The benefits of beat gestures in second language processing and learning

Patrick Louis Rohrer
PhD Research Plan

International Cotutelle Supervised by:
Pilar Prieto
ICREA - Department of Translation and Language Sciences, Universitat Pompeu Fabra

Elisabeth Delais-Roussarie
Laboratoire Linguistique de Nantes (LLING), Université de Nantes

September 17th, 2018
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resum</td>
<td>i</td>
</tr>
<tr>
<td>Resumen</td>
<td>iii</td>
</tr>
<tr>
<td>Abstract</td>
<td>v</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1. Object of analysis</td>
<td>1</td>
</tr>
<tr>
<td>1.2. Literature Review</td>
<td>1</td>
</tr>
<tr>
<td>1.2.1. Co-speech gestures</td>
<td>1</td>
</tr>
<tr>
<td>1.2.2. Prosodic features of co-speech gestures</td>
<td>4</td>
</tr>
<tr>
<td>1.2.3. Properties of beat gestures</td>
<td>7</td>
</tr>
<tr>
<td>1.2.3.1. Prosodic features of beats</td>
<td>8</td>
</tr>
<tr>
<td>1.2.3.2. Pragmatic functions of beats</td>
<td>11</td>
</tr>
<tr>
<td>1.2.4. Cognitive benefits of referential gestures in the first and the second language: behavioral and neurophysiological evidence</td>
<td>14</td>
</tr>
<tr>
<td>1.2.4.1. Behavioral evidence</td>
<td>15</td>
</tr>
<tr>
<td>1.2.4.2. Neurophysiological evidence</td>
<td>20</td>
</tr>
<tr>
<td>1.2.4.3. Stronger gesture effects for second language processing</td>
<td>23</td>
</tr>
<tr>
<td>1.2.5. Cognitive benefits of beat gestures in the first and the second language: behavioral and neurophysiological evidence</td>
<td>25</td>
</tr>
<tr>
<td>1.2.5.1. Neurophysiological evidence</td>
<td>25</td>
</tr>
<tr>
<td>1.2.5.2. Behavioral evidence</td>
<td>29</td>
</tr>
<tr>
<td>2. Goals of the Dissertation</td>
<td>34</td>
</tr>
<tr>
<td>3. Theoretical Framework</td>
<td>35</td>
</tr>
<tr>
<td>3.1. Gesture and speech as one system</td>
<td>35</td>
</tr>
<tr>
<td>3.2. Embodied Cognition</td>
<td>37</td>
</tr>
<tr>
<td>3.3. Dual Coding Theory</td>
<td>38</td>
</tr>
<tr>
<td>4. Hypotheses</td>
<td>38</td>
</tr>
<tr>
<td>4.1. Study 1</td>
<td>39</td>
</tr>
<tr>
<td>4.2. Study 2</td>
<td>40</td>
</tr>
<tr>
<td>4.3. Study 3</td>
<td>40</td>
</tr>
<tr>
<td>4.4. Study 4</td>
<td>41</td>
</tr>
<tr>
<td>5. Experimental Studies</td>
<td>41</td>
</tr>
<tr>
<td>5.1. Study 1: A cross-linguistic examination of the prosodic and pragmatic features of beat gestures</td>
<td>41</td>
</tr>
<tr>
<td>5.1.1. Research questions</td>
<td>41</td>
</tr>
</tbody>
</table>
Resum

S’ha demostrat que els gestos de la parla són beneficiosos per a diversos aspectes de la cognició i influencien positivament en l’aprenentatge (veure Goldin-Meadow, 2011 per a un resum). Mentre que els gestos referencials ajuden a millorar la comprensió (e.g., Morett, Gibbs & MacWhinney, 2012) i la memòria (e.g., Feyereisen, 1998; Tellier, 2005, 2006) i també ajuden en l’adquisició del vocabulari d’una segona llengua (L2) (e.g., Kelly, McDevitt & Esch, 2009; Tellier, 2008), se sap poc sobre l’efecte dels gestos rítmics, o gestos no-referencials que típicament s’associen amb la prominència prosòdica. La recerca prèvia sobre aquest úlim tipus de gest ha mostrat resultats contradictoris pel que fa als seus beneficis cognitius. Estudis recents han mostrat sobre el fet que els gestos rítmics ajuden a una millor memorització de la informació, tant en una primera com en una segona llengua (e.g., Igualada, Esteve-Gibert, & Prieto, 2017; Llanes-Coromina, Vilà-Giménez, Kushch, Borràs-Comes, & Prieto, 2018 per L1; Kushch, Igualada & Prieto, 2018 per L2), com també a la pronúncia d’una L2 (e.g., Gluhareva & Prieto, 2017; Llanes-Coromina, Prieto & Rohrer, 2018). Per contra, altres estudis no han trobat efectes positius en una L1 (Macoun & Sweller, 2016; Austin & Sweller, 2014; Feyereisen, 2006). De la mateixa manera, els gestos rítmics mostren un augment de l’atenció en estudis neurofisiològics (Biau & Soto-Faraco, 2013; Dimitrova, Chu, Wang, Özyürek, & Hagoort, 2016). D’altra banda, la investigació prèvia sobre els gestos referencials ha demostrat que els parlants no nadius semblen beneficiar-se més dels gestos i els integren d’una manera més intensa (amb un efecte de l’N400 més fort) que els parlants nadius (Dahl & Ludvigsen, 2014; Drijvers & Özyürek, 2018). Tanmateix, la comparació entre l’impacte dels gestos rítmics en una L2 i en una L1 no ha estat investigada amb profunditat.

L’objectiu general d’aquesta tesi és avaluar el rol dels gestos rítmics (acompanyats de prominència prosòdica) en el processament de la parla en una primera i segona llengua. Es pretén estudiar com els gestos rítmics beneficien la memòria, la comprensió i la memòria fonològica en parlants nadius i no nadius de tres llengües diferents (anglès, francès i català). La hipòtesi general que guia la tesi és basa en la idea que els parlants nadius integraran els gestos rítmics de manera més intensa que els parlants nadius, i, per tant, s’espera que els efectes de d’aquests gestos en el processament siguin més visibles en una segona llengua que en una primera llengua.

Aquesta recerca inclou quatre estudis que responen a tres objectius principals, els quals estan relacionats amb el paper dels gestos rítmics en el processament de la parla en una
primera i segona llengua. L’Estudi 1 té com a objectiu principal clarificar la relació temporal i pragmàtica entre els gestos rítmics i el discurs en anglès i francès, a través de l'anàlisi d’un corpus de 15 Ted Talks en anglès i 15 Ted Talks en francès, dues llengües que codifiquen la prominència prosòdica de manera diferent (en anglès predomina una accentuació lèxica a inici de paraula, mentre que l’accentuació del francès opera a un nivell de la frase, amb un accent inicial no obligatori i un accent obligatori al final de la frase). L’Estudi 2 utilitzarà un paradigma de potencials evocades (ERP) per analitzar la modulació de dos components ERP (el component P3a i el component N400) per tal d’investigar els efectes de processament dels gestos rítmics en una primera i una segona llengua (Dimitrova et al.; 2016; Drijvers & Özyürek, 2018; Kutas & Hillyard, 1980). L’Estudi 3 investigarà si l’exposició als gestos rítmics durant una tasca de dibuixar una narració verbal (adaptada de Dahl & Ludvigsen, 2014) aportarà una millor memòria i comprensió de la parla als parlants no nadius, comparat amb els nadius. L’Estudi 4 explorarà els efectes de la presència de gestos rítmics sobre la memòria fonològica, de la qual s’espera una millora de la pronunciació en la segona llengua.

Els participants realitzaran una “shadowing task” i les produccions seran avaluades pels nadius tenint en compte el nivell d’accent, la intel·ligibilitat i la fluïdesa. En resum, els resultats dels quatre estudis ens permetran avaluar d’una manera més completa la influència dels gestos rítmics com a markadors multimodals i els seus efectes potencials en el processament i aprenentatge d’una segona llengua.

**Paraules clau:** Gestos rítmics, gestos no-referencials, prosòdia, ERPs o paradigma de potencials evocats, aprenentatge d’una segona llengua
Resumen

Se ha demostrado que los gestos co-verbales son beneficiosos en diversos aspectos de la cognición e influencian positivamente el aprendizaje (ver Goldin-Meadow, 2011 para un resumen). Mientras los gestos referenciales muestran un incremento de la comprensión auditiva (e.g., Morett, Gibbs, & MacWhinney, 2012), aumentan la memoria (e.g., Feyereisen, 1998; Tellier, 2005, 2006), y ayudan en la adquisición de una segunda lengua (L2) (e.g., Kelly, McDevitt & Esch, 2009; Tellier, 2008), se sabe poco del efecto de los llamados gestos rítmicos, o gestos no-referenciales que típicamente se asocian con la prominencia prosódica. Investigaciones sobre estos últimos muestran resultados contradictorios por lo que se refiere a los beneficios cognitivos. Estudios recientes reflejan que los gestos rítmicos ayudan a una mejor memorización de la información en la primera y segunda lengua (e.g., Igualada, Esteve-Gibert, & Prieto, 2017; Llanes-Coromina, Vilà-Giménez, Kushch, Borràs-Comes, & Prieto, 2018 para L1; Kushch, Igualada & Prieto, 2018 para L2), y también con la pronunciación en L2 (e.g., Gluhareva & Prieto, 2017; Llanes-Coromina, Prieto & Rohrer, 2018). En contraposición, otros estudios no han encontrado resultados positivos en la L1 (Macoun & Sweller, 2016; Austin & Sweller, 2014; Feyereisen, 2006). Asimismo, los gestos rítmicos muestran un aumento de la atención en estudios neurofisiológicos (Biau & Soto-Faraco, 2013; Dimitrova, Chu, Wang, Özyürek, & Hagoort, 2016). Asimismo, estudios recientes sobre los gestos referenciales han mostrado que los hablantes no-nativos se benefician más de los gestos, y los integran de una manera más intensa (con un efecto de N400 más fuerte) que los hablantes nativos (Dahl & Ludvigsen, 2014; Drijvers & Özyürek, 2018). Por otro lado, el impacto de los gestos rítmicos en L2 comparado en L1 no ha sido investigado en profundidad.

El objetivo general de esta tesis es de evaluar el rol de los gestos rítmicos (acompañados de prominencia prosódica) en el procesamiento del habla en la primera y segunda lengua. Se pretende estudiar cómo los gestos rítmicos ayudan a la memoria, a la comprensión, y a la memoria fonológica en hablantes nativos y no nativos de tres lenguas diferentes (inglés, francés, y catalán). La hipótesis general que guía la tesis y sirve como base a dos de los cuatro estudios es la idea que los hablantes no-nativos asimilan los gestos rítmicos de manera más intensa que los nativos, y por tanto los efectos de procesamiento deben de ser más visibles en una segunda lengua.

Esta investigación incluye cuatro estudios que responden a los tres objetivos principales delimitados, los cuales están relacionados con el rol de los gestos rítmicos en el
procesamiento del habla en una primera y una segunda lengua. El Estudio 1 tiene como objetivo clarificar la relación temporal y pragmática entre los gestos rítmicos y el discurso en inglés y en francés a través de un análisis de un corpus de 15 Ted Talks en inglés y 15 Ted Talks en francés, dos lenguas que codifican la prominencia prosódica de manera diferente (en inglés contiene un acentuación léxica donde predominan los inicios de la palabra, mientras que la acentuación en francés opera a un nivel de la frase, con un acento inicial no-obligatorio y un acento al final de la frase obligatorio). El Estudio 2 utiliza un paradigma de potenciales evocados (ERP) y analiza la modulación de dos componentes ERP (específicamente el componente P3a y el componente N400), para investigar los efectos de procesamiento de los gestos rítmicos en una primera y una segunda lengua (Dimitrova et al., 2016; Drijvers & Özyürek, 2018; Kutas & Hillyard, 1980). El Estudio 3 indagará si la exposición a los gestos rítmicos durante una tarea de dibujar una narración verbal (adaptado de Dahl & Ludvigsen, 2014) aportará una mejor memoria y comprensión del habla en los hablantes no-nativos, comparado con los hablantes nativos. El Estudio 4 investigará los efectos de la presencia de gestos rítmicos sobre la memoria fonológica, resultando en una mejora pronunciación en la lengua segunda. Los participantes realizarán una “shadowing task”, y las producciones serán evaluadas por los nativos en relación al nivel de acento, la inteligibilidad, y la fluidez. Resumiendo, los resultados de los cuatro estudios evaluarán de una manera más completa la influencia de los gestos rítmicos como marcadores multimodales y sus efectos potenciales en el procesamiento y aprendizaje de una segunda lengua.

Palabras clave: gestos rítmicos, gestos no referenciales, prosodia, ERPs o paradigma de potenciales evocados, aprendizaje de segunda lengua
Abstract

Co-speech gestures have been shown to aid various aspects of cognition and positively influence the learning environment (see Goldin-Meadow, 2011 for a summary). While referential gestures have been shown to increase listening comprehension (e.g., Morett, Gibbs, & MacWhinney, 2012), boost memory (e.g., Feyereisen, 1998; Tellier, 2005, 2006), and aide in second language (L2) learning (e.g., Kelly, McDevitt, & Esch, 2009; Tellier, 2008), less is known about the effects of beat gestures. Research on beat gestures, or non-referential gestures which typically align with prosodic prominence in speech has found contradictory results on their potential benefits. While some studies have shown that beat gestures can be useful in information recall in both first and second language (e.g., Igualada, Esteve-Gibert, & Prieto, 2017; Llanes-Coromina, Vilà-Giménez, Kushch, Borràs-Comes, & Prieto, 2018 for L1; Kushch, Igualada, & Prieto, 2018 for L2) as well as L2 pronunciation (e.g., Gluhareva & Prieto, 2017; Llanes-Coromina, Prieto, & Rohrer, 2018), other studies did not find positive effects of beat gestures in the L1 (Austin & Sweller, 2014; Feyereisen, 2006). Further, beat gestures have been shown to increase attention in neurophysiological studies (Biau & Soto-Faraco, 2013; Dimitrova, Chu, Wang, Özyürek, & Hagoort, 2016). Interestingly, recent studies on referential gestures have shown that non-native L2 listeners benefit more from gestures and process gestures in a stronger fashion (and with a stronger N400 effect) than natives when speech is clear (Dahl & Ludvigsen, 2014; Drijvers & Özyürek, 2018). The impact of beat gestures in L2 as compared to L1, however, has not been directly tested. The general aim of this thesis is to assess the effects of beat gestures (accompanied with prosodic prominence) on first and second language processing of speech, and particularly how beat gestures can be of help in areas such as recall, comprehension and phonological memory involving three languages (English, French, and Catalan). The general hypothesis behind two of the studies in this thesis is that non-native listeners process beat gestures more strongly than natives.

The thesis includes four separate studies that address three main goals related to the role of beat gestures in L1 and L2 speech processing. Study 1 has the goal to elucidate the temporal and pragmatic relationship between beat gestures and speech in French and English via a corpus analysis of a set of 15 Ted Talks in English and a set of 15 Ted Talks in French, two languages that structurally encode prosodic prominences in different ways (e.g., English has typically word-initial lexical stress marked by a pitch accent, while accentuation in French works on a phrasal level, with an optional phrase-initial and obligatory phrase-final accent).
Study 2 will use Event-Related Potentials (ERP) to investigate the modulation of two ERP components by the presence of beat gesture in native and non-native speech, and specifically the P3a component, which is a measure of attention to focused words with beat gestures (Dimitrova et al., 2016), and the N400 component, which is a measure of semantic integration and comprehension (Drijvers & Özyürek, 2016; Kutas & Hillyard, 1980). From a behavioral point of view, Study 3 will test whether exposure to beat gestures during a narrative-drawing task (adapted from Dahl & Ludvigsen 2014) will lead to better recall and comprehension scores for non-native listeners, as compared to native listeners. Study 4 will investigate the impact of beat gestures on phonological memory leading to second language pronunciation improvement. A shadowing task will be carried out and participant productions will be assessed by native raters for accentedness, comprehensibility, and fluency. All in all, the results of the four studies will assess in a comprehensive way the role of beat gestures as multimodal pragmatic markers and their potential effects on L2 speech processing and learning.

