the variance in accentedness ratings was due to pitch range, word stress, and mean length of pauses, while 35% of variance in comprehensibility was due to speech rate alone. Similarly, White and Mattys (2007) found that rhythm properties had a significant influence on native speakers' ratings of foreign L2 speech. Following this line of research, van Maastricht, Krahmer and Swerts (2015) showed that L1 speakers of Dutch were able to distinguish recordings made by native speakers from those made by non-native speakers based on prosodic cues alone.

From a pedagogical perspective, a number of classroom studies in second language acquisition have demonstrated the importance of teaching suprasegmental components of the language in order to improve the overall fluency and comprehensibility of language learners (e.g., Derwing, Munro and Wiebe, 1998; Derwing & Rossiter, 2003). Derwing et al. (1998) confirmed that, in an 11-week English as a Second Language (ESL) course, learners who received global pronunciation instruction[5] obtained significantly

---

[5] The term global instruction has been widely used in pronunciation teaching studies, and has also been called prosodic, or suprasegmental instruction. However, there is a difference in the studies with respect to

better results in spontaneously-produced speech than learners who were exposed to only segmental pronunciation instruction, or those who received no pronunciation-specific instruction at all. Global pronunciation instruction focused on speaking rate, intonation, rhythm and stress at the word and sentence levels - that is, on suprasegmentals. Segmental pronunciation instruction, on the other hand, was centered on individual phonemes and on discrimination of minimal pairs.

Similarly, Derwing and Rossiter (2003) performed a comparison between three instructional methods: segmental, global, and no specific pronunciation instruction, as a control group. In the experimental groups, participants were exposed to 20 hours of pronunciation training over the course of 12 weeks, after which participants' pre-and post-training recordings were evaluated by five judges who were native English speakers and professional ESL experts. Even though none of the groups showed significant improvements in accentedness, only the global instruction group improved significantly in terms of comprehensibility and fluency.

A more recent study by Gordon, Darcy and Ewert (2013) also confirmed that overall comprehensibility after a three-week training period improved significantly in the group that received suprasegmental instruction (focused on stress, rhythm, linking and

---

the specific features they train. We will therefore specify the types of suprasegmental features that are trained in each study.

reductions). Thirty ESL learners were divided into three groups that differed in the type of explicit[6] instruction received; two experimental groups received explicit instruction either on suprasegmental or segmental features. A combination of these features was presented orally to the third group without explicit instruction, that is implicitly. Pre-test and post-test recordings indicated that only the explicit group trained on suprasegmentals improved their comprehensibility scores significantly from pre-test to post-test.

Behrman (2014) compared the effects of both segmental and prosodic training on reducing nonnative speakers´ foreign accent in American English, providing individual instruction to four native speakers of Hindi. While the segmental training focused on the articulation of consonants, the prosodic training centered on four prosodic utterance levels (rise-fall pitch in one-word utterance, rising, falling, and rise-fall intonation in three-word utterances). This study differs from others in that it used individualized instruction, and all participants underwent both segmental and prosodic training, with accuracy being assessed after each session. The results demonstrated that both segmental and prosodic training resulted in increased accuracy of pronunciation and prosody

[6] Explicit instruction has been defined as a process in which conscious attention is given to linguistic forms, and involves awareness; the brain areas activated are located in the post-frontal cortex. Implicit instruction, in contrast, involves no awareness or consciousness in meaningful uses of language, with different brain areas being affected (see Hulstijn, 2005: 131).

patterns, respectively, and that those improvements appeared to be maintained over the short term.

While the aforementioned studies provide strong evidence of the importance of suprasegmental instruction in L2 pronunciation teaching, more research is needed to strengthen the existing findings. Also, to our knowledge, there is almost no work testing different methods and there is little evidence comparing different suprasegmental training approaches. Moreover, there is no agreement on how to implement pronunciation instruction in English as a second language (ESL) and English as a foreign language (EFL) classrooms. The results of a recent large-scale survey on practices in English pronunciation instruction in Europe, reported by Henderson et al. (2012), confirm that very little emphasis is currently placed on acquiring mastery of suprasegmental elements of the language. Additionally, there is no mention of using gesture as a tool for L2 instruction-- a facet of the topic that will be covered in detail later in this paper.

Interestingly, concerning pedagogical materials for the training of suprasegmentals, it is worth noting that one of the teaching methods proposed for the ESL/EFL classroom was Graham's (1978) Jazz chants, which introduced rhythm instruction together with the use of gestures in the EFL classroom: learners were asked to finger-tap the beats of short, poem-like structures in order to follow the target rhythm. The exercises involved rhythmically repeating words and short phrases in the music, while performing finger-tapping gestures or clapping one's hands. The idea behind this this type of task was to raise learners' awareness of rhythmic expression in natural language, while modeling the rhythm of traditional American jazz. These types of practice exercises were part of a new approach to language teaching which appeared at the turn of the 1970s: the communicative approach to language teaching (CLT). It advocated the primacy of speech over writing in language learning pedagogy, with a focus on meaning and communication (Pérez-Vidal, 2009). Besides these Jazz chants, very little has been done in terms of classroom materials to train suprasegmentals in the EFL classroom. From a research perspective, no empirical assessment seems to be available of the effectiveness of this type of classroom practice, and, most importantly for the current study, of how gestures may be used as a tool for L2 rhythm training.

Apart from this, there is evidence that visual enhancement of the auditory signals leads to pronunciation improvement, as evidenced by the fact that displays showing pitch contours have long been used to teach intonation (Fouz-González, 2015). For example, Motohashi-Saigo and Hardison (2009) and Hew and Ohki (2004) (both cited in Fouz-González, 2015) employed waveforms and pitch contour displays to provide visual feedback on a number of problematic aspects of Japanese pronunciation. The studies confirmed that learners who received visual feedback experienced significantly higher improvements in pronunciation in comparison to learners who received only auditory feedback.

Taking into consideration that visual information enhances learning processes, there have been a number of techniques proposed that include visual input. For example, in regards to segments, proposals have included plotting formant data from learners´ output exemplifying articulatory information on a graph, a glossmeter or ultrasound displays (Brett, 2004, Flegem 1989, Gick, Bernhardt, Bacsfalvi & Wilson, 2008 cited in Fouz-González, 2015), as well as videogame interfaces with simplified versions of spectrographic feedback (Gómez et al., cited in Fouz-González, 2015). Also, talking heads, which are commonly used to illustrate sound articulation through animated mouth movements, have been implemented to enhance learners' perception and production of L2 speech. All these techniques were considered convenient given that visual information complements the auditory modality in the perception of speech (Hardison, 2007).

4.1.2. The role of gestures in second language learning

A large body of research exists showing the potential beneficial effects of using gesture as a tool for second language learning in the ESL/EFL classroom. Gullberg (2006) suggested that gestures should be taken into consideration in second language teaching
and that their effects be measured within experimental research on language acquisition. The author highlights that hand gestures may provide language learners with additional input to aid comprehension and overall acquisition. Several experimental studies have addressed the role of gestures for second language

word learning, both in adult and children populations (e.g., Kelly et al., 2009, Tellier, 2008, Macedonia et al., 2011). For example, Macedonia et al. (2011) compared the effects of iconic vs. grooming gestures for L2 word acquisition. In the study, 33 German speaking participants were exposed to a corpus of 92 artificial nouns created on the basis of the Italian language. In the training session, words were accompanied by either iconic gestures, i.e., gestures that represent a semantic meaning (McNeill, 1992), or grooming gestures (for example, stretching the leg or touching one's hair). The training session lasted for four days. The result showed that participants recalled the target words significantly better if they were accompanied by iconic gestures.

Tellier (2008) achieved similar results while investigating the impact of iconic gestures on L2 word learning by children (mean age 5;5). Twenty French-speaking children were presented with 8 common English words. Half of the words were accompanied by iconic gestures and the other half were accompanied by pictures. In the post-test recall task, participants showed higher recall rates for words that were presented with iconic gestures.

Other studies have approached gesture typology and efficacy in order to determine which types of gestures seem to have a larger impact on learning. A case in point is the study by Kelly et al. (2009), which confirmed that words are better memorized when they are accompanied by congruent iconic gestures. In their study, 12 Japanese verbs with common everyday meaning were presented to English-speaking adults with no previous knowledge of Japanese. The words were presented in blocks of three in the following four conditions: a) speech, b) speech and congruent iconic gesture (for example, showing the gesture of drinking while saying the word “drink”), c) speech and incongruent iconic gesture (showing the gesture of washing one´s face while saying “drink”), and d) repeated speech (i.e, pronouncing the word twice). The results of the study showed that participants remembered more words when they were accompanied by congruent iconic gestures than in the rest of the conditions.

While the abovementioned work confirmed the positive role of iconic gestures on foreign language word learning, less is known about the role of non-referential beat gestures (in other words, hand and arm gestures that are typically associated with prosodically prominent positions in speech). A recent study by Kushch, Igualada and Prieto (accepted) showed that observing beat gestures while learning novel vocabulary improves participants’ later recall. Their experiment was conducted with 96 Catalan-dominant native speakers who were asked to learn 16 Russian words presented in four conditions, which alternated the presence or absence of prosodic prominence combined with the presence or absence of gestural prominence (i.e., condition 1- beat gesture together with prosodic prominence; condition 2 – prosodic prominence but no beat gesture; condition 3- beat gesture but no prosodic prominence; condition 4 – no prominence at all). The results of a recall task and a recognition task conducted after the training session showed that the strongest effect corresponded to words presented with beat gestures together with prosodic prominence. Thus, these results indicate that beat gestures, as highlighters of prosodic prominence, enhance second language vocabulary learning.

4.1.3. Gestures and suprasegmental pronunciation instruction.

Recent studies have assessed the effectiveness of the use of gestures on pronunciation learning, specifically of a set of phonological processes such as tonal and intonation learning, as well as phoneme duration, with somewhat controversial results. Different types of gestures have been examined, namely pitch gestures and metaphoric gestures. However, little is currently known about other types of gestures, such as beat gestures.

On the one hand, one group of studies have explored the beneficial use of the so-called pitch gestures (or metaphoric gestures that mimic or represent the melody of speech) produced by the instructor on the learning of L2 tones and intonation, with positive results (Hannah et al., 2017; Yuan, González-Fuente, Baills, & Prieto, 2017; Morett & Chang, 2015; Baills et al., under revision). First, three studies investigated the effect of pitch gestures on L2 lexical tone discrimination and word learning in a target tonal language. Hannah, Wang, Jongman & Sereno (2016) looked at how pitch gestures, which represent the melody in speech, affected non-native Mandarin tone perception by testing 25 English speakers on tone identification. Participants in the study listened to two monosyllabic words with the four tones under four conditions: congruent (audio and video) videos without gestures, incongruent videos without gestures, congruent videos with pitch gestures and incongruent videos with pitch gestures. They were then immediately asked to decide which type of tone they heard (level, dipping, rising, or falling tone). The results of the study showed that participants exposed to congruent videos with the gesture condition obtained significantly better scores at tone identification than participants who saw congruent videos without any gestures.

Similarly, in the study by Morett and Chang (2015), 57 English monolingual participants were asked to learn a total of 20 Mandarin words that were accompanied by either hand gestures illustrating the shape of the tone pitch (pitch gesture), semantic (representational) gesture conveying the meaning of the word, or no gesture. The results showed that watching the videos with the instructor producing pitch gestures helped subjects distinguish between the meanings of Mandarin words that varied in tone. Together, these findings provide evidence that the visuospatial features of such pitch gestures might be facilitating the discrimination between Mandarin words differing in lexical tone and thus indirectly enhance L2 word learning. However, Morett and Chang (2015) did not find that the availability of pitch gestures improved the participants' performance in a pitch identification task.