**Keywords:** beat gestures, non-referential gestures, prosody, Event-Related Potentials, second language learning
1. **Introduction**

1.1. **Object of Analysis**

The research in this thesis broadly examines the prosodic and pragmatic properties of beat gestures (or non-referential gestures that typically align with prosodic prominence in speech; see section 1.2.3. below) and their potential differential and beneficial role in L2 speech processing and learning. Drawing on recent results showing that referential gestures aid non-native speakers more than native speakers in comprehension and memory in a drawing task (Dahl & Ludvigsen, 2014, see also Drijvers & Ozyurek, 2018), this thesis will test the hypothesis that non-referential beat gestures are also potentially processed in a stronger way and are potentially more useful in non-native than in native speech. Two main studies will be conducted to test this hypothesis which will elicit both electrophysiological and behavioral data, namely (a) an ERP study will investigate the online processing of beat gestures and prosodic prominence in native and non-native participants when exposed to beat gesture prominence or prosodic prominence; (b) a behavioral study will make use of a narrative-drawing task to assess information recall and comprehension in native and non-native listeners when exposed to beat gesture prominence or prosodic prominence. The last study will use a shadowing task (i.e., an online and simultaneous repetition of L2 speech, either enacted or not) to investigate the influence of gestural/prosodic prominences on phonological awareness. The first study constitutes a baseline study which aims to examine the temporal and pragmatic properties of beat gestures in two languages that mark prosodic prominence in two different ways, e.g., English and French.

1.2. **Literature Review**

1.2.1. **Co-speech gestures**

Co-speech gesture refers to any bodily movements that are involved in the communication of meaning and co-occur with speech in both a temporal and semiotic fashion. These are distinguished from other types of spontaneous bodily movement that do not directly contribute to the meaning, such as grooming (e.g., McNeill, 1992). In an acclaimed book which has become a standard in gesture studies, McNeill (1992) claimed that co-speech gestures are integrated with speech in three complementary ways. They typically show semantic synchrony (i.e., meaning in gesture corresponds to meaning in speech), pragmatic synchrony (i.e., both speech and gesture serve the same pragmatic function), and temporal synchrony (i.e., the two events co-occur phonologically; see section 1.2.3. for more details).
In this thesis, we will follow the standard categorization of gestures proposed by McNeill (1992). He broadly categorized co-speech gestures into referential gestures (or gestures that make an explicit reference to what is conveyed in speech and are comprised of iconic, metaphoric, and deictic gestures) and non-referential gestures (or gestures that do not convey any clear semantic meaning in speech, also called beat gestures). Iconic gestures typically represent, in an easy-to-recognize manner, the object that is being described in speech. An example may be when describing a tree falling down; the speaker may lift their arm vertically and slowly move it to a horizontal position, clearly representing the tree falling. Metaphorical gestures occur when gesture represents some abstract notion. For example, a speaker calling someone crazy and pointing near their temples, moving their hand in a circular fashion. This gesture could be seen as presenting the metaphor of “scrambled brains” to represent mental disorder. Deictic gestures refer to any sort of pointing gesture, either to concrete or imaginary objects. Beat gestures (described in more detail in the next section) were described by McNeill (1992:15) as “a simple flick of the hand or fingers up and down; or back and forth…” These gestures are not referential and are devoid of semantic content. Instead, they are generally used to mark rhythm or information structure (i.e., focus). Figure 1 shows an example of each of the four types of gestures described by McNeill (1992).

It is particularly important here to note that McNeill (2006, p. 61) claims that these categories are not distinct boxes into which each gesture can be clearly defined. Instead, he considers them as dimensions, where any one gesture can carry differing levels of properties from each category.

Kendon (1980) described a hierarchical breakdown of gestural realization which has also been supported in McNeill (1992). The largest component of the hierarchy is known as the Gesture Unit, or G-unit. It is referred to as the period of time between successive rests of the articulators. The Gesture unit can be broken down into one or multiple concatenated Gesture Phrases (G-phrases). The G-phrase consists of even smaller phases: the preparation phase refers to the movement of the articulator from rest (or the preceding phase in the case of multiple G-phrases), the stroke phase contains the most effort and the “meaning” of the gesture, the apex is the point of maximum extension of the gesture, and the retraction or recovery phase refers to the return of the articulators to rest. Pre- and post-stroke holds may be present, where the hands do not move, consequently delaying the onset of the stroke or the retraction phases. Of all of the gesture phases, only the stroke is seen as obligatory in order to
FIGURE 1: Co-speech gesture types according to McNeill’s classification, examples taken from a Ted Talk by David Keith (2007) [bold indicates the lexical affiliate of the gesture]. The top-left image refers to an iconic gesture when the speaker utters “the particles migrate to over the poles” (8m 40s); the top-right composite image shows a beat gesture when the speaker says “Really we’re doing this” (1m 55s); the bottom-left composite image shows a metaphoric gesture when the speaker says “you do some geoengineering for a little while” with the hand moving away from the speaker, representing time moving forward (13m 26s); the bottom-right image shows a deictic gesture when the speaker says “That impacts the middle of the planet where we live” (8m 47s).

FIGURE 2: The different gesture phases of a deictic gesture from Keith (2007) at 8m 47s. The movement of the preparation phase is captured between the first two images (movement from the offset of previous gesture to the beginning of the stroke). The middle two images show the movement of the stroke, with the apex shown in the third image. The movement of the retraction phase is captured in the two images to the right, where the hand moves from the apex to rest or to start another preparation phase for another gesture.
constitute a G-phrase. Figure 2 illustrates the different phases in the execution of the deictic gesture shown in Figure 1.

Since the early 1990’s, research into the temporal and semantic properties of co-speech gestures has exploded, so much so that it is now recognized under the name “gesture studies”. A large amount of research has gone into the temporal properties and cognitive benefits of the use of co-speech gestures. The following two subsections will assess studies that delve into topics particularly regarding (a) the prosodic features of co-speech gestures, as well as (b) their pragmatic contribution in discourse.

1.2.2. Prosodic features of co-speech gestures

It has been noted early on in gesture research that gestures tend to show a temporal alignment with phonology. For example, Condon & Ogston (1967) noted that a speaker’s changes in body motion tend to co-occur with the articulated segmental organization of their speech, particularly at the level of what they called the etic syllable. Etic syllables are described as pulses, where one etic syllable equals one chest pulse (Pike, 1943, p.53, Abercrombie 1967, p. 35, as cited in Pike, 1984). This idea of co-occurrence between gesture and speech was termed self-synchrony. To highlight this synchrony, the authors give an example from their corpus where a speaker says the word “around”, showing that the movements of her head, eyes, mouth, and trunk all coincide with the tripartite segmentation, or etic syllabic pulses, of the word (a . rou . nd). This co-speech gesture synchrony with speech was consistent across two full utterances by the speaker, lasting 5.5 seconds. The authors claimed that this synchrony falls within a wider organization of behavior, as certain movements may be sustained across numerous etic syllabic groupings (Condon 1976, as cited in Kendon 1980).

This early work led Adam Kendon to describe the temporal relationship between phonology and co-speech gestures. His 1972 study (as cited in Kendon, 1980) analyzed an extended utterance by one speaker that lasted 2 minutes. The speech stream was segmented into prosodic groupings from the tone level to the entire turn level (see p. 210 for details). Kendon found that each level in the speech hierarchy was distinguished also by differences in bodily movement. Specifically he found that body posture changes from Discourse to Discourse (i.e., between speaker turns), arm use varied between locution clusters (“paragraphs” of speech), going from right arm to left arm, to both, etc. Within the clusters, each locution group contrasted with different head movement patterns. Individual, distinctive
movements could thus be seen occurring during each tone group (e.g., an arm extension, rotation, or retraction, etc.). Kendon goes further to include the relationship between the gesture stroke and the nucleus of the tone unit (i.e., the pitch-accented syllable). Analyses showed that in his corpus, the stroke portion of the gesture is always completed either before the pitch-accented syllable, or immediately at its onset.

These findings lead McNeill (1992: 26) to propose the phonological synchrony rule, which states that the stroke must either precede – or end – at the phonological peak syllable of speech. If this synchrony is threatened, the hands will slow, delaying the stroke and thus affording that the synchrony is maintained (see also Rusiewicz et al., 2014). Nobe (1996, 2000, as cited in Loehr 2007) later added a level of granularity to McNeill’s phonological synchrony rule. He defined acoustic peaks as either peaks in fundamental frequency or intensity. He claims that strokes co-occur with the later of the two peaks, when the two peaks do not co-occur themselves.

Since Kendon and McNeill’s early work, a number of studies have investigated the temporal relationship between prosody and gesture, trying to pin down which properties of speech act as anchors for gesture movements. The following paragraphs group these studies into two broad categories, first looking at studies that examine the relationship between rhythm and gestural phases, followed by studies that investigate the relationship between prosodic prominence and gestural prominence.

McNeill (1992: 244) investigated the rhythmicity of gesture movements by looking at the periodicity of gesture strokes and found that they occur at more-or-less equal intervals of one or two seconds, depending on the speaker. From this, Tuite (1993) suggested that there is a rhythmic pulse that underlies both speech production as well as bodily movements. This pulse is manifested as a tonic nucleus (i.e., the most prominent syllable) in speech, and as a gesture stroke in the body.

On the other hand, many studies have investigated the temporal relationship between prosodic prominence (instantiated through pitch accentuation) and gesture prominence and specifically whether the prominent phases of gestures (e.g., strokes and apexes) act as anchoring points for gesture production. Loehr (2007) examined four short clips of American English native speakers during a conversation (a total of 164 seconds were analyzed). He

\[1\] It should be mentioned that most of the studies in the following paragraphs either do not take into account the different gesture types, or do not follow McNeill’s (1992) classification. Gesture type will be indicated accordingly.
found a complex rhythmic synchrony among the apexes of manual and head gestures, and speech. He showed that the different articulators produced peaks of movement (apexes for gesture) and pitch accents at individual regular tempos, which would sometimes align, and sometimes they would not. When investigating the relationship between gesture apexes and pitch accentuation, Jannedy & Mendoza-Denton (2005) found that in a spontaneous discourse filmed at a town hall, 95.7% of an American English speaker’s gesture apexes occurred along with a pitch accent. Conversely, only 69.4% of her pitch accents were marked with a gesture apex. Similar findings were reported by Loehr (2012). In the latter study, the author used the same corpus from his 2007 study, and he claims that pitch accents and gesture apexes “co-occur repeatedly” (p. 81). Upon closer analysis of the distance (in milliseconds) between gesture apex and pitch accent, he found that their distribution is centered very close to zero, and that the overall average is +17 milliseconds. Yasinnick, Renwick, & Shattuck-Hufnagel (2004) looked at the co-occurrence of stressed syllables and gestural hits. These gestural hits are a distinguishing characteristic of a gesture type that they call discrete gestures. Hits are defined as “an abrupt stop or pause in movement, which breaks the flow of the gesture during which it occurs. Hits appear as bouncing, jerky movements, changes in the direction of movement, or as complete stops in movement (p. 98).” In contrast to discrete gestures, continuous gestures do not contain hits, as they are gestures which show no clear, sharp change in movement. Continuous gestures were not included in their analysis. In a sample of 7.5 minutes of an academic lecture by an American English speaker, they found a total of 130 polysyllabic words that co-occurred with a gestural hit. In 90% of these words (117), the word also contained a pitch accent. This rate is much lower for monosyllabic words, where only 65% of the 116 hit-aligned words were also pitch accented. This rate is much lower for monosyllabic words, where only 65% of the 116 hit-aligned words were also pitch accented. It should be mentioned that of the hit-aligned words that were not pitch accented, most were within 100 milliseconds of a pitch accent. Other studies have confirmed that gesture apexes and gesture strokes tend to align with pitch accentuation in languages other than English. Esposito et al. (2007) analyzed 4 minutes of speech by a male Italian speaker, and 4.5 minutes by a female speaker and found that gestural hits overlapped with pitch accents 78% and 84% of the time, respectively. Esteve-Gibert & Prieto (2013) investigated deictic gestures produced by native Catalan-speaking adults. Unlike previous studies which looked at overlap between two phenomena, they compared the alignment of two discrete moments in time: the gesture apex and the intonation peak. They found that the timing of the two events was significantly correlated. Further, they found that upcoming low-boundary tones affect the timing of both intonation
and gesture movements. Esteve-Gibert et al. (2014), head gestures were investigated in semi-spontaneous speech by native Catalan speakers. Much like the previous study, they found that the apexes of head gestures were synchronized with accented syllables, and that upcoming phrase boundary affect the timing of the start, apex, and end of the head gesture.

Despite the overwhelming evidence of synchronization between prosodic phenomena and gestures, there are a small number of studies that have not found such evidence. De Ruiter (1998, study 1, as cited by Esteve-Gibert & Prieto, 2013) analyzed the temporal relationship of pointing gestures and lexically stressed syllables in native Dutch speakers carrying out an image-pointing task. His results showed that gesture apex was not affected by stress position. He even found gesture apexes that did not occur during stressed syllables. Rusiewicz (2010, as cited in Esteve-Gibert & Prieto, 2013) reflected de Ruiter’s findings for English, where metrical structure did not influence gesture timing. Instead, gesture apexes were aligned with the onset of the target word.

To summarize, there has been a large amount of evidence that different kinds of gesture types (e.g., iconics, deictics, beats) tend to tightly temporally co-occur with pitch accents across languages. The previously cited studies have mostly focused on English as the language of study, though there are a few exceptions (e.g., Italian, Catalan, and Dutch). Further, the studies are difficult to compare, as they investigate different aspects of both gesture (hit/stroke/apex) and speech (stressed syllable/pitch accent, lexical affiliate) with very small sample sizes. In any case, it is widely held in the field of gesture research that prosody and gesture are temporally coordinated.

1.2.3. Properties of beat gestures

Beat gestures, the type of gesture investigated in the current work, have also been called “batons” (e.g., Efron, 1941; Eckman & Freisen, 1969) and are generally regarded as gestures that merely mark rhythm, like the conductor of an orchestra. McNeill specifically describes beat gestures as appearing to be beating musical time, with forms relatively the same regardless of content in speech, consisting of “a simple flick of the hand or fingers up and down, or back and forth” (McNeill 1992:15), which is distinguished from other gestures in that they only show a bi-phasic movement. Other authors such as Shattuck-Hufnagel et al. (2016) considered beats as non-referential gestures which share many of the same pragmatic roles with prosody, and in turn propose that they are “prosodic gestures”. While McNeill does claim that the semiotic value of beat gestures lies in the fact that they convey discourse-
pragmatic content (1992:15) (see section 1.2.3.2. below), many researchers have focused exclusively on their rhythmic component. In our view, some recent work has used an oversimplified definition of beat gestures, calling them “rhythmic, non-meaningful hand movements” (e.g., Dimitrova et al., 2016:1255).

In this thesis we adopt a comprehensive definition of beat gestures (see the recent proposal by Prieto et al., 2018) which holds beat gestures refer to any non-referential gesture which is often, though not always, prosodically integrated with speech much like other gesture types. As proposed by Shattuck-Hufnagel et al. (2016) and Prieto et al (2018), the structure of beat gesture phases can be more complex than McNeill’s description of a very simple “out and back” movement. Indeed, Shattuck-Hufnagel & Ren (2018) analyzed a corpus of academic speech and found that of 1263 unambiguously non-referential gestures, only 23 (1.7%) showed the typical handshape described by McNeill (1992). Further, beat gestures can contribute to a variety of important pragmatic and discourse meanings. The next subsections will review some of the prosodic and pragmatic properties of beat gestures.

1.2.3.1. Prosodic features of beats

As previously mentioned, many studies have examined the temporal relationship between prosodic prominence and co-speech gesture. While most of the abovementioned studies do not distinguish between gesture types, few have specifically focused on beat gestures.

One study by McClave (1994) found that the strokes of beat gestures tend to coincide with nuclear pitch accents in English, but not always. Her data suggests that multiple beat gestures are anchored in a tone nucleus, but that they are then generated rhythmically outward. That is, they follow their own rhythm, so that beat strokes occur at regular intervals independent of prosody, thus some strokes may fall on unstressed syllables. Further, this rhythm can even be generated before the tonic nucleus, suggesting that the rhythm is established before the utterance even begins. Loehr’s (2007) study added support to McClave’s findings, claiming that his data showed similar patterns to beat productions as hers, though he offered no further analysis particularly regarding beat gestures. In his thesis, Loehr (2004) commented on the beat gestures in a short 45-second clip of pilot data: He found a total of 8 beat gestures, where 6 of the 8 beats co-occur with an H* tone, while one co-occurs with a L+H* tone and one with a L* tone. In his larger corpus, he found that 43 of
the 66 beat gestures co-occurred within 275ms of a pitch-accent. Further analyses that he reported regarding their proximity to pitch accents did not distinguish gesture type.

Leonard & Cummins (2010) also specifically examined beat gestures and their anchoring points with English speech. They asked one subject to read three fables two times each, where each fable contained 3 naturally stressed words where the reader was instructed to execute a beat gesture. They found that the closest speech landmark to the apex was the peak of the pitch accent in the stressed syllable. The authors concede that while these results are supported by findings from previous studies (e.g., Loehr, 2004 regarding proximity to pitch accents), the methodology was highly experimental, and may not accurately reflect what happens in spontaneous speech.