Yuan, González-Fuente, Baills, and Prieto (2017) investigated how pitch gesture observation can aid in learning difficult Spanish intonation patterns by sixty-four Mandarin Chinese beginner learners of Spanish. Half of the participants received intonation training without gestures, while the other half received the same training with pitch gestures representing nuclear intonation contours. Results showed that observing pitch gestures during the learning phase improved learners' production outcomes to a significantly greater extent than a training without gestures.

Another group of studies by Kelly and colleagues investigated the role of some types of rhythmic gestures that metaphorically map the duration of vowel sounds in L2 Japanese (Hirata & Kelly, 2010; Kelly, Hirata, Manansala and Huang 2014; Hirata, Kelly, Huang, & Manansala, 2014; Kelly & Lee, 2012; Kelly, Bailey & Hirata, 2017). In general, no effect of these gestures was found on the perception of mora length in Japanese. Hirata and Kelly (2010) investigated the role of co-speech gesture perception in the auditory learning of Japanese vowel length contrasts. In the study, participants were exposed to videos of Japanese speakers producing Japanese short and long vowels with and without hand gestures, which were associated with the rhythm of those vowels. A short vertical chopping movement was used for marking short vowels and a long horizontal sweeping movement was used for marking long vowels. The results indicated no significant difference in performance between participants who were exposed to gestures and those who were not.

Kelly, Hirata, Manansala and Huang (2014) and Hirata, Kelly, Huang, and Manansala (2014) explored whether hand gestures influence auditory learning of an L2 at a segmental level. For this purpose they carried out an experiment in which English speakers were trained to learn Japanese words by either observing or producing gestures that coincided with a syllable or mora, as half of the gestures metaphorically represented the information about syllable structure and half of the gestures represented the information about Japanese mora structure. The main finding of the study was that participants across four conditions (Syllabic gesture Observe, Syllabic gesture Produce, Mora gesture Observe, Mora gestures Produce) performed similarly in the auditory identification and vocabulary test performed after the training sessions. Thus, the authors of the abovementioned studies concluded that there is a limited effect of hand gestures when learning durational contrasts in a second language.

There may be a set of reasons that can explain the discrepancy between the results of the abovementioned studies regarding the effects of pitch gestures and metaphoric gestures. First, as Kelly, Bailey and Hirata (2017) note, while pitch gestures have been found to have a systematic positive effect on learning L2 pitch differences, this is not the case with length/duration gestures that represent the length of a sound. To this end, Kelly, Bailey and Hirata (2017) explored the role that metaphoric gestures play in perceiving foreign language speech sounds that differ according to length and intonation. English-speaking adult participants were exposed to videos with a trainer producing Japanese length contrasts and sentence final intonation distinctions accompanied by congruent metaphoric, incongruent and no gestures. The results showed that for intonation contrasts, congruent metaphoric gestures had a positive effect, as identification was more accurate in comparison to other conditions. For the length contrast identification, however, these results were not carried over, and no clear and consistent pattern emerged. In fact, congruent metaphoric gestures made length contrasts identification more difficult.

We also suspect that the difference in the type of gestures used by Kelly, Hirata and colleagues (Hirata & Kelly, 2010; Hirata et al., 2014, Kelly & Lee, 2012; Kelly et al., 2017) might have had an influence. Even though the authors interpreted their results as suggesting a lower limit of speech-gesture integration, they also pointed out that there could be more effective types of gestures (Kelly et al., 2014). Specifically, the mora gestures used in the studies by Kelly, Hirata and colleagues (e.g., the short vertical chopping movements) may be considered as "non-intuitive" to English speakers and, in fact, act as an incongruent combination of speech and gesture that might impede learning durational information in the second language (see Kelly et al., 2009).

More research is needed to investigate whether another type of rhythmic gesture (for instance, beat gestures integrated with prosodic prominence) may promote pronunciation improvement in a second language. Even though McCafferty (2006) pointed out the strong relationship between beat gestures and emerging second language prosody, there is almost no empirical evidence on the role of these types of gestures in the acquisition of suprasegmental elements. While this kind of gesture appears naturally in spontaneous speech, there is still little evidence on the effects of beat gestures on the acquisition of suprasegmental elements.

To our knowledge, the only study investigating this issue is Gluhareva and Prieto (2017), which showed that beat gestures used to mark the rhythm of speech are an effective aid for achieving more native-like pronunciation in an L2. The study consisted of a brief within-subjects training, during which participants were asked to watch an English instructor producing a set of target sentences in English framed in a discourse situation. Half of the utterances were accompanied by rhythmic beat gestures, while others were produced without gestures. Twenty Catalan participants improved their accentedness significantly on the most difficult trained items when the training utterance was accompanied by rhythmic beat gestures. These results strongly suggest that observing beat gestures favored pronunciation improvement on the most difficult items of the study. However, further work is needed to assess the potential beneficial role of not only observing but also performing beat gestures as a method of suprasegmental training.

4.1.4. Self-performing gestures vs. observing gestures

The embodied cognition paradigm (Bersalou, 2008) underlines that cognitive processes are grounded in multiple ways and share mechanisms with actions. A set of studies inside this paradigm highlight the importance of the body in cognition, as bodily states can cause cognitive states and be the result of them (e.g., Barsalou et al., 2003; Barsalou, 2008). From this perspective, gestures are considered an important form of embodiment in language, closely linked to memory. There is evidence in recent work that language and body movements are supported by the same neural substrates (e.g., Glenberg and Kaschak, 2002; Pulvermüller et al., 2005). The cognitive system uses the body as an external informational structure that supports internal representations (Barsalou et al., 2003; Niedenthal et al., 2005). Embodied cognition has important implications for education, as it highlights the importance of appropriate sensory and motor interactions during learning processes for the efficient development of human cognition (see Kiefer & Trumpp, 2012; Wellsby & Pexman, 2014, for reviews).

Gestures are considered under the angle of embodiment. Neurophysiological studies provide evidence that self-performing a gesture when learning verbal information forms sensorimotor networks that represent and store the words in both native (Masumoto, 2006) and foreign languages (Macedonia et al., 2011). In this respect, a specific teaching approach was actually developed in the 1980s (Total Physical Response, proposed by James Asher) within the changes in language teaching traditions which came along with CLT, as referred to above.

Gesture in teaching should be given specific consideration, as there is a lack of consensus in the research. It seems to be clear, however, that redundant information does not always enhance learning (e.g., Yeo, Ledesma, Nathan, Alibali & Church, 2017). The study by Yeo et al. (2017) investigated the role of teacher-produced gestures representing mathematical information in students' learning from lessons about links between linear equations and corresponding graphs. Eighty-two middle-school students completed a pre-test, viewed a video lesson, and then completed a post-test comparable to the pre-test. The lessons that were viewed varied in whether the teacher referred to the equations in gesture and in whether she referred to the graphs in gesture, yielding four conditions: neither equations nor graphs, equations only, graphs only, and both equations and graphs. The gestures were redundant with speech, meaning that the referents of the gestures were also mentioned in speech (e.g., pointing to "2" while saying "2"). The results showed that students learned less when the teacher referred to the equations in gesture than when she did not. The findings of the study put into consideration the redundancy between gesture and speech, and the possibility of "trade-offs" in attention to visual representations. The results of the study underline the need for a more nuanced view of the role of teachers' gestures in students' comprehension and learning.

In gesture literature, there is substantial evidence indicating that producing gestures is more effective in some contexts than merely observing them, for general learning processes (e.g., Beilock and Goldin-Meadow, 2010; Goldin-Meadow, 2014; Goldin-Meadow, Cook & Mitchell, 2009). Beilock and Goldin-Meadow (2010) carried out two experiments that involved solving and explaining the Tower of Hanoi task (TOH) with gestures. Gesturing during the task had beneficial effects on later speech performance; that is, gestures helped participants to change thought, adding action information to their mental representations of the task. Results of the study support the fact that producing gestures can have an effect on changing participants' mental representations and contribute positively to task solving. Likewise, a recent study by Baills, Suárez-González, González-Fuente, & Prieto (under revision) has confirmed that both observing and producing pitch gestures favored both tone discrimination and lexical word identification and recognition by non-tonal learners of Chinese. In the study, Catalan native participants with no previous knowledge of Chinese were asked to observe (Experiment 1) and produce (Experiment 2) pitch gestures during a short multimodal training session on Chinese tones and words. Participants were tested on tone identification and word learning after the training sessions. The results of the study showed positive effects of pitch gesture observation and pitch gesture production in comparison to the no-gesture condition.

4.1.5. Goals of the study

Previous research on the effects of using gestures on L2 pronunciation has revealed contradictory results, and very little research has addressed, on the one hand, the potential beneficial role of beat gestures on pronunciation improvement in a second language, and, on the other, whether methodologies involving production rather than observation practices are more effective for pronunciation learning.

The goal of the present research study is to investigate whether participants show higher gains in accent improvement if they are instructed to observe beat gestures and subsequently imitate the experimenter by producing beat gestures themselves, in comparison to only observing beat gestures. First, following Gluhareva and Prieto (2017), we believe that the presence of visible and natural beat gestures working together with prosody can provide a further benefit for boosting rhythmic information for L2 speakers of English and thus for favoring pronunciation learning. Second, following embodied cognition perspectives, we hypothesize that producing gestures in comparison to only observing them will effectively be more beneficial for pronunciation improvement.

4.2. Methods

The study consisted of a between-subject training paradigm with a pre-test and post-test design. That is, before and after training, the participants' speech production was recorded and perceptually evaluated by native speakers for degree of accentedness. We adopted the definition of accentedness used by Munro and Derwing (1998), "the extent to which an L2 learner's speech is perceived to differ from native speaker (NS) norms" (160) and Flege (1995), according to whom "listeners hear foreign accents when they detect divergences from English phonetic norms along a wide range of segmental and suprasegmental (i.e., prosodic) dimensions" (233). The primary focus of the present study is suprasegmentals.

4.2.1. Participants

Eighteen native speakers of Catalan (4 male and 14 female) (mean age = 21.5 years, SD = 3.327) from Universitat Pompeu Fabra participated in the study. All participants were first-year students in Translation and Interpreting and Applied Languages. Participants were asked to complete a language questionnaire and to provide written informed consent for their data to be processed. All subjects reported having an upper-intermediate level of English, more specifically a B2[7]. Participants also reported using Catalan for an average of 75.7% (SD = 8.5) of their daily communication needs. Participants were randomly assigned to two groups (9 participants in each group): Group 1 – Beat Observation group; Group 2- Beat Production group. Participants received 5 euros as remuneration for

---

[7] Students in the Translation and Interpreting and Applied Languages degrees at the Universitat Pompeu Fabra are required to have at least a B2 level of English (according to the Common European Framework of Reference for Languages) prior to beginning the program.

their participation in the experiment, which lasted approximately 30 minutes.

4.2.2. Materials

This section describes the materials used for the pre-test and post-test phases and during the pronunciation training, namely, a discourse prompt situation and a training video.

(a) *Discourse prompt situations*. 10 discourse prompt situations were used for the pre- and post-training assessments, adapted from Gluhareva and Prieto (2017). For the post-training assessment, 10 new items were added. Each of the prompts consisted of an image of an everyday situation which the participants might face while living abroad in an English-speaking country, and a short set of instructions describing the situation which clarify the speech act they are expected to perform. For example, in one of the items, the participants were shown an image of a group of tourists visiting New York City, trying to find their way with the help of a map. The image was accompanied by the following instruction: “You are trying to find Central Park. You ask a local person for directions”.

(b) *Training videos*. For the training phase, each of the prompts used in the pre- and post-test was accompanied by a training video with a native speaker of American English giving the responses to the situations in the prompts while using beat gestures. For this study we adopted the corpus of training videos used in Gluhareva and Prieto (2017), as presented in Appendix B. In all videos, the instructor accompanied her speech with rhythmic beat gestures which marked the relevant prosodic components of the utterance. The beat gestures consisted of simple open palm hand movements (see Figure 1). In all 10 beat training videos, all of the nuclear pitch accents received full beat gestures, while some non-nuclear stressed syllables were marked with less forceful beat gestures. However, not all stressed syllables were accompanied by beat gestures, because as highlighted by Gluhareva and Prieto (2017), it would have appeared unnatural; thus, the instructor only placed beat gestures on the words with the heaviest semantic weight (see Appendix C for the transcript of the videos with beat gestures).

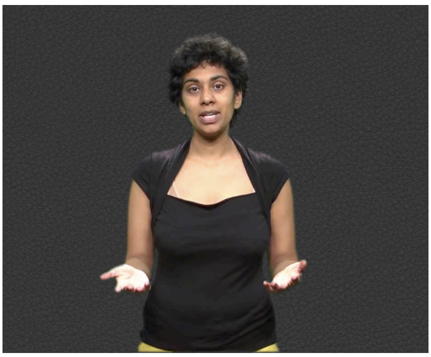


Figure 1. Still image from the training videos in the two conditions illustrating an open-palm beat gesture.

4.2.3. Experimental setup and procedure

The participants were tested and trained individually using a laptop computer at the Language Laboratory Lab at Universitat Pompeu Fabra.

First, participants were randomly assigned to one of the two groups (Beat Observation and Beat Production), as the goal of the study was to see whether producing beat gestures during the pronunciation training could result in more substantial pronunciation improvement in comparison to only observing beat gestures. The experiment involved three phases, namely, a pre-test phase, a training phase, and a post-test phase (see Figure 4.2 for a schema of the experimental procedure).

Phase 1 - Pre-test: Participants were asked to record themselves speaking, alone in a quiet classroom. First the participants were shown an instruction slide with each situation, in which they are told what speech they have to perform (ask for directions, introduce themselves, ask for the time, etc.). In each of the images, a blank speech bubble would appear in order to mark that it was their turn to produce their speech act. In order to elicit natural speech and to avoid having the participants read off the screen while producing the responses, a black screen was introduced. The participants' response to each situation was audio recorded.

Phase 2 - Training: Following the pre-test, the training began. Participants were asked to join their previously-assigned groups and each group was gathered in a different computer room. Participants were asked to sit in front of a computer so that they could follow the instructions and go through the training. The Beat Observation group received instructions to watch the videos with an instructor and repeat the sentences after the instructor without imitating the gestures. The Beat Production group received instructions to watch the videos with an instructor and repeat the speech, but additionally produce the gestures seen in the videos, as follows: "Repeat after the instructor. Accompany your speech with gestures, similar to the ones that the instructor produces".

For both groups of participants, the training began with a habituation phase, during which participants received a short explanation of the structure of the experiment. The habituation phase was performed with the experimenter present in the room. The experimenter presented the participants the first two extra familiarization situations, in order to make sure that all participants understood the task and more specifically that they knew how they were to use gesture/how gesture was used. These two extra prompts were not recorded and were not taken into account when processing the results. After verifying that the participants understood the instructions for the task, the experimenter left the room only for as long as the training phase lasted (around 15 minutes) in order to allow participants to feel totally free to use their hands while repeating. When the participants were left alone, they were video recorded in order to further control if the training had been completed correctly. The training was video-recorded with a Nikon d7000 camera, which was facing the participants and was located two meters away from them.

Phase 3 - Post-test: Following the training phase, the participants were given a 5-minute break, during which they completed the consent form. They then proceeded with a post-test, which included the same 10 items as the pre-test, plus 10 new items. During the training phase of the study, participants were not aware of the task that they were going to perform in the post-test. Moreover, including a set of 10 new items in the post-test allowed us to assess their general improvement after the training session. The responses were audio-recorded.

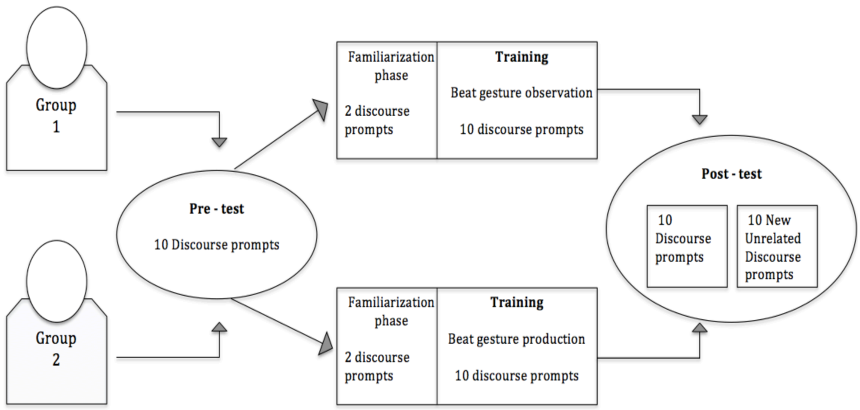


Figure 4.2. Overall experimental procedure.

Figure 4.3 displays still images from the training phase of the experiment: participants in the Beat Observation training group are shown at the top, while the bottom panel shows participants who were assigned to the Beat Production training group.

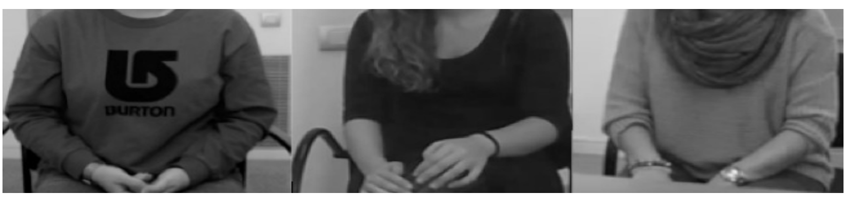


Figure 4.3. Still images from the training phase videos from Beat Observation group (top panel) and from the Beat Production group (bottom panel).

A total of 18 video recordings from the training session were obtained (18 participants). In order to check (a) that the participants had correctly followed the instructions on the Beat Observe and the Beat Produce conditions, and (b) that the beats produced by the participants in the Beat produce condition were adequately and naturally performed with respect to form and rhythmic pattern, the video recordings were reviewed by an independent rater who evaluated the 9 participants in the gesture production group, focusing on how natural they found their gesture performance using a Likert scale (1-badly done, 5-well done). The results were positive for all the participants and thus they were all included in the analysis (M = 3.73, SD = 0.83).

Ratings

The participants' recordings from both the pre-test, post-test and test for new items were rated by five native speakers of American English, one male and four females (Mean age = 26 SD = 2.3). The raters resided in Barcelona, Catalonia. At the time of the assessment, all raters reported having normal hearing. Each rater evaluated a total of 540 participants' recordings (18 participants × 30 items coming from 3 tests: pre test, post-test, 10 new items in the post test). All ratings were performed via a five-part online survey consisting of 168 audio clips. The raters reported that each part of the survey took them approximately 60 minutes to complete, for a total of 5 hours. The stimuli were presented to the native speakers in a random order that included the recordings from the pre- and post-tests.

Following Gluhareva and Prieto (2017), prior to performing the ratings, the raters received instructions on how to evaluate the recordings, based on the speaker's degree of foreign accent, instead of the content conveyed or grammar used in their utterance. Each page of the survey presented the raters with one recording, which they were asked to evaluate on a 9-point accentedness scale, from "1" (native/ no accent) to "9" (very strong foreign accent). The instructions were the following:

> *"The task*
>
> *Your task is to listen carefully to short audio recordings. You will be asked to indicate how native-like (in other words, like it was produced by a native speaker of English) each recording sounds. Then, you will be asked to indicate how "native" each recording sounds, on a scale of 1 to 9.*
>
> *IMPORTANT: When evaluating the recordings, please do not concentrate on their content or length. Instead, focus on the overall pronunciation of each clip."*

Following Gluhareva and Prieto (2017), accentedness was chosen as the target measurement because, as highlighted by van Maastricht, Krahmer & Swerts (2015), while native listeners tend to be very sensitive to nonnative-accented speech, ratings of intelligibility and comprehensibility tend to be less as extreme (in other words, heavily accented speech may still be rated as relatively intelligible.

Additionally, a comprehensive measurement such as accentedness, focusing on suprasegmentals as stated above, was chosen in lieu of asking the raters to evaluate more specific characteristics of the speech samples (e.g. stress patterns, intonation, etc.) because we aimed to assess the participants' pronunciation based on the global impression that it produces in native judges.

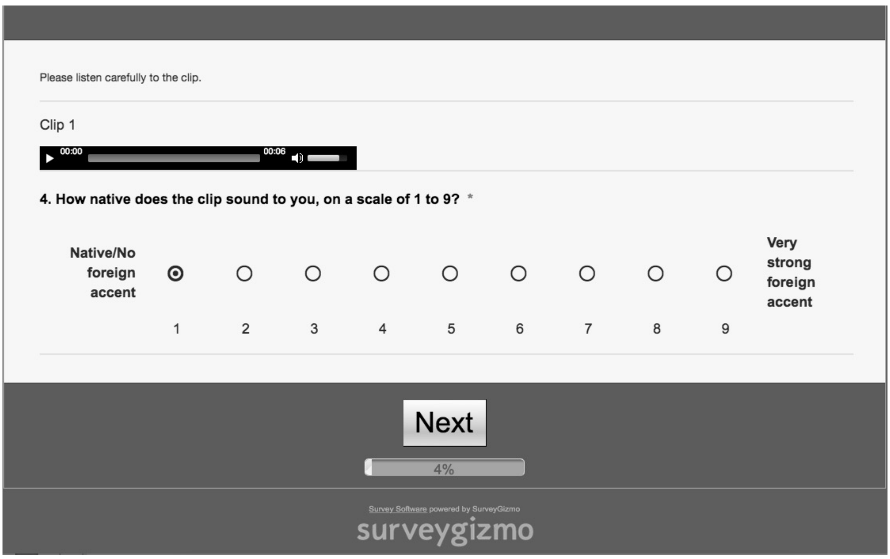


Figure 4.4: Sample page from the online rating survey.

Inter-rater reliability.

Inter-rater reliability was assessed using an intra-class correlation (ICC) analysis for each pre- and post-training test item, and then obtaining an aggregate mean of the results. This yielded a Cronbach's Alpha score of .73, which surpasses the generally-accepted measure of .7 (Larson-Hall, 2010). Therefore, all of the raters' scores were combined to produce a mean rating for each recording.

4.3. Results

A total of 2,700 tokens (5 raters × 3 tests × 10 situational prompts × 18 participants) were submitted to a Generalized Linear Mixed Model (GLMM) using IBM SPSS Statistics 24. The dependent variable was Rating (continuous). The fixed factors were CONDITION (two levels, according to the fact whether participants were asked to repeat the gestures after the instructor or not: 1 - Beat Production condition; 0 - Beat Observation condition), TEST (three levels: pre-test, post-test, new unrelated items), as well as their interaction. A random intercept was set for Rater, with a random slope defined both for Participant and Situation.

The GLMM revealed significant results for the three fixed effects. The main effect of CONDITION ($F(1, 2694) = 7.935$, $p = .005$) indicates that the beat production condition received lower ratings than the beat observation condition ($\beta = .119$, $p = .005$). The main effect of TEST ($F(2, 2694) = 12.661$, $p < .001$) indicates that perceived accentedness was lower in the post-test compared with both the pre-test ($\beta = .523$, $p < .001$) and the new items test ($\beta =$

.480, $p < .001$), whereas no statistical difference was found between the pre-test and the new items test ($\beta = .043$, $p = .263$).