A more recent study by Shattuck-Hufnagel & Ren (2018) examined the relationship between non-referential gestures and pitch accents, as well as their groupings at higher level prosodic constituents in a corpus of academic-style speech. Gestures with clearly defined strokes were annotated and determined to be either referential or non-referential. Sequences of gestures that appeared to be performed as a group based on handedness and trajectory were labelled at Perceived Gesture Groupings (PGGs). They chose this term instead of Kendon’s Gesture Unit (see section 1.2.1. above) and Kendon’s proposed term is said to end with a retraction phase, yet the authors felt that the PGGs may not have this characteristic. Upon analysis for the relationship between gesture strokes and pitch accentuation, they found that non-referential gesture strokes overlapped with a pitch-accented syllable 83.13% of the time. Similar results were found for the strokes of referential gestures in their corpus (82.85% overlap). Regarding their groupings, they found that only half of the PGGs were contained within a Full Intonational Phrase. By carrying out Extended Rapid Prosodic Transcription (E-RPT, a “crowd-sourced” prosodic transcription where inexperienced prosodic labelers demarcate higher level prosodic boundaries), they found 18 higher-level constituents, where PGG hand form and trajectory differ from one higher level constituent to another.

All in all, it can be said that compared to referential gestures much less is known about the temporal association of beat gestures with speech in natural discourse, particularly on how they behave cross-linguistically in languages where phrasal pitch accents have different positions and tone types. In general, the abovementioned studies on the relationship between gesture and prosodic prominence generally use few speakers and very short amounts of discourse. A sample size large enough is lacking to produce reliable results. Secondly, the languages of study tend to focus on English or other predominantly stress-timed languages.
Little is known about how beat gestures function in a language with a fixed phrasal-level encoding of prominence. **Study 1** of this thesis will analyze the temporal behavior of beat gestures in relation to pitch accentuation in two languages where prosodic prominence is encoded differently (*e.g.*, English and French; see Figure 3). English shows typically word-initial stress marked by a rising pitch accent L+H* (or also H*) where placement is determined by the individual lexical item it accompanies, and relative intensity to other stresses within the intonational phrase generally reflect information structure (*i.e.*, focus marking). Accentuation in French, on the other hand, is not distinctive on the word level, but rather works on a phrasal level. The phrase consists of an optional phrase-initial rising pitch excursion and an obligatory phrase-final pitch accent, H* (realized as L% when the phrase occurs at the end of an intonational phrase). The use of the initial and final accent is tied more to the building up of a rhythmic pattern (some researchers have even named this phrase the *rhythmic group*; see Hirst & Di Cristo, 1993) than conveying information structure, as focus is often marked morpho-syntactically rather than prosodically. To our knowledge the only study on beat gestures in the French language is Roustan & Dohen's (2010), which showed that when a word is prosodically focused and accompanied by a pointing gesture, a beat gesture, or a non-communicative movement (*e.g.*, pushing a button), apexes “occur[ed] within or close to the focused element” (p. 4) and that the co-occurrence was closest with pointing gestures.

In essence, **Study 1** will serve as a baseline study to understand the relationship between beat gestures and speech in two different languages (French and English) which will be reprised in a language-learning context in the latter two studies within this work. This study is a part of a larger collaboration with the Natural Language Processing (TALN) research group at the Department of Information and Communication Technologies at UPF (Drs. Mireia Farrús & Leo Wanner), which aims to develop an automatic annotation for gesture phases and will investigate the temporal relationship between gesture phases (and specifically strokes and apexes) and prosodic structure (*e.g.*, different levels of pitch accentuation) in a large scale corpus of English Ted talks.

---

2 For example, “**C’est moi** qui est allé au supermarché.” (*It is I* who went to the supermarket.) is more common than emphasizing [**Je**] in “**Je** suis allé au supermarché.” (*I* went to the supermarket). Bold indicates focused entity.
FIGURE 3: The intonation contour of an example sentence in English (top panel) and French (bottom panel). In the sentence “I went to the supermarket the other day”, there is a nuclear pitch accent ($H^*$) on the first syllable [su] which is a lexical feature of this word (i.e., this syllable is always stressed in the word “supermarket”). In the French sentence, we can notice the phrasing [J’suis allé] [au supermarché] [l’autre jour] where a pitch accent occurs on the last syllable of each phrase and is not a fixed attribute of the word on which it falls.

1.2.3.2. Pragmatic functions of beats

A few studies have commented on the pragmatic role of beat gestures as defined by McNeill (1992). As previously mentioned, McNeill describes their key roles in conveying discourse–pragmatic functions. Indeed the author says that their semiotic value is their relationship to the overarching discourse. First, beat gestures are said to work much like highlighters, working together with prosodic prominence to emphasize certain words, or to mark focused elements in speech (e.g., McNeill, 1992; Loehr, 2007, 2012; Roustan & Dohen, 2010, Krahmer & Swerts, 2007, among many others). 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. Beat gestures have been defined as rhythmic hand and arm movements.
that are typically associated with prominent prosodic positions in speech (i.e., pitch accents) (e.g., Krahmer & Swerts, 2007; McNeill, 1992). Interestingly, Shattuck-Hufnagel et al. (2016) highlighted the fact that (non-referential) beat gestures share many of the same linguistic roles with prosody, like focus marking.

Authors have also highlighted the role of beats in narrative structure and discourse marking. McNeill (1992) stated that beat gestures may mark the addition of new information such as new characters or themes in a narrative. He says “beats mark information that does not advance the plot line but provides the structure within which the plot line unfolds” (1992:15). They may indicate movement between different narrative levels (e.g., narrative, meta-narrative, and para-narrative levels). Additionally, beat gestures can signal frames within the narrative level in which the speaker participates (such as in lexical repairs or adding new information about an already introduced topic). This can signal a change in distance between the speaker and his narration. Thus beats are said to have a meta-pragmatic function, occurring both in narrative and extra-narrative contexts.

Much of the research on the pragmatic functions of gesture does not rely on McNeill’s classification. Indeed, Kendon associated beat gestures with a larger class of “pragmatic gestures”, which “relates to features of an utterance’s meaning that are not a part of its referential meaning or propositions content” (Kendon 2004:158). There are in contrast with substantive gestures which convey utterance content either literally or metaphorically (i.e., referential gestures). In the literature, these pragmatic gestures are often referred to by their hand form or movement (e.g., beats, flips, open-palm gestures, ring/precision grip gestures, among many others). It is important to note here that for this thesis, the proposal by Prieto et al. (2018) will be followed. The proposal subsumed all of the abovementioned gestures under the term “beat gesture”, regardless of hand form or function, as long as they do not have any referential value. That being said, the original “hand form” terminology scheme will be used when reviewing the literature, but it should be remembered that for the present work, we will refer to all of these gestures as beat gestures with particular hand forms.

Kendon (2004) breaks down the various functions of pragmatic gestures into four broad categories according to their functions: performative, modal, parsing, or interpersonal. Gestures that show the type of the speech act (i.e., whether the utterance is expressing a question, a refusal, etc.) is said to be performative in function. The modal function refers to a gesture that indicates how the listener should interpret the utterance (e.g., the speaker’s epistemic stance- their certainty with regards to the utterance). For example, Kendon (2004)
attributes a modal function to hand flip gestures, in that they express “the speaker’s inability or unwillingness to act (...) to offer any suggestions or solutions, to provide meaning or an interpretation of something – to take a stance” (quoted in Ferré, 2011: 3). Figure 4 below shows other examples taken from Prieto et al. (2018), illustrating two beat gestures with different forms and differences in epistemic stance.

FIGURE 4: In the left panel, we see an open palm beat gesture showing low epistemic commitment and detachment (“There are going to be some benefits to climate change -oh yes I think it’s bad”). In the right panel, we see a precision grip beat gesture which shows high epistemic commitment (“that report that landed on Johnson’s desk”).

The parsing function highlights the structure of the spoken discourse without referring to the content within the message. This function can be seen when beats are used to show contrastive focus. Kendon (1995) describes two types of “precision grip” gestures: the “Finger Bunch” gesture which distinguishes the topic from the comment at the utterance level, and the ring gesture (also called “index finger to thumb”, or IFT as per Lempert, 2011) which emphasizes a particular discourse unit within a larger context. Lempert’s (2011) study analyzed the IFT gesture by Barack Obama during political debates and found that indeed they may serve a gestural contrasting function, separating his political position from those of his opponent. He concludes that this is a gestural highlighter of focus.

Regarding these categories, it has been noted by Streeck (2008: 259, as cited in Ferré, 2011: 2) that one single gesture can fulfill multiple pragmatic functions. Ferré (2011) gave the example of a (McNeillian) beat gesture, which could be interpreted as parsing (as it highlights important elements) and modal (showing the speaker’s judgement of importance). Streeck

3 Images taken from a Ted Talk by D. A. Keith (2007) at 3m 19s and 5m 50s respectively, bold and underlined syllables indicating beat gesture placement.
(2008) also showed examples where changes in gestural form may indicate the beginning of a new speech act or discourse segment, which is accompanied by changes in the pragmatic function of speech. Thus the changing of hand shapes across an utterance visually shows the overall pragmatic structure of that utterance. Along the same lines, Bavelas et al. (1992) discussed interactive conversational gestures, in which beat gestures are subsumed. Again, these gestures contribute to “the nature of the dialogue itself, rather than the specific topic of discourse” (p. 476), and changes in form may mark changes in illocutionary force of the discourse and/or speaker stance.

Departing from Kendon’s codification, Ferré (2011) proposed analyzing the pragmatic functions of non-referential open-palm gestures in French into three categories. The first category is a prosodic function, where gestures accompany emphatic stress in speech. She described these gestures as prosodic units, indicating that prosody is made up of both vocal and gestural components. The second category is a discursive function, which occurs when gestures show the discourse structure (e.g., to mark the beginning or the end of a discourse segment). The final category is the modal function. These are non-referential gestures that co-occur with grammatical modality and serve to reinforce it. In this study, Ferré specifically differentiated the open-palm gestures as either beat gestures or hand flips. While she mentioned that beat gestures may have a variety of hand configurations, she considered the basic hand shape of a beat to be an open palmed vertical hand movement (2011: 10).

All in all, even though the pragmatic role of beat gestures has been typically associated with rhythmic structure, discourse structure, and aspects of language at a meta-narrative level, no work until now has been carried out with a systematically labelled large database. The second part of Study 1 will be aimed at determining the pragmatic functions of beat gestures under this broadened definition, in a cross-linguistic fashion. This study will take into account the interplay between articulators, hand forms, and pragmatic meaning. In the next subsections, we review the evidence for the positive cognitive effects of the use of referential and non-referential gestures in the first and in the second language.

1.2.4. Cognitive benefits of referential gestures in the first and the second language: behavioral and neurophysiological evidence

A vast amount of research in gesture studies is dedicated to the investigation of the cognitive benefits of gesture, namely regarding memorization and comprehension. The next
subsections will review this literature from two perspectives, looking at behavioral evidence and neurological evidence.

1.2.4.1. Behavioral evidence

Numerous studies have shown that encoding information with referential gestures boosts memory in the first language. Riseborough (1981) found that verbs that were presented with iconic gestures were recalled the most in both a word list recall task, and a narrative cloze task. Regarding the recall of sentences, young adults and children were also shown to remember sentences produced with iconic gestures more than sentences presented with only acoustic cues or no cues at all (Thompson 1995; Thompson, Driscoll & Markson 1998). These results echoed Cohen & Otterbein’s (1992) findings, where they exposed participants to sentences co-produced with either pantomimic (i.e., referential) gestures or meaningless gestures. In a free recall task, sentences with pantomimic gestures were better remembered. Feyereisen (1998) also found that when exposing listeners to sentences with iconic gestures that match the content in the sentence, or iconic gestures that did not match the content in the sentence (mismatching), sentences with matching gestures were remembered the most. He claimed that this finding is thus more than just the peculiarity of the gesture co-occurring with the sentence, but in fact the gesture is conveying important information related to the sentence in the visual modality, which is added to the verbal information.

Similar benefits of iconic gestures for short-term memory were reported in children. When studying the effects of referential (metaphoric) gestures on long-term memory in children within the context of a mathematics lesson, Cook, Mitchell, & Goldin Meadow (2008) instructed students on how to solve addition math problem (e.g., 4 + 3 + 6 = __ + 6). In a pre-instruction phase, the teacher quickly explained how to solve the problem in three conditions, only speech (“I want to make this side equal to that side”), only gesture (instructor moved his left hand under the left side of the equation, paused, and moved his right hand under the right side of the equation), and both speech and gesture. The students were then asked to repeat after the instructor (and/or to perform the same gesture if the teacher had performed them). They were then fully instructed how to solve the problem (with the instructor using both gesture and speech). Then, children were asked to repeat the action from the pre-instruction phase (i.e., speech, gesture, or both). Then students were given a post-test immediately after instruction, as well as a follow-up test four weeks after instruction. The investigators found that immediately after instruction the students solved a comparable
number of problems correctly. Using a regression model with the difference between pre-test and post-test scores, they then calculated estimates of their scores for the follow-up tests at four weeks. After the follow-up tests, they found that students who were told to gesture during their instruction maintained 58% of their immediate post-test gains, while students in the speech-only condition maintained only 33% of their gains.

Another study by Tellier (2005) directly tested the effect of visual modality on memory by presenting 3 videos of word lists to 32 French children, individually. Each video presented 10 words, and showed (a) a person only pronouncing the words, (b) a person producing the words with an iconic gesture, or (c) an image that represented that word. She found that children performed a free recall task better in both of the visual conditions (gesture or image). In a second study (Tellier 2007) she investigated the role of producing gesture on memory. Following a similar protocol to her previous study, she divided the children up into three groups: one saw no gesture and repeated the words they heard, one that saw gesture and repeated the words, and one that repeated the words while reproducing their corresponding gesture. The children who reproduced the gestures did significantly better on the free recall task than the other two groups. This effect has also been shown to go beyond just the memorization of words or sentences in one’s first language.

A number of studies have explored the influence of referential gestures on memory particularly regarding second language as well (e.g., Allen, 1995; Tellier, 2008; Kelly, McDevitt & Esch, 2009; Macedonia, Müller & Friederici, 2011; Morett, 2014). Kelly et al., (2009) investigated the perception of referential gestures on word learning by native English speakers with no instruction in Japanese. They were exposed to training videos which introduced Japanese words with an iconic gesture that corresponded with the lexical affiliate (matching gesture), an iconic gesture that did not correspond (mismatching), speech alone where the word was presented once, or speech alone which was repeated twice. The mismatching gesture was presented to test whether hand movement alone produced the effect, or if meaning in gesture must be congruent to meaning in speech for learning to occur. The reasoning for the repeated speech condition was so that there was the same amount of information as the speech with matching gesture condition. The participants took part in a free recall task and a recognition task immediately following the training, as well as 2 days and one week later. The results of the memory tasks showed that participants remembered the most words in the condition with speech and matching gesture, followed by the repeated speech condition, the speech alone condition, and the mismatching gesture condition.
respectively. Similar results were found by Macedonia et al. (2011) when teaching German-speaking participants Italian non-words. They were trained by watching videos presenting the new vocabulary items with either iconic or meaningless gestures. Meaningless gesture here refers to physical movement that was did not contain any iconic or symbolic meaning, such as touching one’s head or knee. Again, both short-term and long-term memory tests revealed that participants correctly remembered the words more often when presented with iconic gestures.

Allen (1995) investigated the impact of emblematic gestures (or gestures that have been conventionalized and are culture specific, such as the peace sign) when teaching English students of French common French expressions. The researcher used 50 gestures that were emblems in French culture, while not so in the subjects’ native culture. Subjects were divided into three groups: (a) the experimental group, where emblems were shown and reproduced by the subjects during the teaching of emblems, (b) the control group that was never exposed to emblems, and (c) a comparison group that were shown emblems only during the posttest. The expressions were taught 10 at a time in five 30-minute sessions over a 10-week period. Each session consisted of a pre-test, treatment (training), and posttest to test short-term memory. A final recall test was given which contained 2 expressions from each of the five sessions, which served to test long term memory. She found that students who were presented with emblems recalled the meaning of more expressions than those who were not exposed to emblems, and that the group also remembered significantly more expressions after the 10-week period of instruction.

In another study by Tellier (2008) which distinguished the effects of image perception and gesture re-enactment, twenty French children with no prior knowledge of English were divided up into two groups: a picture only group, and a gesture group. Over four sessions, the children were taught a total of 8 English words. They were further asked to repeat the word (along with the gesture in the gesture group). As expected, the children in the gesture group performed significantly better, confirming that re-enacting gestures leads to better memorization than visual input in the form of images alone. Similar results were found in Morett (2014) when investigating native English adults encoding (storage in memory) and recall (retrieval from memory) of 20 novel Hungarian words in an interactive learning task. Participants were put in pairs and one person was assigned the “explainer” and the other was assigned the “learner”. The explainer saw 10 of the Hungarian words with representational (i.e., iconic/metaphoric) gesture and 10 without, and had to teach the novel vocabulary words to the learner. During this teaching phase, both participants engaged in dialogue and the
session was filmed for gesture production analysis immediately following the teaching phase, both participants were instructed to take a memory recall test. The impact of gesture was determined by examining the relationship between gesture production and target word repetitions during the learning phase and subsequently comparing this relationship to recall scores. She found that the use of representational gesture by the Explainers facilitated their encoding of the new vocabulary, and led to the Learners performing better in the vocabulary recall task. Her results describing the effects of beat gestures are further described in section 1.2.5.2. below.