Finally, the interaction of CONDITION × TEST ($F(2, 2694) = 6.675$, $p = .001$) can be interpreted in two different ways. First, in the sense that there was a significant effect of CONDITION, such that the beat production condition received lower accentedness ratings than the beat observation condition, in the post-test ($\beta = .204$, $p = .001$), but not in the pre-test ($\beta = .104$, $p = .106$) nor in the new items test ($\beta = .049$, $p = .355$). Second, even though the effects of CONDITION were similar in each TEST, there was a size effect difference. The accentedness difference between pre-test and post-test was favorable to the beat production group. The accentedness distance between pre-test and post-test was stronger in the beat production condition ($\beta = .573$, $p < .001$) than in the beat observation condition ($\beta = .473$, $p = .003$); and similarly, the difference between the post-test and the unrelated items test was larger for the beat production condition ($\beta = .558$, $p < .001$) than for the beat observation condition ($\beta = .402$, $p = .003$); meanwhile, no significant differences were found between the pre-test and the new items test either in the gesture condition ($\beta = .016$, $p = .681$) or in the no gesture condition ($\beta = .071$, $p = .241$). Figure 4.5 shows the mean accentedness ratings in the three tests (pre-test, post-test and test with new items) as a function of the beat production and beat observation conditions.

On the whole, the results confirm that training with beat gesture production significantly improved participants' pronunciation in comparison to beat gesture observation.

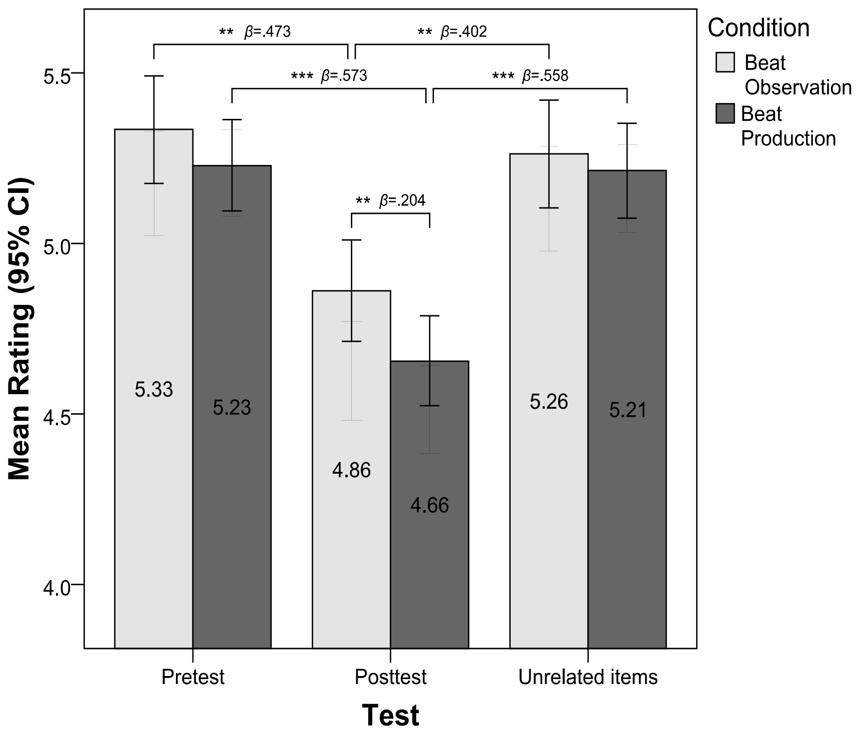


Figure 4.5: Mean accentedness rating scores obtained from the three tests (pre-test, post-test, and test with new items) as a function of the beat production and beat observation conditions. Note that lower scores indicate less accented speech.

4.4. Discussion and conclusions

The present study explored the effects on L2 English pronunciation improvement, as measured through degree of foreign accent, of a

15-minute pronunciation training based on beat gesture production. In a between-subject experiment using a pre- and post-test design, Catalan second language learners of English were exposed to training videos with a female instructor producing short discourse utterances and accompanying her speech with rhythmic beat gestures. Participants in the experiment were randomly assigned to two groups. One group was asked to repeat the utterances after the instructor and accompany their speech with beat gestures. The other group of participants was asked to only repeat the speech. The results showed that producing beat gestures during the training session was beneficial for pronunciation improvement, since the accentedness ratings given by native speaker raters were significantly lower in the post-test for participants who produced gestures during training.

The main goal of our study was to compare the effects of observing vs. producing beat gestures on pronunciation improvement. The results demonstrate that producing beat gestures leads to higher gains in accent improvement in comparison to only observing beat gestures, in the case of the participants tested (see also Gluhareva and Prieto 2017). While both training methods (beat gestures perception and beat gesture perception) had beneficial effects, training with beat gesture production resulted in significantly better outcomes in comparison to only observing beat gestures. The beneficial role reported for beat gestures for pronunciation improvement is in line with the previous literature stating that beat gestures are strongly associated with prosodic prominence in one's first language-- Krahmer & Swerts (2007) showed that beat gestures enhance prominence perception when witnessed by L2 speakers. Thus, we believe that both observing and producing beat gestures together serve as an additional source of linguistic information for L2 speakers and consequently facilitate their language production. It is also important to emphasize that the beneficial effect of producing beat gestures occurred after a short 15-minute training. Most of the studies on pronunciation (e.g., Munro and Derwing, 1995, 1999; Derwing & Munro, 1997; Gordon et al., 2013, among others) have utilized more lengthy training designs and included multiple training sessions (but see Gluhareva & Prieto, 2017).

The results of the present study complement previous studies related to the role of gestures for pronunciation instruction. First, beat gestures seem to behave like pitch gestures, which have been consistently shown to have positive effects on learning tones (Morett & Chang, 2015; Yuan et al., 2017; Baills et al., under revision). By contrast, duration gestures, such as the ones used by Hirata & Kelly (2010), Hirata et al., (2014), Kelly & Lee (2012), and Kelly et al., (2017), did not indicate a beneficial effect of beat gestures on learning durational contrasts in a second language. A possible explanation for this may be that while Hirata and Kelly assessed segmental-level improvements, the present study centered on a different level of language learning, namely, suprasegmental (rhythmic) improvement. Also, beat gestures used in the studies are different from the gestures used in our research. In the studies by Hirata and colleagues, short vertical chopping movements were used. These gestures may act as an incongruent combination of speech and gesture that may impede learning durational information in the second language (see Kelly et al., 2009).

Results from the previous studies had suggested that production of gestures by learners themselves is more effective than merely observing gestures (Goldin-Meadow, 2014; Goldin-Meadow et al., 2009; Masumoto, 2006; Macedonia et al., 2011; Saltz & Donnenwerth-Nolan, 1981). Our research is novel, as to our knowledge, no previous research has investigated the potential beneficial effect of producing beat gestures for pronunciation instruction. The results of the experiment showed that participants who were asked to produce beat gestures together with speech during the training session showed a significantly beneficial effect on their ability to produce English speech in comparison to participants who were instructed to produce only speech during the training. All in all, the findings of our study are in line with research that demonstrates that producing gestures facilitates learning mental tasks more than simply observing them (see, among others, Goldin-Meadow et al., 2012).

The findings of our study give additional support for the importance of an explicit, global approach to second language pronunciation instruction-- more specifically, an approach that targets suprasegmental elements of the language and incorporates gesture use. The innovative method used in the present study provides support for the use of materials that include beat gesture production imitation practices in the L2 classroom.

4.5. Limitations

The present study involved a very short-term pronunciation training design and some limitations need to be acknowledged. Due to the fact that the participants' post-training recordings were taken only 5 minutes after they were shown the training video, no long-term effects of the training were assessed. The training paradigm that we applied does not allow us to evaluate the extent to which the participants will retain the benefits of the training over longer period of time. Also, the significant beneficial effect of producing beat gestures was not transferred to new unrelated items that were included in the post-test, probably due in part to the abovementioned factors.

It is also important to recognize that the present training paradigm involved a relatively homogenous group of participants in terms of age, language background, and level of English; hence, more investigation is needed to assess how the use of beat gestures may affect different groups of learners.

5. GENERAL DISCUSSION AND CONCLUSIONS

5.1. Summary of findings.

The general aim of this thesis was to investigate the effects of beat gestures, e.g., non-referential hand gestures that are typically associated with prosodically prominent positions in speech, on learning processes in a first and in a second language. We adapted different methodological approaches to assess how beat gestures benefit (a) information memorization in one´s native language, (b) vocabulary learning in a second language and (c) accent improvement in a second language. Three independent studies were carried out, each one presented in a separate chapter.. The first two studies (Chapters 2 and 3) focused on the role of beat gestures on word memorization in first and in second language learning[8] respectively. The third study investigated whether producing beat gestures has an additional beneficial effect on pronunciation improvement in a second language (Chapter 4).

The results of the three studies revealed that beat gestures are beneficial for word memorization in the L1 and word learning in the L2. In the first study (Chapter 2) participants memorized novel words better in focused position presented with a stronger prosodic

[8] In this PhD thesis, the term 'second language learning' is used as a cover term that refers to the process of learning another language after the native or dominant one. This is a common strategy in the field, which uses this term to refer to the learning of a third or a fourth language (Gass, 2013).

prominence (in the form of pitch accentuation) than words presented with no prosodic prominence. It was also shown that the beat gestures produced together with prosodic prominence on critical words significantly enhance information recall of these target words in comparison to producing these words with prosodic prominence alone (and without gestures) or with non-prominent prosody.

The results of the second study (Chapter 3) showed that beat gestures are beneficial for L2 word learning. Specifically, the presence of prosodic prominence alone had a clear beneficial effect on novel word learning in our data in comparison with no prominence. Participants in the study remembered more words in the focal pitch accent condition (prominence in speech) than in the non-focal condition (no prosodic prominence), both in free recall and recognition tests conducted after the training. Importantly, the presence of beat gestures (i.e., gestural prominence) had an (optimal) effect only when it was accompanied by prosodic prominence. Also, when target words were produced with both gestural and prosodic prominence the beneficial effects were stronger in comparison to other conditions.

The results of the third study (Chapter 4) showed that training coupled with producing beat gestures improves foreign language pronunciation. More specifically, producing beat gestures in a short pronunciation training session was shown to generate additional beneficial effects in comparison to only observing beat gestures.

The findings of the third study give additional support for the importance of an explicit, global approach to second language pronunciation instruction, which incorporates gesture use. Specifically, the innovative method used in the study provides support for the use of natural materials that include beat gesture production imitation practices in the L2 classroom

In the next sections the previously mentioned findings are discussed against those in the previous literature. It is argued that they contribute to the existing body of research specifically with regard to the following issues, namely (a) how beat gestures and prosodic prominence influence information memorization in the L1 and L2 (section 5.2 and 5.3), (b) how beat gestures lead to pronunciation improvement in the L2 (section 5.4). Finally, in light of the embodied cognition paradigm and multisensory processing theories, we will also discuss the potential cognitive effects of beat gestures in comparison to other movements (section 5.5).

5.2. Beat gestures add positive effects to prosodic prominence in L1 and L2 recall of information

One of the aims of this thesis was to investigate the potential positive effects of beat gestures on word memorization in an L1 (Study 1) and word learning in an L2 (Study 2). While previous research on lexical learning in an L2 has fully acknowledged the positive role of representational gestures (e.g., Kelly et al., 2009;

Macedonia et al., 2011; Tellier, 2008; Quinn-Allen, 1995), less is known about the potential effects of non-referential beat gestures. Generally, beat gestures are tightly aligned to prosodic prominence (e.g., McNeill, 2016, Shattuck-Hufnagel et al., 2016). The aim of the first two studies in this PhD dissertation was specifically designed to disentangle the potential beneficial effects of prosody and gestural prominence. The aim of Study 1 was to investigate the potential positive effects of gestural prominence encoded in beat gestures, on information memorization in a contrastive discourse, controlling for the effects of prosodic prominence (e.g., pitch accentuation). We specifically intended to test whether adding a beat gesture to a prosodically accented item would enhance memorization of the target word. The aim of Study 2 was to test whether the effects of beat gestures (whether accompanied by prominent prosody or not) are beneficial in the context of novel word learning in a second language. The experimental design in this study allowed to independently assess the potential effects of prosodic and gesture prominence on novel word learning. The results of these two studies confirmed that beat gestures add a positive effect to prosodic prominence when memorizing information in first and in second languages.