Regarding comprehension, a number of studies have found that referential gestures enhance speech comprehension and are integrated with speech. In a meta-analysis of over 60 studies regarding the impact of referential gesture on comprehension, Hostetter (2001) found 6 different manners in which gesture may improve speech. First, gestures may improve communication as they particularly lend themselves to convey spatial information much more effectively than speech. Indeed, an early study by Graham & Argyle (1975) asked participants to describe shapes with varying levels of codability (*i.e.*, the ease of describing in the verbal modality) to listeners, who then recreated the shapes as they were described. The researchers found that gesture-accompanied speech improved the accuracy of the shapes produced by the listeners, and that this effect was greatest for the shapes with low-codability. Another study by Driskell & Radtke (2003), showed that when listeners watched speakers describe spatial words (such as *under* or *square*) with or without accompanying gestures, they were able to correctly identify the word being described quicker when gesture was present. This effect disappeared when speakers described non-spatial words, such as *color* or *warm*.

Another way in which gesture may facilitate comprehension is by conveying information that is not present in speech. Indeed, such gestures that give supplementary information to speech are still integrated by listeners. Beattie and Shovelton (1999) recorded participants describing cartoon narratives with spontaneously produced iconic gestures. Many of their gestures did indeed contain supplementary information to speech. They offer the example of someone says “The chairs were around the table” while producing a pointing gesture moving in a large circle (implying that the table was large and circular). The researchers then took the videos and showed them to another group of participants in two conditions: audio-visual and audio only. After watching the videos or hearing the audio, participants were interviewed, where 2 yes-no questions pertained to each item or action described in the narrative, particularly pertaining to the supplementary information in gesture.
Recuperating the example gesture above, the questions were “Is the table square?” to which the correct response would be “No”, and “Is the table large?” to which the correct response would be “Yes.” As one would expect, the audio-visual group were able to correctly answer more questions than the audio-only group.

Gestures may also indirectly affect comprehension. Producing gestures has been shown to aide speakers in producing more fluent speech with fewer pauses (Rauscher, Krauss & Chen 1996). Thus gestures may benefit the speaker by allowing them to produce more fluent speech, in turn facilitating the comprehension by the listener. Gesture may also positively affect comprehension through creating a positive rapport between interlocutors, as listeners tend to like speakers who produce gestures more than speakers who don’t (Kelly & Goldsmith, 2004). Further, listeners who see speakers gesture may be more engaged and pay more attention to speakers (to be described further in section 1.2.4.2.).

It is also considered that when gestures show the same information as speech that their increased redundancy may help in language comprehension, in that they provide additional cues to meaning that the listener can exploit. A priming study by Kelly, Özyürek & Maris (2010) exposed participants to action primes (e.g., someone chopping vegetables) and matching/mismatching multimodal targets (someone saying “chop” with a chopping gesture vs. someone saying “chop” with a twisting gesture). They showed that matching gesture targets were matched with primes faster than mismatching gesture primes and advocated for a strong integration hypothesis between gesture and speech. This effect of redundancy may be particularly useful when speech is difficult to understand, such as for second language learners (Sueyoshi & Hardison, 2005; Dahl & Ludvigsen, 2014, described more in detail in section 1.2.4.3.) or for listeners with lower verbal skills. McNeil et al. (2000), for example, showed that referential gestures were more beneficial to preschoolers than to kindergarteners. Fifteen preschoolers and 14 kindergarteners were individually given 2 tasks where in each task, the experimenter would instruct the child to select certain blocks with animal designs (task 1) or arrange blocks with different images on each face in a certain way (task 2). The experimenter’s instructions used sentences that contained a compound relative clause, so that they were considered difficult for the preschoolers but easy for the older kindergarteners. The experimenter’s speech was accompanied by either reinforcing gesture, no gesture, or conflicting gesture. Preschool children performed the task much better when reinforcing gesture was present, while kindergarteners did not show a difference between the two conditions. To ensure that this was due to the complexity of the sentence, a second experiment
was carried out with the preschoolers with the same task and design, only with simpler sentences. In this case, the reinforcing gestures did not influence performance. Thus they conclude that reinforcing gestures do indeed help when the speech is at a higher level of complexity.

All in all, this section has shown that referential gestures are a useful tool to aide in comprehension and memorization by both native speakers and non-native speakers. Gesture’s impact on comprehension is dependent on a number of factors. The factor described regarding the ease of comprehension depending on the speech complexity is the most pertinent to this thesis. An early study by Rogers (1978) showed that gestures become increasingly useful for comprehension as the speech-to-noise ratio decreases (i.e., as noise is increased). Similar results were found by Riseborough (1981), where the narratives were presented in four conditions (Audio only, Audio-visual no gesture, audio-visual with vague movement, and audio-visual with iconic gesture) with increasing levels of noise. As noise increased, comprehension of the narrative decreased except for participants who were exposed to the meaningful gesture condition. These findings have been supported in more recent neurophysiological studies (see section 1.2.4.2. below). Further, there is a parallel with speech comprehension in second language, where studies seem to confirm that gestures may have a stronger influence in the second language than in the first (e.g., Dahl & Ludvigsen, 2014, to be described in section 1.2.4.3.).

1.2.4.2. Neurophysiological evidence

Neurophysiological studies are complementary to behavioral studies, as they examine the brain’s online reaction to stimuli. A number of neurophysiological and imaging studies have investigated the interaction between speech and referential gesture in language processing. Most of the studies cited below use the Event-Related Potential (ERP) paradigm which refers to the electrophysiological brain response to stimuli. ERPs are generally measured via electroencephalography (EEG) where the resulting ERP waveforms are made up of positive and negative voltage deflections. These deflections are related to components that may reflect different neurological processes, such as attention or comprehension. Generally, the various components are named by their deflection (Negative or Positive) and the post-stimulus latency in milliseconds (e.g., N400 is a negative-going deflection around 400ms post-stimulus). The studies below generally investigate modulation in the N400 component, as has been shown to be a marker of semantic comprehension (Kutas & Hillyard, 1984).
Specifically, the larger the N400 component, the more processing is required for comprehension. If the N400 is smaller, semantic processing is said to be less demanding (i.e., easier to comprehend).

One study by Wu & Coulson (2007) found that iconic gestures that were presented (without speech) after cartoon images that were incongruent with that image elicited an N400 component compared to congruent gestures. Dealing with gestures within the context of speech, Kelly, Kravitz, & Hopkins (2004) exposed participants to speech accompanied by matching, complementary, and mismatching iconic gesture. They found that the N400 was reduced when gestures were matching. Following a similar protocol, Kelly, Creigh & Bartolotti (2010) had participants watch similar videos, but asked them to determine if the speaker was female or male. This was done in order to determine if the integration of gesture is automatic, as the participants were not specifically focusing on the content of speech or gesture. They found that indeed, even when not focusing on the gestures, participants produced a smaller N400 component when gesture and speech were congruent.

When investigating the idea that gestures can facilitate comprehension after exposure, Holle & Gunter (2007) presented participants with sentences where the first clause contained a semantically ambiguous homonym, which was then disambiguated in the second clause of the sentence (e.g., [She controlled the ball] [which during the game...] vs [which during the dance]). The stimuli showed the speaker producing an iconic gesture during the first clause that would correspond to either the dominant or subordinate meaning. ERPs time-locked to the target word in the second clause showed that congruent gestures elicited a smaller N400 than incongruent gestures. This suggests that listeners use gestural cues online to disambiguate upcoming speech.

In addition to ERP paradigms, which show precise temporal relationships between stimuli and brain processes, neuroimaging via functional magnetic resonance imaging (fMRI) show areas of the brain where processing takes place, normally by looking at changes in blood flow. This is also known as the blood-oxygen-level-dependent, or BOLD responses. These responses are coupled with neural activation. Xu et al. (2009) showed that the integration between meaningful iconic gestures and meaningful speech both activate the left interior frontal gyrus (IFG) and the bilateral medial temporal gyrus (MTG). These findings were corroborated by Straube et al (2012) suggesting that these are the areas involved in the comprehension of representational gesture.
It has been argued that gestures tend to occur more frequently in adverse listening conditions (Hoskin & Herman, 2001; Kendon, 2004). A number of neurophysiological studies have, thus, investigated the integration of gesture in degraded speech. Following the design in Holle & Gunter (2007), Obermeier, Holle & Gunther (2011) modified the temporal synchrony between the gesture and its lexical affiliate. They also included a condition for degraded speech. They found that when the gesture is not temporally co-produced with its lexical affiliate and participants were not asked to pay attention to the gesture, speech-gesture integration does not occur (as determined by the lack of any N400 affects across gesture conditions; see also Habets et al., 2011). However, gesture-speech integration did occur with the misaligned gestures were produced in noisy speech. This finding is important as it suggests that integration depends on the communicative situation and adverse situations tend to favor gesture-speech integration.

An fMRI study by Holle et al (2010) investigated the integration of iconic gestures at various signal-to-noise ratios for speech. Participants watched videos of an actor uttering short sentences that were either accompanied by a congruent iconic gesture or not. They found that in both good and adverse conditions, the superior temporal sulcus and superior temporal gyrus were sensitive to stimuli with gesture, and these areas were more active in the speech degraded condition. The authors conclude that attending to gesture in poor listening conditions can indeed enhance comprehension and help disambiguate speech that is hard to interpret.

In general, less work has investigated whether these integration effects are maintained in second language processing (e.g., Macedonia et al. 2011, Kelly et al 2007). An fMRI study by Macedonia et al. (2011) investigated the neural substrates of iconic and meaningless gestures in foreign language word learning during a word recognition task. Native German-speaking participants were taught novel words from a pseudo-language following Italian phonotactic constraints. These words were taught by video where the actor said the target words accompanied by either an iconic gesture or a meaningless gesture (e.g., touching their knee). During the imaging session, participants were presented with the target words (in both text and audio formats) and asked to press a key if they encountered an unknown word. Much like previous studies, the iconic gestures activated the premotor cortex bilaterally, while meaningless gestures activated a vast brain network with the most activation in the left cuneus, which the authors claim reflect metacognitive processes. This data, coupled with the results from their behavioral experiment (as described in section 1.2.4.1.) lead the authors to
conclude that it is the semantic meaning attached to the representational gesture that drives the memory performance of newly learned words, and not merely movement itself. One ERP study that specifically focused on second language learners by Kelly et al. (2007, as cited in Kelly et al. 2008) taught English learners of Japanese a number of novel Japanese verbs with or without iconic gestures. Results showed that iconic gestures produced a deeper and stronger late positive complex, which is said to reflect the strength of memory coding (Rugg & Curran, 2007). These results suggest that encoding new vocabulary with representational gestures help form longer lasting memories (see section 3.3, for the theoretical basis).

The work reviewed in this section has revealed (a) important brain signatures indicating gesture-speech integration both in L1 and L2 language processing; and (b) a tendency for adverse speech and communicative situations to favor gesture-speech integration. The next section will present behavioral and neurophysiological data that suggest that second language learners may in fact benefit more from meaningful gestural input than native speakers.

1.2.4.3. Stronger gesture effects for second language processing

For second language learners listening comprehension demands are challenging. An often cited study by Sueyoshi & Hardison (2005) compared the effects of gesture and facial cues on listening comprehension in both low and high proficiency learners of English. Participants were assigned to one of three conditions: audiovisual with gesture and face visible, audiovisual with only face visible, and audio only. The results suggested that manual gestures (including beat gestures) significantly improved the results of a multiple-choice comprehension task for the low proficiency learners. High proficiency learners saw the most gains from facial cues. While this study has contributed to the understanding of gesture’s effects in L2 comprehension at different levels of L2 ability, their participants were already living in the L2 environment, and there were no controls to see if similar results would be found in L1. A similar study by Dahl & Ludvigsen (2014) was carried out to control for these issues. They used a drawing task in order to avoid depending on language reading and decoding to convey understanding. Further, the task allowed for the discrimination of what they recall in explicit information and what they comprehend of implicit information. Native English and Norwegian middle school students were exposed to English videos of a speaker describing 4 single-panel cartoon images in two conditions, with and without gesture. Students were then asked to draw the image from the description in the video. They found that
native speakers included more implicit and explicit information than non-natives. When looking at differences between conditions, they found no difference between the Gesture and No-gesture conditions for native speakers, while non-natives included significantly more elements when gestures were present. For implied elements, non-natives scored comparably to natives in the gesture condition.

In an ERP study comparing the integration of iconic gestures with figurative speech between low-proficiency Spanish learners of German and high proficiency Spanish learners of German, Ibañez et al. (2010) found that low-proficiency learners showed smaller amplitudes and a greater latency in processing than the high-proficiency learners. They comment that the high-proficiency group showed waveforms similar to those by native German that were acquired in an earlier study using the same paradigm. This suggests that proficiency level is important to take into account when examining second language paradigms.

A more recent ERP study carried out by Drijvers & Özyürek (2018) specifically investigated the processing of matching vs mismatching iconic gestures in native and non-native speakers of Dutch in both clear and adverse listening conditions. Twenty-four Dutch participants and 23 German advanced learners of Dutch participated in the experiment where they were exposed to videos of a speaker uttering highly frequent Dutch action verbs with congruent or incongruent gestures, which were in turn presented in either clear speech or degraded speech. This led to four conditions: in clear speech with matching iconic gesture, clear speech with mismatching gesture, degraded speech with matching gesture, and degraded speech with mismatching gesture. Regarding the use of gestural input in non-natives, they hypothesized that the N400 effect may be stronger for them than for natives as they will make more use of sensory input from gesture, requiring more neural resources. Drijvers & Özyürek’s (2018) study found that indeed in clear speech conditions, non-native speakers did indeed exhibit a larger N400 component than their native counterparts. As they pointed out, “This may be due to the fact that they pay more attention to gestures when they are unsure about the language” (p.15). In degraded speech, however, the non-native participants did not show a difference in the N400 component between matching or mismatching gestures, both having a smaller amplitude which resembled that N400 in the mismatching gestures with clear speech condition. Native listeners also showed a reduced N400 component in degraded speech. The authors claim that “when signal quality suffers and there are less auditory cues to map semantic information to, non-native listeners are less able than native listeners to benefit from semantic information from the gesture to boost comprehension and resolve the degraded
These results are consistent with other studies which found that the N400 component in native speakers decreases linearly as sensory input is diminished (Obleser & Kots, 2001; Strauß, Kotz & Obleser, 2013).

The results reviewed in this section suggest that adverse speech conditions (be it the presence of degraded speech, or the fact of not being native in the target language) lead to a strong use of referential gestures to better comprehend and recall L2 speech. However, most of the previously cited studies specifically investigate referential gestures that convey semantic content. In general, less is known specifically about the role of beat gestures in the recall and comprehension of speech, especially on the potential effects of beat gesture in non-native speech processing as compared to native speech processing. Studies 2 & 3 of this PhD project will explore this research question particularly regarding the processing of beat gestures in native and non-native speech with both behavioral and neurophysiological responses. The next section will review the studies focusing on the cognitive benefits of beat gestures in the first and the second language.

1.2.5. Cognitive benefits of beat gestures in the first and the second language: behavioral and neurophysiological evidence

Beat gestures have received relatively little attention in the literature regarding their cognitive benefits compared to referential gestures (see section 1.2.4. above). However, a number of studies have indeed taken on the role of investigating this issue, mostly reporting conflicting results. This section will review the previous behavioral and neurophysiological studies that have focused on the benefits of beat gestures for information recall, language comprehension, and phonological memory.

1.2.5.1. Neurophysiological evidence

Research in cognitive neuroscience has focused on the processing effects of beat gestures in the native language, showing that beat gestures help the listener to direct the focus of attention (Biau & Soto-Faraco, 2013; Dimitrova et al., 2016), ease semantic processing (Wang & Chu, 2013) and disambiguate the sentences that contain complex syntactic structure (Holle et al., 2012). In the study by Hubbard et al. (2008), beat gestures were shown to be integrated with speech. Speech accompanied by gesture showed increased activity in the bilateral non-primary auditory cortex, and beat gestures provoked more activity in the left superior temporal gyrus/sulcus when compared to nonsense movements. A third area, the
right plenum temporale was found to be a multisensory integration site for beat gestures and speech. These findings suggest that there is a common neural substrate for processing both beat gesture and speech, reflecting their joint communicative role.

Biau and Soto-Faraco’s (2013) study exposed participants to a discourse by a Spanish politician with a wide range of beat gestures in two conditions: audio-visual and audio only. By measuring participants’ Event-Related Potentials (ERPs), the researchers found a positive shift at an early sensory stage in the audiovisual condition (before 100 ms post word onset) in frontal and centro-parietal regions that may reflect early processing of audiovisual stimuli, and also an enhanced P2 component in the parietal region. This component is typically associated with auditory integration (Näätänen et al., 1997) and may be modulated by attention (Hillyard, Hink, Schwent, & Picton, 1973). There was no difference in the ERP components when the same words were heard 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 brain 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.

The study by Holle et al. (2012) explored the use of beat gestures in disambiguating syntactically ambiguous sentences. They centered their attention on the P600 component, which is related to syntactic processing and is typically larger when exposed to less preferred syntactic structures (for example “The woman persuaded to answer the door”) (Osterhout & Holcomb, 1992). In the study by Holle et al. (2012) participants were exposed to ambiguous sentences in German in respect to subject and object role (for example “Peter sagt dass die Frau die Männer gegrüßt” – literally translated as “Peter says that the woman the men greeted have”). Stimuli were presented either in the preferred SOV structure (above) or the less preferred OSV structure (“Peter says that the men the woman greeted have”). They found that the P600 effect disappeared when the non-preferred noun phrase was produced with a beat gesture. These findings suggest that beat gestures enhance the credibility of the non-preferred syntactic structure. However, it remains unclear whether results of the previously mentioned studies are due to the presence of pitch accentuation that occurs together with beat gestures.

To our knowledge only few studies have centered their attention on how brain integrates beat gestures with co-occurring prominence in speech. In the study by Wang and Chu (2013) participants had to watch videos of a person producing co-speech gestures. The critical words were accompanied by a beat gesture, a control hand movement or no gestures.
The critical words were produced either with pitch accent or without pitch accent. While participants watched the video their ERPs were recorded. The results showed modulation of the N400 component, which indexes semantic integration of a target word, being larger when semantic integration is more difficult (Kutas & Hillyard, 1980). Beat gestures and control hand movements elicited smaller negativities in the N400 time window in comparison to the no hand movement condition. However, beat gestures elicited smaller negativities in the N400 over right posterior electrodes than control hand movements. Taken together, this suggests that while both beat gestures and control hand movements facilitated the semantic integration of the target word, beat gestures contain special features for increased semantic processing. Interestingly, no interaction between beat gestures and pitch accent was observed in the study. That is to say that these effects were independent of whether beat gestures and pitch accentuation co-occurred or not. These results suggest that prominence in gesture and prominence in speech trigger the attentional system separately for semantic processing. The results of the study conducted by Wang and Chu (2013) may be, from our point of view, influenced by the type of stimuli used in the experiment. In Wang and Chu’s (2013) videos only the torso and limbs of the person saying sentences were visible in the video recordings. The participants could not see the mouth of the speaker in the video. The authors’ decision to hide lip movement in the videos could have had a negative effect on the natural integration of prominence coming from speech and gesture. Also, the type of gesture chosen for a control hand movement (moving hand to the right palm-opened), complies the characteristic of beat gestures.

Another recent ERP study by Dimitrova, Chu, Wang, Özyürek, & Hagoort (2016) investigated how beat gestures modulate the processing of accented-focused vs. unaccented-non focused words in context. In the study, participants had to watch video recorded dialogues. The target word in each dialogue was presented under the 6 experimental conditions: 1) pitch accent + beat gesture; 2) pitch accent + grooming movement; 3) pitch accent + no hand movement; 4) no pitch accent + beat gesture; 5) no pitch accent + grooming movement; 6) no pitch accent + no hand movement. All beat gestures were manipulated so that the gesture apex co-occurred with the word onset (hereafter labeled as target onset). In the sentences in conditions 4), 5), 6) the pitch accent was placed on another word in the sentences. The results of the study showed that at -400 to -300 ms to target onset, they found that focused targets (regardless of gesture) elicited a negativity relative to non-focused targets. Further, stimuli with hand movements (both beat gestures and grooming movements) show a
marginally significant negativity relative to targets with no gesture. At -300-0 ms to target onset, they found no effect of focus, but positive clusters by both beat and grooming gestures. This early positivity is in line with previous studies (Biau & Soto-Faraco, 2013; Wang & Chu, 2013). At 200-600 ms after target onset, they found a positivity for focused targets relative to non-focused targets. Again with gesture, there was a similar positivity. These results replicate those of Biau & Soto-Faraco’s (2013) study, where they found a late positivity 200 ms after word onset. In light of these results, the authors hold that beat gestures are nonverbal cues to focus. They also conclude that the positivity around 200ms after target onset may reflect an increase in listener’s attention. The authors call this the ‘novelty P3a’ component, which shows the allocation of attention to relevant information in speech. Moreover, while the beat gestures and focused words did not interact in the P3 time window, the P3a effect for focus was largest in trials with a beat gesture. “This numerical difference suggests that listeners’ overall attention to focus is increased when the speaker produces a beat gesture” (p. 1267).

All in all, the early positivity obtained in focused words with beats reported by Biau & Soto-Faraco (2013) and Wang & Chu (2013), is attributed to enhanced attention to the visual stimuli. This positivity lasted until the gesture was retracted. Authors hold that grooming can be a distraction or even impede the comprehension of information from speech. However, none of the previously cited studies explored the effects of beat gesture in non-native language processing. To the author’s knowledge, only one ERP study (described in section 1.2.4.3.) has investigated the processing of iconic gestures in second language (Drijvers & Özyürek 2018). They found that in clear speech conditions, non-native speakers did indeed exhibit a larger N400 component than their native counterparts, suggesting that non-natives may process gesture more strongly than natives in clear speech.

The aim of Study 2 of this thesis will be to investigate the impact of prosodic (pitch accents) and gestural prominence (beat gestures) integration on contrastive information memorization in a naturally produced discourse in native and non-native speech. Using event related potentials (ERPs) we aim to explore potential neural response differences between prosodic and gestural prominence in native and non-native speech. This study is being carried out as a collaboration with the Multisensory Research Group at the Center for Brain and Cognition at UPF (Lab Director: Salvador Soto-Faraco). The general hypothesis is that non-native listeners might process beat gesture more strongly than natives (see Drijvers & Özyürek’s, 2018 results for iconic gestures). The research questions are as follows: Does the observation of beat gesture in discourse lead to stronger attentional and integration
improvements over pitch accent alone (which would be seen as early positivity and specifically a P3a effect, as per Dimitrova et al. 2016) and over observation in L1 discourse? Does the observation of beat gesture in L2 discourse lead to stronger attentional and integration improvements than in L1 discourse? Regarding the latter, we expect that the processing of beat gestures will be stronger when participants are processing non-native vs. native speech (as per Drijvers & Özyürek, 2018).

1.2.5.2. Behavioral evidence

In contrast with the consistent evidence showing the benefits of referential gestures, studies investigating beat gestures have found more conflicting evidence regarding recall, comprehension, and phonological processing. While some behavioral studies have shown that beat gestures aid information recall and comprehension both by children and by adults (Igualada, Esteve-Gibert, & Prieto, 2017; So, Sim Chen-Hui, & Low Wei-Shan, 2012; Austin & Sweller, 2014; Llanes-Coronina et al., 2018; Kushch & Prieto, 2016; Kushch, Igualada & Prieto, 2018) others have not (see Feyereisen, 2006; Morett (2014); Macoun & Sweller, 2016).

Recall

One study by Feyereisen (2006) showed a set of 26 sentences to adult participants. Ten of these sentences were shown with an accompanying representational gesture, while 10 sentences contained a non-representational gesture (6 sentences were filler items). The author found that in a free-recall task, participants could remember sentences better when they were accompanied by representational gesture than when accompanied by non-representational gesture. However, this study did not take prosody into account. Indeed, independent evidence has shown positive effects of prosodic prominence (L+H* pitch accent) on both the ability to recall novel information and on the comprehension of information (Fraundorf, Watson, & Benjamin, 2010 for English; see also Kushch & Prieto 2016 for Catalan). Recent studies controlling separately for the effects of prominence in speech and prominence in gesture show that beat gestures together with pitch accentuation lead to better information memorization in comparison to just prosodic prominence. In the study by Kushch & Prieto (2016), Catalan native speakers were presented with short discourses which contained two sets of contrastive pairs. The target word within each of these contrastive pairs was accompanied by either pitch accentuation (L+H*) or pitch accentuation with a beat gesture. They found that target words
Another set of studies have shown contradictory results of the use of beat gestures in memory recall in preschoolers. A study by So et al. (2012) investigated the importance of semantic meaning encoded in gestures for memory recall. They tested both adults and children, and found that while beat gestures were beneficial for adults, children did not show any benefits. In their design, however, words were presented in isolated target sentences without context. Two subsequent studies found positive results for beat gestures in recall by children when they were presented in pragmatically relevant contexts. Austin & Sweller (2014) investigated the effects of both referential and beat gestures in the recall of spatial directions in both children and adults. They found that children recalled the directions better when they were presented with gesture, regardless whether they were referential or not. Igualada et al. (2017) investigated the issue in a narrative context, where target words produced either with or without beat gesture contrasted with adjacent words. They found that children remembered more words in the beat condition, and that this recall was localized to target words and not their adjacent words (i.e., there was no global recall effect). These results were subsequently confirmed by Llanes-Coromina et al (2018), who tested the differential effects of beat gesture and prosodic prominence on the recall of contrastively focused words in preschoolers. Participants were exposed to three short stories with target words presented in three conditions; prominence in both speech and gesture, prominence in speech only, or non-prominent speech. They found that children remembered words best then they were accompanied by beat gestures which integrate both prosodic and gestural prominence.

Regarding the role of beats in the second language vocabulary recall, two studies in particular have shown somewhat conflicting results. Morett’s (2014) study, as described in section 1.2.4.1., also analyzed the use of beat gestures for target vocabulary encoding and recall by both the “Explainer” and “Learner” participants in a vocabulary teaching task. The results showed that the production of representational gestures by Explainers facilitated both (a) their own encoding of the target words and (b) the recall of those target words by the Learners. However, Explainer’s production of beat gestures facilitated their own recall and enhanced the Learner’s encoding. Thus the results of this study imply that beat gesture production is more important for recall than merely viewing them in input.

In another study, Kushch, Igualada & Prieto (2018) investigated the impact of viewing beat gestures with or without prosodic prominence on Catalan native speakers learning
Russian in a within-subject design. The participants were taught the Russian nouns by watching videos of a native Russian speaker producing target vocabulary in a standard context sentence (e.g., *Bossa es diu “suma” en rus* – *Bag is called “sumka” in Russian*). The target words were presented in 4 conditions: no prominence in either speech or gesture, prominence in speech alone, prominence in gesture alone, or prominence in both speech and gesture. During the training phase, a total of 16 words (where each word always appeared in the same condition) were presented to the participant 4 times. A free recall task and an audio-cued task were administered to the participants. They found significant results for both the prominence in speech and the bimodal prominence conditions, with the latter showing the greatest benefit. This indicates that seeing beat gestures have a positive effect on recall when they are produced with prosodic prominence.

All in all, the results on this section show that even though all gesture types may aid in memory, they may function in different ways. Moreover, contradictory effects are found with respect to the effects of beats in the second language. To our knowledge, no study has directly compared the effect of beat gestures on recall in native and non-native participants. For this reason, Study 3 of this thesis will test the potential stronger effects of beats in non-native speech than in native speech (see section 5.3 below). Moreover, Study 2 will also collect behavioral data in parallel to the neurophysiological data with a short memory quiz after the presentation of the stimuli. Both Studies 2 and 3 will allow us to compare the neural patterns and the behavioral effects of pitch accent and beat gesture on recall in both native and non-native speakers.

**Comprehension.**

To our knowledge, very few studies have examined the effect of beat gestures on comprehension, and with conflicting results (Macoun & Sweller, 2016 vs Llanes-Coromin et al., 2018). Perhaps this is due to the fact that they do not directly convey the semantic content that is lexically affiliated. However, beat gestures do convey information and discourse structure (see section 1.2.3.2. above).

Macoun & Sweller (2016) tested the effects of different gesture types on preschooler’s comprehension of a short narrative. One hundred and one preschool children were exposed to 2-min long narratives where speech was either accompanied by iconic gesture, a deictic gesture, a beat gesture, or no gesture. After watching the narrative, the children were asked a free recall question followed by 15 specific questions relating to what was said in the
narrative. They found that answers given in the free recall question were significantly more likely to come from deictic and iconic gestures, while there was no difference between beat gestures and the control condition. Similarly, children performed much better on the question items pertaining to general comprehension in the iconic and deictic conditions. This study suggests that preschoolers only take into account referential gestures. It must be noted, however, that the beat gestures they used were only on contrastively focused words.

Llanes-Coromina et al.’s (2018) study also investigated the role of beat gestures in comprehension by preschoolers, and took into account the various pragmatic uses of beat gestures. They presented 5- and 6-year-olds with narratives which contained target items that were either accompanied by speech prominence alone, or speech prominence with beat gesture. Beat gestures were placed on contrastively focused words, as well as on discourse markers, such as after, that, and then. A comprehension task showed that the children comprehended the story better when they contained beat gestures.

All in all, the previous studies regarding the cognitive benefits of beat gesture seem to suggest that beat gesture may indeed be a valuable tool for comprehension and recall in both native and non-native populations. However, no studies have specifically investigated the effects of beat gestures in second language input on comprehension and recall. Study 3 of this thesis will fill this gap by looking at the adult language learning population, and by specifically using beat gestures in input to determine their capacity for both comprehension and recall in a narrative drawing task adapted from Dahl & Ludvigsen (2014) (see section 5.3 below).

**Phonological memory.**

Beat gestures have also been shown to aide second language learners with pronunciation in the L2. Most studies are qualitative in nature, showing how teachers use gestures in the classroom to teach second language pronunciation (e.g., Smotrova, 2017) or how students use beat gestures themselves in order to manifest the abstract nature of prosody in concrete terms for their learning (e.g., McCafferty, 2006). Some studies have taken these qualitative results and investigated them further in more experimental settings.

One study by Gluhareva & Prieto (2017) investigated the impact of observing beat gestures on pronunciation. Learners of English were exposed to a training video where the L2 instructor responded to discourse prompts. For example, the discourse prompt would show an image of two people, with one person pointing at their watch. Underneath the image reads the
sentence “You are in the metro and would like to ask a stranger for the time.” The instructor responds “Excuse me, what time is it?” while either producing beat gestures or not. The participants recorded their own responses to the same prompts before and after the training session. Their recordings were then rated by native speakers for accentedness on a scale of one to seven. The results showed that the beat-training group significantly improved their pronunciation from pre-test to posttest.

Another study by Llanes-Coronina, Prieto & Rohrer (2018) focused on the effects of producing beat gestures on reading performance. In a pre-test/training/posttest design, students were asked to read 5 texts aloud (2 for the pre-test, 2 for the training, and 1 for the posttest). In the training session, half of the students were given no gesture instruction, while the other half was encouraged “to move their hands while reading”, ultimately producing non-referential beat gestures. The posttest text was syntactically more complex and longer than the pre-test and training texts. Participant recordings were rated by American English native speakers on four pronunciation measures, namely accentedness, comprehensibility, fluency, and expressiveness. Participants in the gesturing group received significantly better pronunciation scores of accentedness, comprehensibility, and fluency on the posttest than the non-gesturing group.

Thus regarding pronunciation, the few studies that investigate the issue seem to suggest that both observing and producing beat gestures lead to better pronunciation in L2 speech. However, none of these studies investigated the impact of beat gesture on pronunciation when the input is immediate and when the speaker listens to the input and enacts it in an online, temporally coordinated manner. **Study 4** will use a shadowing paradigm, which is a technique originally intended to train simultaneous interpreters to be able to quickly translate and repeat what was said. In 1992, the technique was integrated into EFL classrooms in Japan (Tamai, 1992, as cited in Hamada, 2018). Hamada (2018) described shadowing as a technique in which learners must simultaneously reproduce what they hear following exposure to small chunks of audio stimuli. Because the repetition is done as quickly as possible, it is considered an online task (Shiki et al., 2010; Bailly, 2003). Further, learners need not pay attention to the meaning of the words they hear, only the sounds. By focusing learner’s attention on phoneme perception skills, cognitive resources that would normally be devoted to word recognition are freed. Thus it is considered a bottom-up learning process in that they attend to individual sounds which later improves word recognition in listening comprehension (Hamada, 2016a).
While most studies focus on the benefits to listening comprehension, a few studies have looked at improvement in pronunciation through shadowing, with mostly positive results. While Hamada (2016b, as cited in Hamada, 2018) found little improvement after Japanese students of English were trained with a shadowing task for 20 minutes, twice a week, during a month, Foote & McDonough (2017) trained English learners of various L1 backgrounds with shadowing for eight weeks (20 minutes twice a week) and found improvement in imitation skill, comprehensibility, and fluency (with no affect for accentedness). Another study by Hsieh, Dong & Wang (2013) examined the use of shadowing at both the word and sentence level to teach English intonation to Taiwanese learners and found that the shadowing group performed significantly better than the control group in terms of pronunciation, fluency, and intonation. However none of the studies specifically used audio-visual input aimed at incorporating gesture. Study 4 aims to combine the audio-visual aspect from Gluhareva & Prieto (2017) with the online training technique offered by shadowing. The results of Study 4 will show the immediate effects of beat gestures on phonological awareness, which will manifest as immediate gains in pronunciation (see section 5.4 below).