Previous research in the field of gesture has mostly centered on representational gestures and their role in information memorization, both in first and in second languages (e.g., Goldin-Meadow, 2003; Quinn-Allen, 1995; Austin and Sweller, 2014); however, little is known about the potential effects of beat gestures.

Also, previous studies investigating the effects of beat gestures on information recall (e.g., So et al., 2012; Igualada et al., 2017; Austin and Sweller, 2014) did not control for the potential effects of prosodic prominence. We claim that the novelty of our research lies in the fact that it is the first study to test whether naturally produced beat gestures accompanying prosodic prominence are beneficial in the context of information recall in one's native language as compared to prosodic prominence with no beat gestures. The results showed that naturally occurring beat gestures produced together with prosodic prominence strengthen the effect of prosodic prominence and induce better mnemonic effects.

First, the results of Study 1 reveal that L1 novel words in focused positions presented with stronger prosodic prominence (in the form of pitch accentuation) are recalled better than words presented with no prosodic prominence. These results are consistent with previous studies that have reported a positive role of prosodic prominence on information comprehension and memorization in an L1 (e.g., Bock & Mazzella, 1983; Fraundorf et al., 2010). Second, the results also showed that adding beat gestures together with prosodic prominence over critical words significantly improves information recall of these target words in comparison to prosodic prominence alone (and with no gestures) or with non-prominent prosody. Study 2 showed that the same result with beat gestures obtains with word learning in a second language. These results are in line with Krahmer and Swerts's (2007) study which found that adding visual beats to pitch accentuation leads to a stronger perception of prominence. Interestingly, the results of study 2 also confirmed that presenting words with beat gestures but no prosodic prominence did not have beneficial effects on word learning. We explain this result by the fact that naturally produced beat gestures are almost invariably linked to prosodic prominence in speech (Shattuck-Hufnagel et al., 2016; Yasinnik et al., 2004; see Wagner et al., 2014 and Jannedy & Mendoza-Denton, 2005 for a review).

The findings in these two studies (Study 1 and Study 2) reveal that the beat gesture condition leads to a stronger perception of prominence, which in turn induces higher recall rates. Together with Biau & Soto-Faraco (2013) and Holle et al. (2012) we consider beat gestures as attention-getters that serve to highlight a discourse function of information focus. We believe that due to beat gestures (or gestural prominence) the attention that a listener pays to a particular piece of information automatically increases, and as a consequence this leads to higher gains in recall. Also, it is known that redundant multimodal integration cues facilitate speech perception, thus facilitating language processing (Lewkowicz & Hansen-Tift, 2012; van Wassenhove, et al., 2007). Hence, the effect of beat gestures could be attributed to the integration of cross-modal perception processes (Biau, et al., 2016; Hubbard, et al., 2009). There is also research that has shown that the ability to selectively attend to specific elements of speech while disregarding others facilitates language development in its early stages (de Diego-Balaguer, Martinez- Alvarez, & Pons, 2016).

5.3. Why beat gestures have a mnemonic effect

As Andric and Small (2012) states, in contrast to other gesture subtypes, beat gestures lack abstract semantic content. Then it would be possible to suggest that the presence of visual movements (or nonsense hand movements), which are not hand gestures, might have similar effects on memory recall than beat gestures. However, a number of studies have shown that beat gestures act different than any other nonsense hand movements (e.g., Hubbard et al., 2009; Biau & Soto-Faraco, 2013; Dimitrova et al., 2016; Wang and Chu, 2013). In the fMRI study conducted by Hubbard et al. (2009) thirteen adult subjects were exposed to videos with spontaneously-produced speech accompanied by either beat gestures, nonsense hand movements, or no movements. The bilateral non-primary auditory cortex showed greater activity when speech was accompanied by beat gestures than when speech was presented alone. Also, the left superior temporal gyrus/sulcus revealed stronger activity when speech was presented together with beat gestures in comparison to the speech plus nonsense hand movement stimuli. The findings of the study suggest a mutual substrate for processing speech and gesture, the left posterior temporal lobe, which is known to also be sensitive to speech prosody (Humphries et al., 2005). Biau and Soto-Faraco (2013) investigated the time course of beat-speech integration in speech perception. In this study, participants were asked to watch a discourse by a Spanish politician with a wide range of beat gestures, in two conditions, audio-visual and audio only. By measuring participants' ERPs, the researchers found a positive shift in ERPs at an early sensory stage in the audiovisual condition. There was no ERP difference when the same words were heard only with the audio, without viewing the video. Thus, the results of this study suggest that beat gestures are integrated with speech early on in time, and that they modulate word-evoked potentials in early stages of speech processing. The results of the study also support the idea that beat gestures can serve as a highlighter, as they help the listener to direct his or her focus of attention. Finally, the results by Dimitrova et al. (2016) showed that beat gestures behaved differently from other hand movements such as grooming gestures, which served as a control measure over visual activity. As expected, the late positivity effect was only present during beat gesture observation while grooming hand movements didn't interact with focus processing.

As has been discussed in section 1.2.2 of this thesis, naturally produced beat gestures almost invariably occur together with prosodic prominence in speech. This fact could provide an explanation as to why beat gestures without prosodic prominence did not have a beneficial effect on memory in Study 2. At the same time, the results of Wang and Chu (2013) seem to highlight the independent role of hand gestures relative to prosodic prominence. In this experiment, participants were asked to watch videos of a person speaking and gesturing. Target videos contained critical words, which were presented under six experimental conditions containing a combination of two factors, namely hand movement and pitch accent (accented and unaccented). The results showed that both beat gestures and pitch accentuation presented separately elicited smaller negativities in the N400 time window. These results suggest that prominence in gesture, like prominence in speech, triggers the attentional system separately for semantic processing. Thus Wang and Chu's results seem to contradict the results of our experiment, showing that beat gestures have positive effects only when accompanied by prosodic prominence. Interestingly, aside from differences between the tasks performed in Wang and Chu (2013) and those in the present study, in Wang and Chu's experiments participants could not see the mouth of the speaker. One might argue that the authors' decision to hide lip movements might have had a negative effect on the natural integration of prominence coming from speech and both articulatory and hand gestures.

On the other hand, the results of Study 1 in this thesis confirmed the fact that there is an optimal effect of the joint association of visual and prosodic prominence (as compared with prosodic prominence alone). We consider that the visual and speech prominence encoded by beat gestures marking linguistically relevant functions (e.g., focus marking) have a potential effect on language processing and learning. This means that, first, in a situation of prominence, as stated in Terken (1991), a particular element needs to stand out from the surrounding elements, and second, that the focused element reflects certain properties of the discourse context (Büring, 2007). In this thesis we propose that prominence expressed with beat gestures is dependent on surrounding speech elements (pitch accents), to fully express the semiotic value of the beat gesture (McNeill, 1992). We consider, however, that beat gestures are isolated from semantic, they serve to add pragmatic information regarding the relevance of a particular element in discourse (McNeill, 1992) and in this way promote the learning of this element.

5.4. The beneficial effects of beat gestures on pronunciation improvement

Study 3 in this thesis explored the effects of beat gesture production in the pronunciation training by foreign language speakers. In a between-subject experimental design, non-native speakers of English using a pre- and post-test design were exposed to training videos with a female instructor producing short discourses and accompanying her speech by rhythmic beat gestures. Participants in the experiment were randomly assigned to two groups. Participants from one group were asked to repeat the discourses after the instructor and accompany their speech by gestures. Another group of participants was asked to repeat only speech without gestures. The results showed that in general, training with beat gestures is beneficial for pronunciation improvement, since the accentedness ratings were significantly lower in post training for all groups of participants. The results are in line with Gluhareva and Prieto (2017) research that confirmed that training with observing beat gestures significantly improved the participants' accentedness ratings.

The results of our research extend the findings by Gluhareva and Prieto (2017) as they demonstrate that encouraging learners to produce gestures themselves adds a beneficial effect in comparison to only observing beat gestures. Our results support the previous findings that suggested that the production of gestures by learners themselves is more effective than merely observing gestures. (Goldin-Meadow, 2014; Goldin-Meadow et al., 2009; Masumoto, 2006; Macedonia et al., 2011; Saltz & Donnenwerth-Nolan, 1981). The novelty of our research lies is the fact that, to our knowledge, this is the first study with beat gestures in relation to this topic. Previously, other types of gestures (i.e., mora gestures, pitch gestures) had been investigated.

Our findings are in line with the previous literature on the relationship between gesture and prosody in one's first language. Previous research has suggested a close relation between beat gesture and prosodic prominence in natural interactions (e.g., Biau & Soto-Faraco, 2013; Krahmer & Swerts, 2007; Loher, 2012; Wagner et al., 2014). For example, the study by Biau and Soto-Faraco (2013) showed that beat gestures help the listeners regulate the parsing of a stream of speech and to focus the attention on the most relevant parts of the discourse. Krahmer and Swerts (2007) found that highlighted words are perceived as more prominent when accompanied by beat gestures.

The results of Study 3 of this thesis has demonstrated that beat gestures reveal similar beneficial effects in second language as in the first one. The results of our research demonstrated that beat gestures enhance prominence perception when witnessed by L2 speakers as they do in first language (as shown by Krahmer & Swerts, 2007). Observing beat gestures provides an additional source of linguistic information for L2 speakers and consequently facilitates their language production. Also, the results of our research are in line with the studies of Morett and Chang (2015) and Baills et al., (under revision) who confirm that producing rhythmic pitch gestures benefit foreign language learning. In their study and Chang (2015) showed that the production of pitch gestures by participants had positive results on learning words with different in lexical tones in Mandarin Chinese. Baills et al. (under revision) examined whether a short training session could enhance the learning of Mandarin lexical tones and words when participants were asked to observe and also mimic pitch gestures. The results of their study, and the results confirmed that producing gestures leads to higher gains than merely observing them.

However, we must notice that the results of Study 3 contrast with the findings in Hirata and Kelly (2010) and Hirata et al. (2014). The results of these studies yield no beneficial effect of beat gestures on learning in a second language. However, it should be taken into consideration that while Hirata and Kelly assessed segmental-level improvement, the present study centered on a different level of language learning, namely suprasegmental (rhythmic) improvement. In addition to this, the form of beat gestures used in the studies is different (as discussed in section 1.7.).

All in all, the findings in Study 3 may give additional support to the importance of an explicit, global approach to second language pronunciation instruction. Importantly, the findings of our study demonstrate that producing beat gestures facilitates learning tasks more than simply observing them. These findings are in line with the previous research that confirms the same effect but for other types of gestures (see, among others, Goldin-Meadow et al., 2012). Also, the innovative method used in the present study provides support for the use of natural materials, as well as for spontaneous (non-read) methods of eliciting speech from L2 speakers.

5.5. Final conclusions: Implications for multisensory processing theories and their implementation in the second language classroom

Different theories have been put forward over the years which have sought to explain the fact that when gesture and speech are produced in an aligned manner, specific processing mechanisms seem to take place in our cognitive system, reinforcing multimodality channels. The results of our three studies put together seem to back up this line of reasoning.

Our research centered on assessing the effects of the presence of visual information of prominence expressed by beat gestures on information memorization and pronunciation improvement. It is a common fact that we express and perceive information multimodally, through different channels, e.g. aurally and visually. The notion of multisensory processing refers to the interaction of signals that arrive from different sensory modalities. During the last decades, several researchers have discussed how multimodality (the co-occurrence of several modalities) can reinforce memorization. There exist different theories about the effects of multimodal encoding of information on memory. A number of studies have claimed that memory enhancement is triggered by gestures in terms of the depth of encoding. For example, Quinn-Allen (1995) states that gestures provide a context for verbal language that results in deep processing and internalization of verbal information. According to the dual coding theory, proposed by Clark and Pavio (1991), learning processes can be improved when both auditory and visual modalities work together. Baddeley's (1990) model claims that information is better coded using the two modalities because the combination of different modalities leaves more traces in the memory system. The model posits three independent components: 1) the articulatory loop, which is a speech-sound-based storage system of a limited quantity of phonological items; 2) the visual-spatial sketchpad, which encodes non-verbal visual and spatial information; and 3) the central executive device, which coordinates the two other components and directs attention to incoming stimuli. Thus, according to Baddeley (1990), working memory consists of separate auditory and visual working sub-memories, and consequently the representation of auditory and visual information occurs in independent systems. Within this logic, the addition of redundant visual information (for example, gestural prominence) to speech information creates multimodal memory traces and learning can be improved when the information is presented visually and aurally. This idea is also supported by Moreno and Mayer (2000), who claim that multimedia learning is beneficial, because it includes both auditory and visual information. The cognitive theory of multimedia learning, elaborated by Moreno and Mayer (2000) states that working memory includes independent auditory and visual working memories that are separated systems for processing verbal and non-verbal information.