2. Goals of the dissertation

The current PhD thesis has three main goals. The first goal is to explore how beat gestures function in two different languages in regards to their prosodic and pragmatic features. This will enable us to have a clearer picture of the properties of beats for native speakers vis-à-vis beat gesture production. The second goal is to determine the benefits of beat gesture in both first and second language from both a neurophysiological perspective and a behavioral perspective (in regards to recall and comprehension effects). This will ultimately reveal the cognitive benefits of beat gestures, and their relative strength in L1 and L2. The third goal is to explore the benefits of beat gesture on phonological awareness in second language learners. All in all, this work intends to empirically assess the benefits of beat gestures for second language processing across different areas and the real world applications of these benefits in the second language classroom. The results of these studies will probably show that beat gestures are far from “meaningless” gestures and deserve more attention in future studies.

4 Foote & McDonough’s (2016) study gave participants audio clips of popular television shows on an iPod, and the corresponding videos were available online if the participant wished to watch the video while doing the task.
To achieve these goals, four empirical studies will be carried out. The following research questions will be addressed in each study: (Study 1) Are there any potential cross-linguistic differences in the temporal and pragmatic use of beat gestures in two languages that structurally encode prosodic prominence in different ways? (Study 2) Do beat gestures increase attention and modulate semantic processing in second language? Further, are these effects stronger in L2 than in L1? (Study 3) Will beat gestures in speech input facilitate recall and comprehension in a narrative-drawing task? Further, will non-native listeners make more effective use of beat gestures than native listeners? (Study 4) Can the presence of beat gestures in L2 input lead to better second language pronunciation measures during a shadowing task?

Our general hypotheses are the following: (1) Study 1: Beat gestures will show a tight temporal affinity to pitch accents in both languages, with potential differences according to pitch accentual strength (e.g., nuclear vs pre-nuclear), positioning (e.g., phrase-initial vs phrase-final), and their pragmatic function (e.g., rhythmic marking vs information structure marking); (2) Study 2: Beat gestures will increase attention and modulate integration in both first and second language, with these effects being more robust in non-native listeners; (3) Study 3: The presence of beat gestures will aide comprehension and recall in a narrative drawing task in both native and non-native listeners, with effects being stronger in the non-native group; and (4) Study 4: The presence of beat gestures will lead to significant gains in pronunciation ratings for French learners of English in a shadowing task.

3. Theoretical framework

3.1. Gesture and speech as one system

Many theorists hold that gesture and speech come together to form one system. In an early article by McNeill (1985), the author argued that speech and gesture are parts of the same psychological structure and that they share a computational stage. This argument was based on the three synchronies that he described in further detail in McNeill (1992), as well as synchronies in development and pathology. As previously mentioned, co-speech gestures show synchrony with speech at three levels; a temporal synchrony occurs where gestures tend to be produced with prominent positions in speech. That is, the stroke phase of the gesture tends to occur at or just before the stressed syllable of the lexical affiliate. They also synchronize on a semantic level, where meaning in gesture often parallels meaning in speech. This is particularly true regarding iconic gestures, where significance in speech is often
conveyed in gesture (for example, the uttering “crawling up a pipe” while hands rising upwards, mimicking the crawling action). Speech and gesture also share the same pragmatic function. Here, McNeill specifically mentioned beat gestures in that they mark discourse-oriented functions. For example, beat gestures may mark changes from narrative to extra-narrative discourse.

Beyond these three synchronies, McNeill (1985) also mentioned two more arguments, namely that they develop together with speech in children, and break down together with speech in aphasia. Indeed, children often point to objects, allowing them to refer to objects before they have words for them. These can then the coupled with one-word utterances to convey two pieces of information, one in each modality. For example, a child may point at a cup and say “mine”. Furthermore, it has been shown that the occurrence of pointing with one word utterances is predictive of two-word utterance production (see Iverson & Goldin-Meadow, 2005). Regarding their breakdown in aphasia, McNeill claimed that patients with Broca’s aphasia are able to use referring terms, yet they cannot combine these terms into larger grammatical wholes. Similarly, these patients continue using iconic gestures yet produce few or no beat gestures (which would indicate relationships beyond referential terms). The reverse is true for patients with Wernicke’s aphasia, where patients have difficulty forming coherent semantic plans, but are able to construct sequences of words. Almost all of their gestures are considered non-referential beat gestures.

The integration of gesture into speech production models has led to three models of multimodal speech production. The first is the Free Imagery Hypothesis, which holds that representational gestures are generated before linguistic formulation processes. They may either be generated from spatial images in working memory (Krauss, Chen & Gottesman, 2000) or from a “conceptualizer” where a pre-verbal message is conceived and later processed into linguistic formulations (De Ruiter, 2000). In direct contrast to the Free Imagery Hypothesis, the Lexical Semantic Hypothesis claims that gestures are generated after the linguistic formulation of speech. In other words, gestures are produced as a result of the semantic information in lexical items that have already been selected (Butterworth & Hadar, 1989). The Interface Hypothesis, however, claims that gestures are generated from a combination of both imagistic and linguistic processes (Kita & Özyürek, 2003). In this model, the “Communication Planner” generates the communicative intent and decides what modalities are to be used. This information is sent to a “Message Generator” and an “Action Generator”, which exchange information in an online, bi-directional manner to determine
what information is to be expressed which modality. This latter gesture/speech production is the predominant hypothesis in the domain today. While this model is designed to capture gesture production in general, the authors only commented or described representational gestures.

3.2. Grounded and Embodied Cognition

Theories of grounded (Barsalou 2008, 2010) and embodied cognition (Ionescu & Vasc, 2014) state that the body and mind are influenced by each other in cognitive processes. Though both theories are very similar, the former claims that cognition is based in modal simulations, bodily states, and situated action, rejecting the idea that amodal symbols represent knowledge in semantic memory. Accounts for Grounded Cognition are may be based on the role of the body, particularly that bodily states can cause or be effects of cognitive states (Barsalou et al., 2003) or on the role of simulation (or reenactment) “of perceptual, motor, or introspective states acquired during interaction with the world, body, and mind” (Barsalou, 2008: 618). The author described the act of sitting in a chair, for example. He claims that as an experience occurs (such as easing into the chair), the brain captures states across all modalities and integrates them into memory as a multimodal representation (for example, the look and feel of the chair, the act of sitting, the feelings of comfort and relaxation). Thus, when knowledge is required to represent a chair, the brain reactivates all of these multimodal experiences associated with it. Similarly, embodied cognition holds that cognition is based in sensory-motor processes, our body’s morphology, and internal states (Ionescu & Vasc, 2014). What distinguishes it from Grounded Cognition is that instead of the body creating mental representations, it is the body itself replaces these mental representations. As the differences between these two theories go beyond the scope of this thesis, these concepts will be simply referred to as one under the blanket term “Grounded and Embodied Cognition.”

Building upon these theories, Hostetter & Alibali (2008, 2010) developed the Gesture as Simulated Action (GSA) framework. The GSA framework proposes that the activation of the motor system during speech production (i.e., representational gestures) stem from action simulation. The authors give the example “I picked the pepperoni off the pizza”, where a picking gesture with thumb and fore-finger is performed. Simply put, the speaker thinks of a picking action, and their motor system simulated it based on past real-world experiences, much like it would do if she read the word pick. The iconic gesture is, thus, the result of this
neural simulation of picking. While the GSA framework specifically accounts for representational gestures, beat gestures can also be seen as having a role in Grounded and Embodied Cognition. One study by Carlson et al. (2007) investigated the effects of pointing while counting. Participants were presented with a certain number of asterisks on a screen and asked to count them either while pointing or not pointing. They found that pointing tended to lead to more correct answers. Interestingly, when participants were not allowed to point, they replaced the pointing gestures with head nods (considered beat gestures as per Prieto et al., 2018) 75% of the time. Further, in non-pointing trials, participants who nodded performed significantly better than participants who did not nod. Thus following these theories, cognition and the body are intimately linked and gestures may play a critical role in exploring this interaction.

3.3. Dual Coding Theory

One possible explanation for the encoding visual and auditory information in memory is the Dual Coding Theory (Paivio, 1971, Baddeley, 1990). This theory holds that when information is encoded in two different modes, that learning is improved. Baddeley’s (1990) model held that when information is encoded in multiple modalities, deeper traces are left in the memory system, allowing for better recall. His original model consisted of three independent components. The phonological loop acts as a storage system for phonological stimuli. The visuospatial sketchpad encodes non-verbal visual/spatial information. Finally, the central executive device manages attention and coordinates the other two components.

In 2000, Baddeley refined his model to include the episodic buffer component. This component is used for the temporary storage of information that has been presented through multimodal code. This, in turn, is encoded as a single episodic representation, rejecting the previous idea that auditory and visual information were stored separately. The Dual Coding Theory has been used in a number of gesture studies to explain why word recall is improved when words are presented with gestures (Tellier, 2008; Kushch, Igualada & Prieto, 2018).

4. Hypotheses

The review of the literature has laid the groundwork for the research questions to be addressed in this PhD thesis. As mentioned before, the main underlying hypothesis of this thesis is that beat gestures are indeed meaningful gestures that can benefit listeners even
though they do not convey direct referential meaning to speech. Further, these gestures will be even more beneficial to non-native listeners than native ones. We base this hypothesis on two complementary studies. First, the study by Dahl & Ludvigsen (2014) showing that in a narrative drawing task the presence of referential gestures aids non-native listeners more than native listeners in regards to memory and comprehension. The second neurophysiological study by Drijvers & Özyürek (2018) corroborated these results by showing that that in clear speech, non-native listeners showed a larger N400 component than native listeners, indicating their stronger levels of processing. Taken together, these results make up the basis of our hypotheses in Studies 2 and 3 regarding the interaction between native-speaker status and the effects of beat gestures. The sections that follow will briefly lay out the hypotheses for the research questions in each study individually.

4.1. Study 1

This study will serve as a baseline study and will investigate the prosodic and pragmatic features of beat gestures in a cross linguistic fashion, focusing on French and English. As mentioned before (see section 1.2.2.), previous work has shown that gesture strokes tend to be anchored to pitch accents (e.g., Loehr, 2007, 2012; Esteve-Gibert & Prieto, 2013, among others). However, not much work has been carried out cross-linguistically and with languages which display differences in phrasal prominence assignment, like English or French. Similarly, little work has been carried out in the role of nuclear and pre-nuclear prominences across languages. In the case of English and French, even though we expect beat gestures to occur with pitch accents in both languages, the beat gesture is most likely to occur on the nuclear pitch accent in English, and the initial pitch rise in French (particularly during emphasis of a particular item). We believe this to be the case for French, as the initial pitch excursion in French is attributed an emphatic function under certain contexts (see Astésano, 2007; Di Cristo, 1999, 2000). Finally, we expect correlations between particular hand forms and pragmatic functions of beats (e.g., “open palm away” for low epistemic commitment, see Figure 4 in section 1.2.3.2.) for each language, which may or may not be language-dependent. This study will add to the literature regarding the interaction between beat gestures, prosody, and pragmatics, particularly regarding beat gesture production in French. Our general hypotheses are the following: (1) Beat gestures will show a tight temporal affinity to pitch accents in both languages, with potential differences according to pitch accentual strength
(pre-nuclear vs nuclear) and positioning (initial vs. final); (2) Beat gestures can show differences in the beat form and semantic meaning correspondences across languages.

4.2. Study 2

As mentioned before (see section 1.2.5.2.), neurophysiological and neuroimaging studies show that non-referential beat gestures seem to increase attention and benefit comprehension for the listeners in L1 speech (see Biau & Soto-Faraco, 2013; Dimitrova et al., 2016; Wang & Chu 2013). However, these effects have not been tested in non-native populations. The goal of Study 2 is to assess the effects of beat gesture on attention (reflected in the P3a component described in Dimitrova et al., 2016) and semantic integration (reflected in the N400 component as per Wang & Chu, 2013) in both native and non-native populations of Catalan and English speakers. We expect the effects of beat gestures in non-native populations to be greater than in native populations. Specifically, the two components will show larger amplitudes, indicating that non-natives are processing gesture in a stronger manner than natives. Additionally, we will collect behavioral data from the memory quizzes, where we expect the gesture condition to be remembered most, followed by the prosodic prominence condition and the baseline condition remembered least. This study will add to the literature by being the first to directly compare the processing of beat gesture in native and non-native populations.

4.3. Study 3

The studies mentioned in the review of the literature suggest that beat gestures may indeed be a tool that can be exploited by both native and non-native listeners to boost recall and aide comprehension (e.g., Igualada, Esteve-Gibert, & Prieto, 2017; Llanes-Coromina, Vilà-Giménez, Kushch, Borràs-Comes, & Prieto, 2018 for L1; Kushch, Igualada, & Prieto, 2018 for L2). The results previously reported seem to indicate that these benefits are influenced by two independent factors. First, they seem to require being presented in a pragmatically relevant context, where they serve to highlight discourse and information structure (as previously described). Second, their benefits are boosted by the presence of prosodic prominence. However, no study has directly compared these effects in first and second language. Following up on some studies suggesting that referential gestures may be more beneficial in adverse speech conditions and in second language speech (e.g., Dahl & Ludvigsen, 20114; Drijvers & Ozyurek, 2017), the aim of Study 3 is to determine if the
beneficial cognitive effects of beat gestures are augmented in the second language as opposed to the first language. It is believed that the presence of beat gestures during narrative exposure will lead to participants better recalling and comprehending that narrative, as reflected in the drawing task. Further, it is believed that non-native listeners will have greater gains than native listeners when beat gestures are present.

4.4. Study 4

The final study investigates the effects of beat gesture on phonological awareness via a shadowing task. It is believed that the presence of beat gestures in L2 input will lead to better pronunciation ratings for accentedness, comprehensibility, and fluency (e.g., Gluhareva & Prieto, 2017; Llanes-Coromina et al., 2018). What distinguishes this work from previous studies is the fact that in the current study, participants will have an active training session where they will be asked to imitate in a simultaneous fashion the native speaker various times. Given the beneficial effects of the online shadowing techniques for imitation skills, comprehensibility, and fluency (Foote & McDonough, 2017; see Hamada, 2018 for a review), we expect that enacted shadowing techniques will allow for a deeper encoding in memory for the better recall of correct phonological forms during the posttest. This study, together with Studies 2 and 3, will have direct implications for L2 pedagogy.

5. Experimental Studies

The following section will describe each study to be carried out in detail, covering the research questions, materials, methodology, and expected outcomes for each study. As Study 2 has already begun, an additional section discussing the interim results will be included.

5.1. Study 1: A cross-linguistic examination of the prosodic and pragmatic features of beat gestures

5.1.1. Research Questions

The first study is a baseline study which aims to examine the temporal and pragmatic features of beat gestures from a cross-linguistic point of view. As mentioned in section 1.2.3., previous literature on the prosodic and pragmatic features of beat gestures is limited. Regarding the prosodic features, the few abovementioned studies have mostly focused on
English as the language of study with very limited sample sizes. Little is known about how beat gestures function in a language with a fixed phrasal-level encoding of prominence. Regarding the pragmatic features, most research is qualitative in nature and focuses on very specific hand forms. Little is known about these functions cross-linguistically.

This study aims to answer two questions. First, how do beat gestures temporally align with prosody in two languages that structurally encode prosodic prominence in different ways from the point of view of phrasal positioning and tone type? Second, are these differences related to differences in pragmatic function (e.g., rhythmic marking vs information marking)?

In essence, our question is whether there are any potential cross-linguistic differences in the temporal and pragmatic use of beat gestures. In order to explore these questions, a systematic prosodic and gestural corpus annotation of academic discourses by native English and French speakers will be carried out in collaboration with the Natural Language Processing Research Group (TALN) at the Department of Information and Communication Technologies at Universitat Pompeu Fabra (Dra. Mireia Farrús & Dr. Leo Wanner).

5.1.2. Materials

The baseline corpus will be based on a total of 30 Ted Talks, half of which will be executed by native English speakers and half by native French speakers. Ted Talks were chosen due to their academic nature, which tends to favor the use of non-referential beat gesture. The English subset of the corpus will be taken from the database developed by Farrús, Lai & Moore (2016), which offers a time-aligned transcription of the Ted Talks. The French subset of videos will be collected either directly from www.ted.com or from other websites which host videos from Ted-affiliated events (TedX).

5.1.3. Coding

The corpus is currently being manually annotated for gesture in ELAN (Wittenburg et al., 2006) - a program that allows for precise annotation of audio-visual recordings and it is of standard use in the field of gesture research. A coding scheme established by Prieto et al. (2018) will be used to code for co-speech gesture annotation and its prosodic integration.
Let us take the following example into consideration. The beat gesture that occurs on the second word in the sentence\(^5\) "This \textit{geo}engineering idea, in its simplest form, is the following…” will be analyzed (see Figure 5).

First, in our coding scheme, a set of five main tiers are used. The first tier includes the orthographic transcription, the rest of the main tiers classify the type of articulator (manual, head, eyebrow, others) and establish the gesture type within this articulator. Each tier contains the information:

- **Main Tier 1**: Orthographic Transcription

- **Main Tier 2**: Manual Gesture Type. Here McNeill’s (1992) standard classification of referential gestures into iconic, metaphoric, deictic gestures, and an expanded definition for beat gestures (as described in sections 1.2.1. and 1.2.3.2.) is used.

- **Main Tier 3**: Other articulators. Here we code for other articulators, such as the torso.

Four different aspects of each gesture articulator are hierarchically coded as sub-tiers under each of the parent (main) tiers based on each articulator. The sub-tiers are described as follows:

- **Sub-tier 1**: Gesture form. A simplified version of MIT Coding Manual (“http://web.mit.edu”) which labels the \textit{hand shape}, \textit{orientation}, and \textit{trajectory} components for the stroke is used. For non-manual gestures, only trajectory is coded.

- **Sub-tier 2**: Gesture phases. Shattuck-Hufnagel et al.’s [9] proposal for coding the phase structure of manual gestures is used. Phases will include preparation phase, beginning and end of the gesture stroke, retraction phase, as well as optional pre- or post-stroke holds. A sub-tier codes for the gesture apex separately. For non-manual articulators, only the apex is coded. The images of the speaker in Figure 5 show the main phases of a bimanual beat gesture.

- **Sub-tier 3**: Prosodic Component. This sub-tier will contain two types of information, represented in two further sub-tiers. The first sub-tier is perceptual in nature; the annotator assesses whether the gestural prominence (\textit{e.g.}, gesture strokes or apexes are associated with prosodic prominence (+) or not (-). The second sub-tier will

\(^5\) \textbf{BOLD} indicates the syllable which contains the gesture apex in Figure 5
contain previously annotated ToBI labeling imported from PRAAT in order to independently analyze the gesture-speech temporal alignment in finer detail.

- **Sub-tier 4: Pragmatic Component.** This tier specifies the various pragmatic functions of the beat gesture. The pragmatic sub-tier will be manually coded using the coding scheme proposed by Ferré (2011) which classifies the pragmatic functions of beat gestures as prosodic, discursive, and modal.

![FIGURE 5: The ELAN interface showing the coding scheme to be used for the analyses of beat gestures (Other Articulators tier not shown). The current example was taken from Keith (2007) at 3m 32s, where the images (from left to right) correspond to the preparation, onset of stroke, apex, and retraction.](image)

Prosody will be automatically coded for pitch accents and boundary tones. The AuToBI automatic coding system will be used for English (Rosenberg, 2010) and the Analor automatic coding system will be used on the French dataset (Avanzi, Lacheret & Victorri 2008).

As of the writing of this research plan, the gestures produced during a total of 9 minutes of speech from the David Keith TedTalk (Keith, 2007) has already been manually
coded following the abovementioned scheme. After manually coding one hour of gesture annotation (around December 2018, see section 6 with PhD work plan), an automatic coding algorithm will be developed in a prospective collaboration with the Natural Language Processing Research Group (TALN) at the Department of Information and Communication Technologies at Universitat Pompeu Fabra (Drs. Mireia Farrús & Dr. Leo Wanner). This algorithm will be designed to automatically detect certain phases of the beat gestures, namely the stroke and apex. This automatic analysis will then be run on the rest of the corpus, and then be double-checked manually and corrected for any mislabeling issues that are found. This method will allow for gestural annotation to be done on a large scale and within a short period of time.

5.1.4. Data Analyses

A series of analyses will be carried out in order to answer the two main research questions in the study. First, how do beat gestures temporally align with prosodic structure in two languages that structurally encode prosodic prominence in different ways from the point of view of phrasal positioning and tone type? Following Loehr (2012) and others, an initial analysis will be carried out on the entire corpus between the temporal relationship between the location of the apex and the peak of the pitch accents. In French, the temporal alignment of pitch accents will be distinguished between initial rises and final accents (annotated as Hi and H*/L* in Figure 3, respectively). The second research question is whether are there any systematic correspondences between the manual form of beat gestures and their pragmatic function between the two languages of study.

5.1.5. Expected Results

Regarding the first research question, we expect that for both languages, most of the gestural strokes will co-occur with pitch accents (more than 85%), and there will be a tight temporal alignment with gesture apexes (within 300ms of each other), as described in Loehr (2012). There might be an interesting differential effect of pitch accent type (nuclear vs. non-nuclear) and position (phrase-initial vs. phrase-final) in the two languages. In the French subset, it is believed that a larger number of gestural strokes will align with the initial pitch rise than the final full pitch accents. Regarding the second question, it is believed that these gestures will be primarily marking focus in English speech, while in French they will mainly serve a rhythmic marking or emphatic function.
5.2. Study 2: The neurolinguistic processing of beat gestures in native and non-native speech

5.2.1. Research Questions

This study will explore the role of pitch accentuation with and without beat gestures, in both first and second language. Previous findings suggest that beat gestures improve attentional processes, manifesting as a P3a component, and ease semantic processing in speech, which manifests as a decreased N400 component (e.g., Dimitrova et al., 2016; Wang & Chu, 2013). According to Drijvers & Özyürek (2018), non-native listeners process iconic gestures stronger than native speakers in clear speech. This study will determine whether or not the same is true for L2 speech processing in regards to beat gestures.

5.2.2. Participants

A total of 32 participants will take part in the study. Participants will give written informed consent after being briefed about the nature of the experiment and the future use of the data. The participants included 16 native English speakers with an intermediate level of Catalan, and 16 native Catalan speakers with an intermediate level of English. The L2 level was self-reported in a short questionnaire (see Appendix A) that asked participants to evaluate their language ability in five areas (listening comprehension, reading comprehension, speaking fluency, pronunciation, and writing) on a scale of one to four (one being native-like and four being least proficient). All participants reported no significant hearing, vision, or language issues, no history of neurological impairment or trauma, and were not taking any medication for psychological disorders.

5.2.3. Materials

The experimental materials consisted of total of 240 short dialogues were filmed in both Catalan and English, consisting of a question and an answer. The question sequences

---

6 I would like to give a special thanks to Chloé Monnier, Olga Kushch, and Evi Kiagi for their help with preparing the material and placing the triggers. I would also like to thank Salvador Soto-Faraco for his guidance and for giving me the resources to carry out the experiment, Marta Papai for helping me with running the pilots and becoming independent in the CBC lab, and Xavier Mayoral for his technical assistance.
introduced a choice between two contrastive answers (e.g., “Did the teacher give books or magazines to the students?”). **Figure 6 (top panel)** shows a screenshot of a question sequence. In the answer sequence, one of the two possible answers was given, serving as the focused element and the target word. (“The teacher gave BOOKS to the students”, CAPITALIZED words indicate the presence of a contrastive pitch accent.). The stimuli were recorded by two bilingual Catalan/English speakers (one male and one female) in a sound attenuated room. The target word was recorded in three different conditions. In the beat gesture condition, (**Figure 6, bottom left panel**), the target word was produced with both a pitch accent as well as a beat gesture. In the pitch accent condition, (**Figure 6, bottom center panel**), the target word was produced with a pitch accent but no gesture. The baseline condition consisted of answers with neither pitch accent nor gesture (**Figure 6, bottom right panel**). Of the 240 question/answer dialogues, 120 of them were answered by the male actor and the other 120 were answered by a female actress. Further, half of the target words for each speaker were in the first position in the question, while half were in the second position. The beat gesture used in all of the stimuli was a bi-manual, up-and-down movement of the hands with palms facing each other, performed in the central gesture space.

Each item was edited in **Adobe Premiere Pro CC 2017** in order to blur the faces of the actors and to place the audio triggers. The audio trigger consisted of a short beep which signaled to the EEG recording program (Brain Vision) the moment that a stimulus occurs. The videos were saved as an .mp4 file, with audio saved in the 5.1 sound system. This allowed for the placement of triggers on two separate audio channels, which ensured a precise synchrony between the video, the audio, and the triggers. Three triggers were placed on the target word: One at word onset (created by putting overlapping beeps on both audio channels 1 and 2), one at the onset of the stressed syllable (created by putting one beep on audio channel 2), and one at the beat gesture apex (created by putting one beep on audio channel 1). In this way, the program used to capture the EEG data could differentiate between the different stimuli: the word onset trigger was denoted as either “WO” when the first syllable was not accented or “WS” when the first syllable was accented. A word-medial stressed syllable as denoted as “SS” and the beat gesture apex as “G” Each marker was subsequently followed by a number that corresponded to the experimental condition. It is important to note that in order to keep a more natural aspect of the stimuli material, neither the video nor the audio were manipulated to perfectly align beat gestures with pitch accents. In this manner, stimuli are as close to real-world conversation as possible.
FIGURE 6: Screenshots of a question/answer discourse sequence presented to participants during the production of the target word 'book'. The top panel shows the question sequence “Did the teacher give books or magazines to the students?” The bottom panel shows the answer sequence “The teacher gave books to the students” in the beat gesture condition (left), the pitch accent condition (center), and the no beat condition (right).

In addition to experimental items, there were also a total of 120 filler items to serve as a distraction for the participants. There were a total of 60 fillers in each language. The fillers differed from the experimental items in that the answer was a third option, not present in the original question. Further, the answer was produced with either 1 or 2 beat gestures. As these are simply a distraction, EEG measurements were not taken during filler items. All in all, a total of 1560 videos were created. Finally, target and filler stimuli were distributed across 6 lists of 360 items each, according to the latin-square procedure so that no participant would see more than one version of the same dialogue (i.e., each list contained each experimental item in only one of the two languages and one of the three conditions). Each list was further
balanced for speaker gender (male/female), and answer position (first/second). Each list was then separated into 4 balanced blocks by language (English – Catalan – English – Catalan or vice-versa). Fillers were added and each block according to language. Each block was then pseudo-randomized individually, ensuring that one filler appeared at least every five stimuli, and that each condition appeared at least every ten stimuli. Additionally, no condition could occur more than two times in a row. Finally, ten filler items were placed at the beginning of the first bloc in each list, and a fixation cross lasting one second was placed in between each item.

In order to capture behavioral data regarding information memorization by participants and to induce an attentive state to the experimental materials, quizzes were created for each block (see Appendix B for a copy of the quiz that was given to participants after watching the first block of lists 1-3). In paper format, the quizzes consisted of 24 items balanced for condition, answer position, and speaker gender. The quiz was given to the participants immediately after completing each block, where quiz items were presented in a random order. Each quiz item was presented as a multiple choice question: it showed the question as originally asked in the stimuli, followed by the two possible answers in the same order as they were presented in the question. Participants were asked to circle the correct answer given in the video.

5.2.4. Experimental Procedure

Participants were presented with the stimuli in an electrically shielded, sound attenuated room in the Center for Brain Cognition lab (see Figure 7). The stimuli were displayed on an LCD screen and audio was presented via headphones. They were instructed both verbally and by an initial instruction screen that they were to watch the videos attentively, as they were going to take a quiz after each block to see how many answers they could remember. The experimental material was then presented to the participants, being played as a playlist (.m3u file) with VLC media player. An initial instructions screen was shown for 30 seconds, followed by a one second fixation cross, and then alternations of experimental material and fixation crosses thereafter. In the first block, the first 10 items were filler items, which allowed the participant to get used to the stimuli. Each block lasted on average about 15 minutes, and after each block, participants were given a 5 minute break, during which they had to complete the memory quiz. Participants were remunerated 10 € per
hour. With sessions lasting about 3 hours (1h 30 m for setting up the electrodes, 1h 30m for running the actual experiment).

![Figure 7: Images of the room where EEG recording took place.](image)

5.2.5. Electrophysiological Recordings

Electrophysiological data was recorded using an ActiCAP system and Brain Wave Visualizer to record the EEG data. Data was collected from 64 electrodes placed in a standard configuration. An electrode placed on the tip of the nose was used as the online reference, Cz as the ground reference, and both mastoids for offline referencing. Impedances were kept below 10kΩ, and the EEG signal was filtered off-line using a 0.5 Hz high-pass filter and a 30 Hz low-pass filter. Eye blink and movement artefacts were removed by applying semi-automatic inspection on raw EEG data. Epochs from -1000 ms to 600 ms in relation to stressed syllable onset were extracted for analyses. In each condition, the grand average was obtained by averaging individual average waves.

5.2.6. Data Analyses

In order to determine the statistical difference between each condition, cluster based random permutation tests were performed (Maris & Oostenveld, 2007) to control for type 1 error. First, each electrode is tested with a simple dependent t-test. Spatially adjacent electrodes that exceed a 5% significance level are grouped into clusters. Then, a cluster-level test statistic is done using the sum of the t statistics of each electrode. Third, the conditions are
randomly assigned to one of two sets (assuming no difference between conditions), creating a null distribution. Finally, observed cluster-level statistics are compared to the null distribution, and all clusters that fall within the highest or lowest 2.5% are cases where the null hypothesis can be rejected.

5.2.7. Expected Results

It is believed that the occurrence of pitch accents with gesture will elicit a P3a component, leading to additional attentional improvements, as reported by Dimitrova et al. (2016). It will also reduce the N400 component in native listeners, facilitating semantic integration of the stimuli, as reported by Wang & Chu (2013). We expect similar modulations for the P3a component across speech/gesture conditions in second language perception, however it is believed that larger amplitudes of both the P3a and N400 components under will be elicited, implicating that non-native listeners rely more heavily on sensory input than native listeners. It is hypothesized that this will be reflected in the behavioral data, where the questions from the beat gesture condition will be remembered more than the prosody only condition and the no prominence condition, respectively.

5.3. Study 3: The effects of beat gesture on recall and comprehension in native and non-native listeners

5.3.1. Research Questions

Previous studies have found contradictory results when investigating the effects of beat gestures on recall and comprehension in both adult and children populations in the first language (for positive results, see Llanes-Coromina et al., 2018; Igualada et al., 2017; for negative results, refer to Feyereisen, 2006; Macoun & Sweller, 2016). For the second language, Kushch et al. (2018) found positive effects of viewing beat gesture in word learning by non-native listeners. However, to our knowledge no study has looked at the differential effect of beat gestures on recall and comprehension between native and non-native listeners (see Dahl & Ludvigsen, 2014 regarding iconic gestures).

The aim of this study is to determine (a) whether or not beat gestures in speech input will facilitate recall and comprehension in a narrative-drawing task; (b) whether non-native listeners will make a more effective use of non-verbal markers than native listeners. A
between-subject experiment involving a narrative drawing will test the effects of using or not using beat gestures.

5.3.2. Participants

A total of 96 participants will be asked to take part in this study. Of these participants, 48 will be intermediate-level (B2) French learners of English, and the other 48 participants will be intermediate-level (B2) English learners of French. Participants will be recruited from Nantes University.

5.3.3. Materials

The materials for the experiment are adapted from Dahl & Ludvigsen (2014) and will consist of short narratives (in both English and French) describing the scenes in a short comic strip (see Figure 7 for an example comic strip with its accompanying English narrative). Embedded in the narratives will be a set number target items. These target items will consist of a set amount of either discourse structure markers or contrastively focused words. The narratives will be filmed by a native bilingual actor or actress in the two languages and under the following two conditions. The first condition will have prosodic prominences (L+H* pitch accents) on target items, and the second condition will have both prosodic prominence (L+H* pitch accents) and beat gestures on the target items. Thus a total of 4 videos will be created per comic strip (2 conditions for each language).

5.3.4. Experimental Procedure

Each group of native and non-native participants will be split up into two sub-groups of 24, one for each condition. Following Dahl & Ludvigsen (2014: 818), participants will be instructed as follows:

First you will view a short video clip. The clip is a description of a short story from a comic strip. Listen to the description, and create a picture in your mind of what this comic strip looks like. You are not allowed to draw while you are listening. After the description, try to draw the comic strip that you just heard described. Try to draw big and fill the page. The quality of your drawing skills is not the most important thing. What is important is how much you remember of the comic strip that was described and that you show that through what you draw. Try to include as much as possible in the drawing. In case something is hard for
you to draw or some element in the drawing seems unclear in the comic strip, you can write and draw arrows next to the element to clarify what it is.