However, a more recent approach by Baddeley (2000) no longer maintains a strict separation between modality-specific subsystems. Baddeley (2000) proposed an extension to the working memory model presented in Baddeley (1990) by introducing a component called episodic buffer which includes a temporary storage of information that is presented through multimodal code, and which transforms the information into a unitary episodic representation. One of the main claims in Baddeley's new theoretical model is on the process of integrating the modality-specific information, rather than maintaining the isolation of the information in different subsystems. Thus, the model proposed by Baddeley (2000) does not maintain the idea of storing the complex images in two separate slave systems, that are responsible for the maintenance of verbally cued images, but rather suggests the existence of a store that draws information from the slave system and from long-term memory and holds it in an integrated way.

Furthermore, there is a line of research that highlights the effect of enactment and of the motor modality on memorization. Recall of enacted information was demonstrated to be superior to recall of information without enactment (Engelkam & Cohen, 1992; Cohen & Otterbein, 1992; Engelkamp & Zimmer, 1985). Engelkamp and Zimmer (1985) posited an enactment effect on memorization by placing a motor system above the visual and verbal memory systems, as encoding of enacted events involve a verbal modality, a visual modality and a motor modality. Thus, enactment makes the memory trace more prominent and consequently easier to find at recall. Neuroimaging studies confirm that brain activity is higher during recall after enacted encoding in comparison to recall after verbal encoding only (e.g., Nyberg et al., 2002).

The grounded or embodied cognition theory (Barsalou 2008) also underlines the beneficial role of enactment to the cognitive processes. A main principle of grounded or embodied cognition theory (Barsalou, 2008) is that cognition is grounded in multiple ways and that it shares mechanisms with not only perception and introspection but also with action. A set of studies within this paradigm highlight the importance of the body in cognition, as bodily states can cause cognitive states and be the result of them (e.g., Barsalou et al., 2003; Lakoff & Johnson, 1980; Smith, 2005 cited in Barsalou, 2008). Gesture is considered an important form of embodiment in language, and it is closely linked to memory (Barsalou, 2008). Recent work on embodied cognition states that language and body movements are supported by the same neural substrates (e.g., Glenberg and Kaschak, 2002; Pulvermüller et al., 2005). The cognitive system utilizes the environment and the body as external informational structures that support internal representations (Barsalou et al., 2003; Niedenthal et al., 2005). There is neurophysiological evidence that self-performing a gesture when learning verbal information forms sensorimotor networks that represent and store the words in both native (Masumoto, 2006) and foreign languages (Macedonia et al., 2011). Additionally, there is also evidence that not only gesture production, but also gesture observation leads to the formation of motor memories in the primary motor cortex (Stefan et al., 2005), which is considered a likely physiological step in motor learning.

In conclusion, the contribution of this thesis to the study of gesture and prosodic prominence is twofold. Firstly, our research confirms previous findings supporting the idea that adding gestures to speech as a visual modality favors information coding and, as a consequence, leads to learning improvement. To this end, we carried out three studies focusing on one specific type of gesture: beat gestures. Beat gestures have not been largely examined before. Indeed, the three studies presented in this thesis report on the beneficial effect of beat gestures for word memorization, word learning, and pronunciation learning. Previous research had already confirmed the importance of gestures in second language acquisition, and specifically in teaching language (see Stam, 2014, 2016 for review; Busà, 2015). As Busà (2015) points out, the study of body language should be integrated in the syllabus of foreign language teaching and learning. However, research centered on beat gestures is scarce. Secondly, these results have implications for instructional practices, as it is common to see a profuse use of gesture in the second language classroom (e.g., Smotrova & Lantolf, 2013). Our findings suggest that the teachers' use of prosodic and gestural prominence may help students to acquire novel words in a second language and improve pronunciation. From a pedagogical perspective, our findings support the use of teaching and learning methods that implement beat gestures as one of the embodied cognition strategies in the second language classroom. Finally, our studies contribute further evidence to the discussion around the effect of enactment and of the motor modality on the mnemonic effects of gesture and embodied cognition. Our results prove learning's beneficial effects of adding a specific type of body movement, both in perception, and specifically in production.

All in all, the results of this thesis show that observing or producing beat gestures can be an equally efficient practice to word learning and pronunciation improvement. From a pedagogical perspective, our findings support the use of teaching and learning methods that implement more active audio-visual and embodied cognition strategies in the second language classroom. For example, teachers can choose to use beat gestures while teaching novel vocabulary for the first time or when teaching pronunciation, asking learners to pay attention to the gesture while listening to the word. Once learners have observed the teacher performing gestures, the teacher may ask them to repeat the speech accompanied with the rhythmic beat gesture, in order to practice lexical and oral skills. Although more applied research is clearly needed in this domain, the results of this thesis constitute an incentive to start implementing more effective multimodal approaches with beat gestures in the classroom.

6. REFERENCES

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (pp. 2-31). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Alibali, M. & diRusso, A. (1999). The function of gesture in learning to count: More than keeping track. *Cognitive Development*, *14*, 37-56.

Alibali MW, Heath DC, & Myers H.J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, *44*, 169–88.

Allport, D. A. (1985). Distributed memory, modular systems and dysphasia. In S. K. Newman & R. Epstein (Eds.), *Current Perspectives in Dysphasia* (pp. 30–60). Edinburgh: Livingstone.

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*, 529–555.

Andric, M. & S. L. Small (2012). Gestures neural language. *Frontiers in Psycholody 3*, 1-11.

Austin, E. E. & Sweller, N. (2014). Presentation and production: The role of gesture in spatial communication. *Journal of Experimental Child Psychology*, *122*, 92-103.

Austin, E. E. & Sweller, N. (2017). Getting to the elephants: Gesture and preschoolers' comprehension of route direction information. *Journal of Experimental Child Psychology, 163*, 1-14.

Baddeley, A. (1990). *Human memory: theory and practice*. East Sussex: Lawrence Erlbaum.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory. *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baills, F., Suárez-González, N. González-Fuente S., Prieto, P. (submitted). Observing and Producing Pitch Gestures Facilitates the Learning of Mandarin Chinese Tones and Words. *Studies in Second Language Acquisition*.

Barsalou, L. W., Simmons, W.K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84-91.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617-645.

Beattie G. & Coughlan J. (1999). An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the- tongue phenomenon. *British Journal of Psychology, 90*, 35-56.

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science*, *21(11)*, 1605-1610.

Behrman, A. (2014). Segmental and prosodic approaches to accent management. *American Journal of Speech-Language Patholo*gy, *23*, 546–561.

Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia*, *44*(2), 178-190.

Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, *124*(2), 143–152.

Biau, E & Soto-Faraco, S (2015). Synchronization by the hand: the sight of gestures modulates low-frequency activity in brain responses to continuous speech. *Frontiers in Human Neuroscience, 9, 1-6.*

Biau, E., Fromont, L.A., & Soto-Faraco, S. (2017). Beat Gestures and Syntactic Parsing: An ERP Study. *Language Learning*, *10*, 1-25.

Billmyer, K., & Varghese, M. (2000). Investigating instrument-based pragmatic variability: effects of enhancing discourse completion tests. *Applied Linguistics*, *21*(4), 517-552.

Birch, S. L., & Clifton, C. (1995). Focus, accent, and argument structure: effects on language comprehension. *Language and Speech, 38*, 365-91.

Birch, S. L., & Clifton, C. (2002). Effects of varying focus and accenting of adjuncts on the comprehension of utterances. *Journal of Memory and Language*, *47*, 571-88.

Borghi, A. M., & Caruana, F. (2015). Embodiment Theory. In J.D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 420–426). Amsterdam: Elsevier.

Blum-Kulka, S., House, J., & Kasper, G. (1989). Investigating cross-cultural pragmatics: An introductory overview. In S. Blum-Kulka, J. House, & G. Kasper (Eds.), *Cross-cultural pragmatics: Requests and apologies* (pp. 1–34). Norwood, NJ: Ablex.

Broaders, S., Wagner Cook, S., Mitchell, Z., Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology, 136* (4), 539-550.

Bock, J. K., & Mazzella, J. R. (1983). Intonational marking of given and new information: some consequences for comprehension. *Memory & Cognition*, *11*(1), 64 –76.

Borghi, A. M., & Caruana, F. (2015). Embodiment Theory. In James D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 420-426). Oxford: Elsevier.

Busà, M.G. (2015). Teaching learners to communicate effectively in the L2: Integrating body language in the students' syllabus, *Lingue e Linguaggi*, *15*, 83-98.

Butcher, C., & Goldin-meadow, S. (2000). Gesture and the transition to from one to two-word speech: When hand and mouth come together. In D. McNeill (Ed.), *Language and gesture* (pp. 235-257). New York: Cambrige University Press.

Capirci, O., Iverson, J.M., Pizzuto, E. & Volterra, V. (1996). Gestures and words during the transition to two- word speech. *Journal of Child Language, 23*, 645- 673.

Cebrian, J. (2002). Phonetic similarity, syllabification and phonotactic constraints in the acquisition of a second language contrast. PhD Dissertation. Toronto Working Papers in Linguistics Dissertation Series. University of Toronto, Toronto, Canada.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics, 34*, 371-387.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review* 3, 149–210.

Cohen, R. L. (1989). Memory for action events: the power of enactment. *Educational Psychology Review, 1,* 57–80.

Cohen, Ronald L. & Nicola Otterbein (1992). Mnemonic effect of speech gestures: Pantomimic and non-pantomimic gestures compared. *European Journal of Cognitive Psychology, 4 (2)*, 113–139.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106*, 1047-1058.

Cruttenden, A. (1997). *Intonation*. Cambridge: Cambridge University Press.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*, 141-201.

Dahan, D., Tanenhaus, M. K., Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*, 292-314.

De Jong N. H., Steinel, M.P., Florijn A. F., Schoonen R. & Hulstijn J.H. (2012). Facets of Spekaing Proficiency. *Studies in Second Language Acquisition, 35 (1)*, 5-34.

De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Eds.), *Language and Gesture* (pp. 284-311). Cambridge University Press.

De Ruiter, L. E. (2015). Intonation status marking in spontaneous vs. read speech in story-telling tasks: Evidence from intonation analysis using GToBI. *Journal of Phonetics*, *48*, 29–44.

Derwing, T.M., & Munro, M.J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, *19*, 1–16.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamental: Evidence-based Perspectives for L2 Teaching and Research.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning, 13(1)*, 1–17.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54(4),* 655-679.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*(3), 393–410.

Derwing, T.M., & Munro, M.J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition, 19*, 1–16.

Dimitrova, D. V., Chu, M., Wang, L., Ozyurek, A., & Hagoort, P. (2016). Beat that word: How listeners integrate beat gesture and focus in multimodal speech discourse. *Journal of Cognitive Neuroscience*, *28*(9), 1255-1269.

Dobrogaev, S. M. (1929). Ucnenie o reflekse v problemakh iazykovedeniia [Observations on reflexes and issues in language study]. Iazykovedenie i Materializm, 105-73.