Participants will then watch the video narrative in their assigned condition, which will be projected on the white board in the classroom. After watching the video, they will be given time to complete their illustration of the narrative event. The drawings will then be collected for analysis.

5.3.5. Data Analysis

Following Dahl & Ludvigsen (2014), recall will be measured by checking off all of the included target elements in the drawing that were originally presented in the video. Based on these elements, it is possible to tabulate four indicators for comprehension: Explicit recall (the number of items explicitly described in the narrative), implicit comprehension (number of elements included that are derived from logical inferences), distortion (elements that were explicitly described or logically implied but incorrectly placed in the drawing), and composite comprehension (the sum of explicit and implicit scores, to be interpreted together with amount of distortion).

5.3.6. Expected Results

It is hypothesized that the presence of beat gestures in both languages (and in both native and non-native speech) will lead to greater scores of both implicit comprehension and explicit recall, and fewer distortions. As for effects of nativeness, it is further hypothesized that beat gestures will show the greatest benefits in non-native speech.

5.4. Study 4: The effect of beat gestures in language input on phonological memory in the L2 in a shadowing task

5.4.1. Research Questions

Few studies have quantitatively explored the benefits of observing and imitating beat gestures to teach pronunciation (see Gluhareva & Prieto 2017; Kushch et al. 2018; Llanes-Coromina et al., 2018). However, these studies either trained participants by passively observing beat gestures, or were not formally trained but asked to move their hands. Study 4
will instead focus on actively training participants to produce synchronized beat gestures during a repetition task. The goal of this study is to determine the effects of online enactment of beat gestures on phonological awareness in the second language, which can ultimately be reflected in pronunciation. Thus the research question for this study is: does the online, synchronized reenactment of beat gestures as afforded by the shadowing task lead to better second language pronunciation measures?

5.4.2. Participants

The 48 non-native speakers from Study 3 will be asked to also take part in Study 4.

5.4.3. Materials

A short English narrative will be produced with a set number of embedded target words in each. The words will range in ease of pronunciation depending on a number of phonological factors (e.g., word stress placement, vocalic reduction, segmental aspects, etc.). Both narratives will be produced by an English actor in two conditions. The first condition will be produced with rising pitch accents (L+H* pitch accents) on target items, and the second condition will have both prosodic prominence (e.g., L+H* pitch accents) and beat gestures on the target items. The full length videos will then be edited to obtain videos that correspond to shorter fragments for the shadowing of smaller segments (see section 5.4.4. below). Figure 8 below shows an example narration for the shadowing task.

<table>
<thead>
<tr>
<th>LOVE RINGS TWICE by Bill Loguidice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theirs was a true love, thought Tony, one to stand the test of time. Eva, his soul-mate, was somehow even more beautiful than the day they first met, he realized. And even though she always seemed to say something interesting, it ultimately didn't matter, as he was happy just to hear her soft, melodic voice. Someday soon, he imagined, they would have children together, and their love would blossom as a family. These wonderful thoughts made him feel warm and tingly inside. Suddenly, without warning, Tony was yanked from his daydream by the doorbell. He let out a long sigh, realizing that that would be the courier with the divorce papers.</td>
</tr>
</tbody>
</table>

FIGURE 8: An example narration for the shadowing task

5.4.4. Experimental Procedure

In a between subject design, participants will participate in a narrative shadowing task. Before participating, all subjects will take a musical aptitude test (Profile of Music Perception...
Skills, Law & Zentner, 2012) to control from any bias stemming from musical ability. Participants will be divided into two training groups: training with beat gestures, or “the beat group” versus training without beat gestures, the “no-beat group”. For the pretest, students will be given a transcript of the narrative individually. They will be given a few minutes to familiarize themselves with the text. Then, the participants will be instructed to read the narrative aloud. After the pretest is completed participants will be asked to participate in a shadowing task. The participant will be shown segments of the original videos in their condition and be told to repeat the sentences while trying to imitate the speaker’s pronunciation. While the “no-beat” group will be mimicking the pronunciation, the “beat group” will be given an additional instruction – they are to also mimic the body language of the speaker in the video (i.e., produce beat gestures). Each phrase will be presented 4 times. After the training, the participant will be given the transcript a second time re-read and re-familiarize themselves with the text a second time. After, they will watch the entire video a second time, and will be encouraged to continue imitating the pronunciation quietly in their head. It is important to note that at any moment the participant is watching the video, the text will not be available. This is done in order to ensure that their attention is focused on the audio-visual stimuli as opposed to the text. After this training session, the participants will be asked to reproduce the narrative for a confederate who is not familiar with the narrative. The entire session will be video and audio recorded for later analyses and speech rating.

5.4.5. Data Analysis

The audio recordings from the pre- and posttest will be rated by 6 native speakers of English for various aspects of pronunciation, particularly accentedness, comprehensibility, and fluency. Raters will be asked to rate these measures on a Likert scale of 0 to 7. Raters will be presented with the individual words alone to rate accentedness and comprehensibility. Then they will be presented with the entire narrative in order to rate fluency.

5.4.6. Expected Results

We expect that the beat gesture group will improve their pronunciation ratings, specifically their accentedness, comprehensibility, and fluency scores more than the no-beat group.
6. Work plan

Previous work (September 2017 – September 2018)

- Establish collaboration for Study 1 with the Natural Language Processing (TALN) research group at the Department of Information and Communication Technologies at UPF (Drs. Mireia Farrús & Leo Wanner)
- Development of beat gesture labelling protocol to be used in Study 1
- Establish collaboration for Study 2 with the Multisensory Research Group at the Center for Brain and Cognition at UPF (Lab Director: Salvador Soto-Faraco)
- Organization of experimental design for Study 2
  - Stimuli randomization, distribution, creation of lists, etc.
- Completion of stimuli materials for Study 2
  - Filming, re-editing, placing triggers
- Carrying out of pilots and data collection for Study 2 in the Center for Brain and Cognition lab at UPF.
- Poster presentation at L2 Phonetics and phonology of L1 Romance Learners (L2PHROL): Turin, Italy, November 16-17. Title: Grounded cognition of prosodic elements via gesture: Preliminary evidence in a French learner of English.

September 2018 – December 2018

- Submission of the PhD Research Plan
- Construction of a corpus for Study 1
- Manual gesture annotation of English Ted Talk corpus (1 hr)
- Continue data collection for Study 2
- Preliminary data preprocessing for Study 2
Artifact rejection, grand averaging, etc.

- Defense of the PhD Research Plan (September 17th, 2018)
- Complete development of stimuli for Study 3
- One month stay in Université de Nantes (December)

**January 2019 – April 2019**

- Begin running automatic annotations for Study 1, manually double checking accuracy
- Statistical analyses for Study 2
- Pilot tests of stimuli for Study 3
- Begin developing stimuli for Study 4

**May 2019 – July 2019**

- Continue running automatic annotations for Study 1, manually double checking accuracy
- Begin writing for Study 2
- Continue developing stimuli for Study 4

**September 2019 – December 2019**

- Attendance to the International Congress of Phonetic Sciences (ICPhS) in Melbourne (Australia). 4-10 August, 2019.
- Continue running automatic annotations for Study 1, manually double checking accuracy
- Begin data collection for Study 3
- Pilot testing for Study 4
- Four month stay in Université de Nantes (December)
- Attendance to Conference EuroSLA 2019 - Location and Dates TBD

**January 2020 – April 2020**

- Data Analysis of Study 1
● Continue data collection for Study 3
● Begin Data collection for Study 4

May 2020 – July 2020

● Attendance to Conference International Society for Gesture Studies (ISGS) – Location and Dates TBD
● Writing for Study 1
● Data analysis for Study 3
● Data analysis for Study 4

September 2020 – December 2020

● Writing for Study 3
● Writing for Study 4
● Begin writing for PhD thesis
● One month stay in Université de Nantes (December)

January 2021 – April 2021

● PhD Thesis writing

May 2021 – July 2021

● PhD Thesis Defense

7. References

7.1. Selected References


Dimitrova et al. (2016) is one of the few studies that investigated the integration of beat gestures with experimental stimuli. This study confirmed Biau & Soto-Faraco’s (2013)
findings that beat gestures elicit an early positivity, and both attribute this effect to increased attention. Dimitrova et al. (2016) is particularly important for Study 2, as the aim is to reproduce the study for L1. What sets Study 2 apart from Dimitrova et al.’s study is that we will explore these same effects in L2 speech. Thus the experimental stimuli follow the same format of Question-Answer sequences with contrastive focus. However, there are a few important differences in our Study 2. First, the original article contained twice as many experimental conditions (six in total: no gesture/beat gesture/grooming gesture x focus/no focus). Our study was not interested in the integration of grooming gestures, and we decided not to include a condition [beat gesture / no focus] as this is very unnatural (see sections 1.2.2. and 1.2.3.1. above). Thus our study contains only 3 experimental conditions. Secondly, the original article manipulated the experimental videos so that the beat gesture apex always occurred at word onset (with the gesture onset occurring 520ms before target word onset). Study 2 does not manipulate the videos in order to obtain natural speech/gesture timing. Finally, the original article time-locks the ERP waveform to gesture onset. The article makes no mention of the prosodic structure of the target words. For Study 2, initial data analyses will be carried out with ERP waveforms time-locked to the onset of the stressed syllable, regardless of where it occurs in the target word.


Dahl & Ludvigsen (2014) is another study that is very important for the current PhD thesis in two ways. First, this study laid the groundwork for the main hypothesis behind two of the proposed studies in this thesis: that gestures may be more beneficial to non-native speakers than to native speakers. Indeed, their study found that non-native listeners correctly recalled and comprehended the narrative when referential gestures were present more than their native counterparts. Secondly, the methodology for Study 3 is adapted from this study. The main difference is that non-referential beat gestures are used in the narrative discourses instead of referential gestures. Also, the original study used single images in the narrative drawing task, whereas Study 3 will use a comic strip with 3-4 frames in order to show narrative development. In this way, beat gestures in the discourse will be appropriate
according to their pragmatic function (information and narrative structure, see section 1.2.3.2.)


The study by Drijvers & Özyürek (2018) supplements the behavioral findings from Dahl & Ludvigsen (2014) by offering neurological evidence that suggests the stronger processing of gesture by non-native listeners than native ones. Drijvers & Özyürek (2018) include a number of conditions and report on the differences in integration between matching and mismatching gesture in clear and adverse speech conditions by native and non-native listeners. The finding that is most pertinent to this thesis is that in clear speech with matching gestures, non-native listeners showed a larger N400 component than that of their native counterparts. They say “non-native listeners possibly recruit the visual semantic information more than native listeners, which is possibly due to the fact that they pay more attention to gestures when they are unsure about their language proficiency” (p. 15). Additionally, the results of Study 2 (regarding attention to beat gestures) may corroborate the idea that it is this increased attention that leads to better performance in recall and comprehension (Study 3).


The study by Gluhareva & Prieto (2017) is the only study to the author’s knowledge to directly study the impact of watching beat gestures on pronunciation. The study found that training while passively watching second language input with beat gestures improved pronunciation in discourse demanding-situations. These findings suggest that beat gestures may serve as a bootstrapping device for language learners regarding pronunciation, as beats highlight the prosodic structure of lexical items. Study 4 of this thesis will add to these findings with an active training via a shadowing task. Having participants imitate speech
online may lead to more robust results concerning the effects of beat gestures on pronunciation.


Prieto et al. (2018) proposed expanding the definition of beat gestures from simple flick up-and-down of the hand (McNeill, 1992). The authors explain that beat gestures can come in a variety of hand shapes, with complex phasing structures and a variety of pragmatic functions in speech. The current PhD thesis will adopt this expanded definition for beat gestures. Thus the beat gestures considered for analyses, particularly for **Study 1**, will include any gesture regardless of hand form or phasing structure, as long as it is not referential to speech in any clear, discernible manner. Additionally, the authors propose a coding scheme to independently code for beat gesture form, temporal relationship with prosody, and pragmatic function. A modified version of this coding scheme will be adopted for **Study 1**, allowing for the robust analysis of manual beat gestures in two languages.
7.2. Full Bibliography


APPENDIX A: Language questionnaire

Language Survey

Full Name ________________________________        Telephone______________________
E-mail___________________________________________   Age_______________________
Place of birth ___________________________  Current place of residence___________________
If this is not where you were born, how many years have you been living there? ______
What language did you grow up speaking (your first language)? ______________________
Please list below any other languages you speak. For each language, please rate your ability in each area on a scale of 1 to 4 (1 being native-like and 4 being the least proficient):

Language 1:_______________________
      | Listening comprehension | 1 | 2 | 3 | 4 |
-----|--------------------------|---|---|---|---|
Reading comprehension       | 1 | 2 | 3 | 4 |
Speaking fluency            | 1 | 2 | 3 | 4 |
Pronunciation               | 1 | 2 | 3 | 4 |
Writing                     | 1 | 2 | 3 | 4 |

Language 2:_______________________
      | Listening comprehension | 1 | 2 | 3 | 4 |
-----|--------------------------|---|---|---|---|
Reading comprehension       | 1 | 2 | 3 | 4 |
Speaking fluency            | 1 | 2 | 3 | 4 |
Pronunciation               | 1 | 2 | 3 | 4 |
Writing                     | 1 | 2 | 3 | 4 |

Percentage of your day spoken in this language: _____ Percentage of your day spoken in this language: _____
**Language 3:**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening comp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading comp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaking fluency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronunciation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Percentage of your day spoken in this language: _____ Percentage of your day spoken in this language: _____

In which language do you feel the most comfortable? _______________________

How old were you when you started learning Catalan? _______________________

How long have you been living in a Catalan-speaking environment?_____________

Are you right-handed or Left-handed? ______________________________________

Do you have any significant vision, hearing, or language issues? ______________

Have you had any history of neurological impairment or trauma? ______________

Do you or have you ever taken any medication for any psychological disorders, such as schizophrenia, bipolar disorder, or psychosis? ______________________________________
APPENDIX B: Memory Quiz
Block A.1

Please read through the questions and circle the answer that was given in the videos
Si us plau, llegeix les preguntes i marcar la resposta que ha escoltat en el video

1. El teu germà toca el violí o el piano a l'escola de música? –
   el violí / el piano

2. Vols demanar pizza o pasta per sopar? –
   pizza / pasta

3. Has vist les piràmides o els museus a Egipte? –
   les piràmides / els museus

4. Prefereixes ajudar-me amb les matemàtiques o amb la literatura? –
   Matemàtiques / Literatura

5. Els científics han descobert una tomba o un esquelet al desert? –
   una tomba / un esquelet

6. T'agrada llegir novel·les o contes quan estàs sola? –
   novel·les / contes

7. L’Anna parla l’alemany o l’espanyol a la feina? –
   l’alemany / l’espanyol

8. Netejaràs la cuina o el bany demà? –
   la cuina / el bany

9. La policia va detenir un lladre o un drogoaddicte? –
   un lladre / un drogoaddicte

10. Estàs imitant les vocals o les consonants de l’alemany? –
    les vocals / les consonants

11. Li agraden els espaguetis o les patates per dinar? –
    els espaguetis / les patates
12. Sucaràs la galeta o la xocolata a la llit? –
la galeta / la xocolata

13. Repararàs la porta o la cadira primer? –
la porta / la cadira

14. Prefereixes veure comèdies o drames quan estàs de mal humor? –
comèdies / drames

15. La Susi va llençar la bossa o la cartera al sofà? –
la bossa / la cartera

16. Penjaràs un quadre o un cartell a la paret? –
un quadre / un cartell

17. Li donaràs un llibre o una revista als teus amics? –
un llibre / una revista

18. La Liz va jugar a tennis o al basquet l’estiu passat? –
tennis / basquet

19. Has rebut una carta o un correu electrònic del teu amic? –
una carta / un correu electrònic

20. Em donaràs monedes o targeta per a pagar? –
monedes / targeta

te / cafè

22. T’agrada nadar o córrer a l’estiu? –
nadar / correr

23. El mestre va donar llibres o revistes als estudiants? –
llibres / revistes

24. Vas a veure un documental o una comèdia al cinema aquesta nit? –
un documental / una comèdia
Appendix C: Sample comic strip narrative for Study 3

Two cats are on a fence. There is a bush behind a **white** cat towards the left. A second cat on the **right** has black spots on his **head**, his **left** leg, and his **tail**. The **spotted** cat is angrily hissing at the other cat, who looks worried. Then, the spotted cat moves **closer** to the white cat and has his **right** paw raised in the air, as if he’s about to attack. The cat on the **left**, still worried, has his **left** hand behind his back, out of view. Next, we see the cat on the **left** has pulled out a spray bottle from behind his back and is squirting the **spotted** cat, which is jumping in the air and has a surprised face. Finally, we see the **white** cat lying down on the fence with a pleased look on its face.

*Images taken from the comic strip “Simon’s Cat” by Simon Tofield. Target items in the text are indicated by bold and underlined type. Target items may change in function of early pilot data for experimental materials.*