Dohen, M. (2009). Speech through the ear, the eye, the mouth and the hand. In A. Esposito, A. Hussain, & M. Marinaro (Eds.), *Multimodal signals: Cognitive and algorithmic issues* (pp. 24-39). Berlin/Heidelberg: Springer.

Edmunds, P. (2010). ESL speakers' production of English lexical stress: The effect of variation in acoustic correlates on perceived intelligibility and nativeness. Unpublished PhD dissertation, The University of New Mexico, Albuquerque, NM, USA.

Efron, D. (1941/1972). *Gestures, race and culture*. New York: King's Crown Press.

Ekman, P. (1979). About brows: emotional and conversational signals. In M. von Cranach, K. Forra, W. Lepinies, & D. Ploog (Eds.), *Human ethology: Claims and limits of a new discipline:*

Contribution to the Colloquium (pp. 169-248). Cambridge: Cambridge University Press.

Ekman, P., & Friesen, W.V. (1969). The repertorie of nonverbal behavioral categories. *Semiotica, 1*, 49-98.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements. *Journal of Speech, Language, and Hearing Research*, *56*, 850-865.

Esteve-Gibert, N. & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication, 57*, 301-316.

Ekman, P. (1979). About brows: emotional and conversational signals. In M. von Cranach, K. Forra, W. Lepinies, & D. Ploog (Eds.), *Human ethology: Claims and limits of a new discipline: Contribution to the Colloquium* (pp. 169-248). Cambridge: Cambridge University Press.

Feyereisen, P. (2006). How could gesture facilitate lexical access? *Advances in Speech-Language Pathology*, *8,* 128-133.

Félix-Brasdefer, J. C. (2010). Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In A. Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance:*

Theoretical, empirical, and methodological issues (pp. 41–56). Amsterdam & Philadelphia, PA: John Benjamins.

Feyereisen, P. (1998). Le rôle des gestes dans la mémorisation d'énoncés oraux. In Serge Santi, Isabelle Guaïtella, Christian Cavé, & Gabrielle Konopczynski (Eds.), *Oralité et ges- tualité. Communication multimodale, interaction. Actes du colloque Orage* (pp. 355–360). Paris: L'Harmattan.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*, 399–423.

Flege, J. E. (1995). Second language speech learning: theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 233-272). Timonium, MD: York Press.

Fouz-González, J. (2015). Trends and Directions in Computer-Assisted Pronunciation Training. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English Pronunciation* (pp. 314-342). UK: Palgrave Macmillan.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how contrastive accenting really is. *Journal of Memory and Language*, *63*(3), 367–386.

Gass, S. M. (2013). *Second Language acquisition: An introductory course*. London: Routledge.

Glenberg, A.M., & Kaschak, M.P. (2002). Grounding language in action. *Psychonomic Bulletin and Review* 9, 558–65.

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

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.

Goldin-Meadow, S., & Butcher, C. (2003). Pointing: Where language, culture, and cognition meet. In S. Kita (Eds.), *Pointing toward two word speech in young children* (pp. 85-107). Hillsdale, NJ: Lawrence Erlbaum Associates.

Goldin-Meadow S, Kim S, Singer M. (1999). What the teachers' hands tell the students' minds about math. *Journal of Educational Psychology*, 91, 720–30.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing Gives Children New Ideas About Math. *Psychological Science : A Journal of the American Psychological Society, 20(3),* 267–272.

Goldin-Meadow, S., Levine, S.L., Zinchenko, E., Yip, T.K.-Y., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science, 15(6)*, 876–884.

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

Goodwyn, Susan W., Linda P. Acredolo, & Catherine A. Brown (2000). Impact of symbolic gesturing on early language development. *Journal of Nonverbal Behavior, 24 (2)*, 81- 103.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing Gives Children New Ideas About Math. Psychological Science : *A Journal of the American Psychological Society, 20(3)*, 267–272.

Gordon, J., Darcy, I., & Ewert, D. (2013). Pronunciation teaching and learning: Effects of explicit phonetic instruction in the L2 classroom. In: J. Levis, & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 194–206). Ames, IA: Iowa State University.

Graham, C. (1978). *Jazz chants: Rhythms of American English for students of English as a second language*. New York: Oxford University Press.

Graziano, M. & Gullberg, M. (2013). *Gesture production and speech fluency in competent speakers and language learners*: Tilburg University.

Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (Homage to Adam Kendon). *International*

Review of Applied Linguistics in Language Teaching, 44(2), 103–124.

Gullberg, M. (1998). Gesture as a Communication Strategy in Second Language Discourse: A Study of Learners of French and Swedish. Travaux de l'institut de linguistique de Lund, 35. Linguistics and Phonetics. Dissertation.

Gullberg, M. (2010). Methodological reflections on gesture analysis in SLA and bilingualism research. *Second Language Research, 26*, 75-102.

Gullberg, M., deBot, K., & Volterra, V. (2008). Gestures and some key issues in the study of language development. *Gesture, 8(2)*, 149–179.

Hannah, B., Wang, Y., Jongman, A., & Sereno, J.A. (2016). Cross-modal association between auditory and visual-spatial information in Mandarin tone perception. *The Journal of the Acoustical Society of America, 140(4)*, 3225.

Hardison, D. M. (2007). The visual element in phonological perception and learning. In M.C. Pennington (Eds.), *Phonology in context* (pp. 135-158). Basingstoke, UK: Palgrave Macmillan.

Heugten, M., Dautriche, I., & Christophe, A. (2014). Phonological and Prosodic Bootstrapping. In P. J. Brooks, & V. Kempe (Eds.), *Encyclopedia of language development* (pp. 447-451). Thousand Oaks, CA: SAGE Publications.

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*, 298–310.

Hirata, Y., Kelly, S., Huang, J., & Manansala, M. (2014). Effect of hand gestures on Auditory Learning of Second-Language Vowel Length Contrasts. *Journal of Speech, Language, and Hearing Research*, *57*, 2090-2101.

Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., & Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Frontiers in Psychology*, *3*, 74.

Hruska C, Alter K, Steinhauer K, Steube A (2001). Misleading dialogues: Human's brain reaction to prosodic information. In C. Cavé, I. Guaitella, S. Santi (Eds.), *Orality and Gestures* (pp. 425-430). Paris: L'Harmattan.

Hu, F. T., Ginns, P., & Bobis, J. (2015). Getting the point: Tracing worked examples enhances learning. *Learning and Instruction*, 35, 85-93.

Hubbard, A. L., Wilson, S. M., Callan, D. E., & Dapretto, M. (2009). Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapping*, *30(3)*, 1028–1037.

Hulstijn, J. (2005) Theoretical and empirical issues in the study of implicit and explicit second language learning: Introduction. *Studies in Second Language Acquisition*, 27, 129–140.

Humphries, C., Love, T., Swinney, D., & Hickok G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Human Brain Mapping, 26*, 128– 38.

Igualada, A., Esteve-Gibert, N., & Prieto, P. (2017). Beat gestures improve word recall in 3- to 5-year-old children. *Journal of Experimental Child Psychology, 156*, 99-112.

Iverson, J. M. & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature, 396*, 228.

Iverson, J. M. & Thelen, E. (1999). Hand, Mouth and Brain. *Journal of Consciousness Studies, 6(11–12)*, 19–40.

Jannedy, S. & Mendoza-Denton, N. (2005). Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure, 3*, 199–244.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System, 38*, 301–315.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. Key (Eds.), *The relationship of verbal and nonverbal communication* (pp. 207-227). The Hague: Mouton.

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

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.

Kelly, S. D., Ozyürek, A., & Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21(2)*, 260–7.

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

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.

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

Kiefer, M., & Trumpp, N. M. (2012). Embodiment theory and education: The foundations of cognition in perception and action. *Trends in Neuroscience and Education*, *1*(1), 15–20.

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.

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal Behavior and Nonverbal Communication: What do Conversational Hand Gestures Tell Us? *Advances in Experimental Social Psychology, 28*, 389–450.

Kushch, O., Igualada, A., & Prieto, P. (accepted). Prominence in speech and gesture favor second language novel word learning. *Language, Cognition and Neuroscience*.

Laan, G. (1997). The contribution of intonation, segmental duration, and special features to the perception of a spontaneous and real speaking style. *Speech Communication*, *22*, 43–65.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge and New York, NY: Cambridge University Press.

Larson-Hall (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

LeBreton, J.M., & Senter, J.M. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*, 815–852.

Leonard, T. & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes, 26*, 1295-1309.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society*, 369(2013- 2030).

Loehr, D. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, *3(1),* 71-89.

Llanes-Coromina, J., Vilà-Giménez, I., Kushch, O., & Prieto. P. (under revision). Beat gestures help preschoolers recall and comprehend discourse information. *Journal of Experimental Child Psychology*.

Macoun, A., & Sweller, N. (2016). Listening and watching: The effects of observing gesture on preschoolers' narrative comprehension. *Cognitive Development, 40*, 68-81.

Macedonia, M., Müller, K. & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*, 982–998.

Marcos, L. R. (1979). Nonverbal behavior and thought processing. *Archives of General Psychiatry, 36(9)*, 940-943.

Masumoto, K., Yamaguchi, M., Sutani, K., Tsuneto, S., Fujita, A., & Tonoike, M. (2006). Reactivation of physical motor information in the memory of action events. *Brain Research, 1101(1)*, 102–109.

Matatyaho-Bullaro D., Gogate, L., Mason, Z., Cadavid, S., Abdel-Mottaleb, M. (2014). Type of object motion facilitates word mapping by preverbal infants. *Journal of Experimental Child Psychology, 118*, 27–40.

Meara, P. M. & Miralpeix, I. (2017). *Tools for researching vocabulary*. Bristol: Multilingual Matters.

Meyer, M., Hard, B., Brand, R., McGarvey M, Dare A. Baldwin, D.A. (2011). Acoustic Packaging: Maternal Speech and Action Synchrony. *IEEE Transactions on autonomous mental development*, 3-1.

Miralpeix, I. & Meara, P. M. 2014. The written word. In J. Milton & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge* (pp. 30-44). Palgrave: Macmillan.

McCafferty, S. G. (2006). Gesture and the materialization of second language prosody. *IRAL:International Review of Applied Linguistics in Language Teaching, 44*, 197–209.

McClave, E. (1998). Pitch and Manual Gestures. *Journal of Psycholinguistic Research*, *27*(1), 69-89.

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

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Mora, J. C. (2007). Learning context effects on the acquisition of a second language phonology. In C. Pérez-Vidal, M. Juan-Garau, & A. Bel (Eds.), *A portrait of the young in the new multilingual Spain* (pp. 241-236). Clevedon: Multilingual Matters.

Mora, J. C., & Fullana, N. (2007). Production and perception of English /h9/-/H/ and /æ/-// in a formal setting: Investigating the effects of experience and starting age. *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany, 1613-1616.

Mora, J. C. & Levkina, M. (2017). Task-Based Pronunciation Teaching and Research: Key Issues and Future Directions. *Studies in Second Language Acquisition*, *39*, 381-399.

Morett, L., M. & Chang L.-Y. (2015). Emphasising sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, *30*(3), 347-353.

Munro, M., & Derwing, T. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*, 73–97.

Munro, M., & Derwing, T. (1998). The Effects of Speaking Rate on Listener Evaluations of Native and Foreign-Accented Speech. *Language Learning*, *48*, 159-182.

Munro, M.J., & Derwing, T.M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*, 73–97.

Muñoz, C. (2013). Explicit Learning in Second Language Acquisition. In C. A, Chapelle (Ed.), *The encyclopedia of Applied linguistics* (pp. 1-6). Oxford: Wiley Blackwell.

Namy, L.L., Acredolo, L., Goodwyn, S. (2000). Verbal labels and gestural routines in parental communication with young children. *Journal of Nonverbal Behavior, 24*(2), 63- 79.

Nathan, M. J. (2008). An embodied cognition perspective on symbols, gesture and grounding instruction. In M. DeVega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment and meaning: a debate* (pp. 375-396). Cambridge: Oxford University Press.

Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, 9, 184–211.

Nicoladis, E., Pika, S., Yin, H.U.I. & Marentette, P. (2007). Gesture use in story recall by Chinese–English bilinguals. *Applied Psycholinguistics, 28*, 721–35.

Özcaliskan, S. & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition, 96(3*), 101-113.

Paivio, A. (1990). *Mental representations: A dual coding approach.* New York, NY: Oxford University Press.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart & Winston.

Pérez-Vidal, C. (2009). The integration of content and language in the classroom : A European approach to education (The second time around). In E. Dafouz & M. Guerrini (Eds.), *CLIL Across Educational levels: Experiences from primary, seconday and tertiary contexts* (pp. 25-40). London: Richmond Publishing.

Pérez-Vidal, C. (2014). Study abroad and formal instruction contrasted: The SALA project. In C. Pérez-Vidal (Ed.), *Second language acquisition in study abroad and formal instruction contexts* (pp. 17-57). Amsterdam/Philadelphia: John Benjamins.

Poggi, I. (2007). *Mind, hands, face and body. A goal and belief view of multimodal communication*. Berlin: Weidler.

Pons, F., & Bosch, L. (2010). Stress pattern preference in Spanish-learning infants: The role of syllable weight. *Infancy*, *15*(3), 223–245.

Prieto, P., Puglesi, C., Borràs-Comes, J., Arroyo, E., & Blat, J. (2015). Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. *Journal of Phonetics, 49*, 41-54.

Prieto, P., & Roseano, P. (2010). *Transcription of Intonation of the Spanish Language*. München: Lincom.

Pulvermuller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience 21*, 793–797.

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of workin memory. *Frontiers in Human Neuroscience*, *9*, 197.

Quinn-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*, 521–529.

Ravizza S. (2003). Movement and lexical access: Do noniconic gestures aid in retrieval, *Psychonimic Bulletin & Review, 10*, 610-615.

Rauscher, F.H., Krauss, R.M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*, 226 – 231.

Richards, J.C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, *10*(1), 77-89.

Rochet-Capellan, A., Laboissière, R., Galván, A., & Schwartz, J. (2008). The Speech Focus Position Effect on Jaw-Finger Coordination in a Pointing Task. *Journal of Speech Language and Hearing Research, 51(6)*, 1507-1521.

Rusiewicz, H. L. (2010). The Role of Prosodic Stress and Speech Perturbation on the Temporal Synchronization of Speech and Deictic Gestures. Doctoral Dissertation, University of Pittsburgh.

Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2013). Effects of prosody and position on the timing of deictic gestures. *Journal of Speech, Language, and Hearing Research, 56*, 458-470.

Rusiewicz, H. L. & Esteve-Gibert, N. (in press). Set in time: Temporal coordination of prosodic stress and gesture in the development of spoken language production. In P. Prieto & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language acquisition*. John Benjamins.

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, *33*(1), 3–11.

Saltz, E., & Donnenwerth-Nolan, S. (1981). Does motoric imagery facilitate memory for sentences? A selective interference test. *Journal of Verbal Learning and Verbal Behavior, 20(3),* 322–332.

Sanaoui, R. (1995). Adult learners' approaches to learning vocabulary in second languages. *The Modern Language Journal*, 79(1), 15–28.

Senkfor, A.J., Van Petten, C. and Kutas, M. Episodic action memory for real objects: an ERP investigation with perform, watch, and imagine action encoding tasks versus a non-action encoding task. *Journal of Cognitive Neuroscience*, 2002, *14*(3), 402-419.

Shattuck-Hufnagel, S., Ren, A., Mathew, M., Yen, I., Demuth, K. (2016). Non-referential gestures in adult and child speech: Are they prosodic? *Proceedings of Speech Prosody 2016*. Boston, USA, May 31-June 3, 2016.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, *12*(6), 914–919.

Skoruppa, K., Pons, F., Bosch, L., Christophe, L., Cabrol, D., & Peperkamp, S. (2013). The Development of Word Stress Processing in French and Spanish Infants. *Language Learning and Development*, *9* (1), 88-104.

Smotrova, T., & Lantolf, J. P. (2013). The function of gesture in lexically focused L2 instructional conversations. *Modern Language Journal*, 97, 397– 416.

So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, *5*, 665-681.

Stefan, K., Cohen, L. G., Duque, J., Mazzocchio, R., Celnik, P., Sawaki, L., Ungerleider, L. & Classen, J. (2005). Formation of a motor memory by action observation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(41), 9339–9346.

Swerts, M., & Krahmer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics*, *36*(2), 219–238

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics, 30*, 629–654.

Tellier, Marion (2005). L'utilisation des gestes en classe de langue: comment évaluer leur e et sur la mémorisation du lexique? In

Michel Billières, Pascal Gaillard, & Nathalie Spanghero-Gaillard (Eds.), *Actes du Premier colloque international de Didactique Cognitive: DIDCOG 2005*. Université de Toulouse–Le Mirail, 26–28 Janvier 2005. Proceedings on CD-ROM.

Tellier, Marion (2007). How do teacher's gestures help young children in second language acquisition? *Proceedings of the meeting of International Society of Gesture Studies, ISGS 2005*: Interacting Bodies, June 15–18, ENS Lyon, France.

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

Ulbrich, C. (2013). German pitches in English: Production and perception of cross-varietal differences in L2. Bilingualism: *Language and Cognition*, 16, 397–419.

Vallduví, E. (1991). The role of plasticity in the association of focus and prominence. *ESCOL* 7, 295-306.

Van Maastricht, L., Krahmer, E., & Swerts, M. (2015). Native speaker perceptions of (non-)native prominence patterns: Effect of divergence in pitch accent distributions on accentedness, comprehensibility, intelligibility, and nativeness. *Speech Communication, 83, 21-33.*

Vilà-Giménez, I, Igualada, A., & Prieto, P. (under revision). Training with rhythmic beat gestures improves children's narrative discourse skills. *Developmental Psychology.*

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, *57*, 209–232.

Wang, L., & Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: An ERP study. *Neuropsychologia*, *51*(13), 2847–2855.

Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology*, 1–10.

White, L., & Mattys, S.L. (2007). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 237-257). Amsterdam: John Benjamins.

Willems, R.M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology, 2*(116), 1-11.

Wong, J. W. S. (2015). The Impact of L2 Proficiency on Vowel Training. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English Pronunciation* (pp. 219-239). UK: Palgrave Macmillan.

Wundt, W. (1921/1973). *The language of gestures*. The Hague: Mouton.

Yasinnik, Y., Renwick, M., Shattuck-Hufnagel, S. (2004). The timing of speech-accompanying gestures with respect to prosody," *Proceedings of the International Conference: From Sound to Sense: +50 Years of Discoveries in Speech Communication*, *MIT*, Cambridg, 10-13 June, C97 – C102.

Yeo, A., Ledesma I., Nathan M. J., Alibali M., Church, B. (2017) Teachers' gestures and students' learning: sometimes "hands off" is better. *Cognitive Research: Principles and implications*, *2*, 41.

Yuan, C., González-Fuente, S., Baills, F. & Prieto, P. (2017). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, 1-28.

Yoshida, H. & Smith, L. B. (2007). Meaning From the First-Person View: A Role for Hands. In Hanako Yoshida (Chair) *Meaning From the Hands*. Symposium conducted at the biennial meeting of the Society for Research on Child Development, Boston, MA.

Yu, C., Ballard, D. H., & Aslin, R. N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science, 29*, 961–1005.

Zammit, M., & Schafer, G. (2011). Maternal label and gesture use affects acquisition of specific object names. *Journal of Child Language, 38(1),* 201-221.

Appendix A

Russian word		Catalan translation	English equivalent
сумка	/sumka/	bossa	bag
ветка	/vetka/	branca	branch
кофта	/kofta/	brusa	blouse
песня	/pesnja/	cançó	song
вишня	/vishnja/	cirera	cherry
белка	/belka/	esquirol	squirrel
дырка	/dirka/	forat	hole
дудка	/dutka/	flauta	flute
вилка	/vilka/	forquilla	fork
вафля	/vaflja/	gofre	waffle
кепка	/kepka/	gorra	cap
место	/mesto/	lloc	place
кукла	/kukla/	nina	doll
булка	/bulka/	pa	bread
шорты	/shorti/	pantalons	trousers
судно	/sudno/	vaixell	vessel

Appendix B:

Materials for the pre-/post-training task

2 Familiarization items

You are at your classmate's birthday party. You meet her boyfriend for the first time and introduce yourself to him.

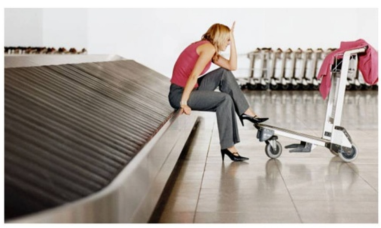
You are trying to find Central Park. You ask a stranger for directions.

10 Experiemental ítems

You are in a restaurant and would like to order a steak with French fries and a glass of red wine.

You are in the metro and would like to ask a stranger for the time.

You are at the market. You want to ask the price of the necklace and ask if you can get it for $5.



You arrive at the airport in New York. You realize that your luggage is lost. You ask an airport employee for help.

You are at the pharmacy. You would like to tell the pharmacist that you have a sore throat and a fever and ask her to prescribe something for you.

You are trying to find an apartment in your new city. You want to ask the agent if this apartment gets a lot of light in the mornings.

You call a pizzeria. You would like to place an order for delivery—two large pizzas with cheese and pepperoni.

You get into a taxi. You would like to ask the driver to take you to the airport as fast as he can, because you are running late for your flight.

You are in a lecture at the university. You didn't hear what the professor just said and would like to ask your friend to repeat it for you.

You are in a clothing store. You would like to tell the clerk that you are looking for this shirt in a bigger size, and ask her if they have it in the back of the store.

10 New unrelated items in the post-test

You are at your university's Student Administration Centre. You would like to tell the secretary that you applied for a new student ID card last week and ask her if it is ready yet.

You go to the computer repair shop. You would like to tell the technician that your computer has been running slowly and ask him to figure out what the problem is.

You go to the phone store. You ask the clerk to show you their newest phone model.

You are at the bank. You would like to ask the teller how to apply for a new student bank account and what documents you need to provide.

You go to the cinema. You want to ask the clerk if there is a discount for students.

Your new roommate has been leaving her dirty dishes in the sink. You want to ask her to clean up after herself.

You go to your professor's office. You want to ask him why you got question number 7 wrong on the test.

You see your classmate at the library. You ask her if she wants to study for the Economics test together.

The roof in your apartment is leaking. You call the repairman and ask him when is the earliest he can come.

You go to a gym in your new neighborhood. You ask the employee how much a new membership costs, and if the gym is open late on Sundays.

Appendix C

Training video transcript

1. HI, I’m MAya. It’s GREAT to meet you.

2. ExCUSE me, we are looking for Central PARK. Could you TELL us where to GO?

3. HI, I’d like to place an ORder for deLIvery. Two large PIzzas with CHEESE and peppeROni.

4. SORRY, what did the professor just SAY? I couldn’t HEAR him.

5. How much is this NECKlace? Can I get it for five DOllars?

6. ExCUSE me, what TIME is it?

7. My LUggage is LOST. Could you HELP me?

8. I’d like to get a STEAK with FRENCH fries, and a glass of red WINE, please.

9. I’m looking for this SHIRT in a bigger SIZE. Could you check and SEE if you have it in the BACK?

10. Can you TAKE me to the AIRport? As fast as you CAN please. I’m LATE for my flight.

11. Does this aPARTment get a lot of LIGHT in the mornings?

12. I have a sore THROAT and a FEver. Could you presCRIbe something for me?

Notes

Full beats are marked with capital letters and intermediate beats are underlined. Emphasis was placed on getting/showing/producing video recordings that appeared as natural as possible; therefore, not all stressed syllables were marked with beat gestures.