USING THE HANDS TO EMBODY PROSODY BOOSTS PHONOLOGICAL LEARNING IN A FOREIGN LANGUAGE

Florence Baills

TESI DOCTORAL UPF / 2021

DIRECTOR DE LA TESI

Dra. Pilar Prieto Vives

DEPARTAMENT DE TRADUCCIÓ I CIÈNCIES DEL LLENGUATGE

To Pedro, Nuria, and Guillem.

À ma famille.

Acknowledgments

This shouldn't be the most challenging part of the writing of a thesis but somewhat it is to me! Why… I am so afraid to sound conventional when I am unmeasurably in debt with all the people that will be mentioned in the next few lines, and so profoundly grateful. This is not a common feeling, I can tell you, to think about all of you while I am writing these lines. And this is why, while you will be reading your name, I wish that you can experience this feeling that someone esteems you for the wonderful person you are and how I honestly rely on you with all my trust.

My first and most important acknowledgement goes to my supervisor Pilar Prieto. Since I started the Master's thesis under your care, it has been a beautiful journey, and I am so grateful to you. I have learned so much! Thank you for this opportunity, thank you for sharing your knowledge, thank you for allowing me to discover how to do research and to train on the necessary skills, thank you for nurturing my curiosity, and most importantly, thanks for being this beautiful, smart, admirable person!

Next, I would like to warmly thank the GrEP team, the ones that are still here: Ïo, Júlia, Marussia, Patrick, Xiaotong, Yuan - you are the dream team, and the ones that have already left: Alfonso,

Ingrid, Iris, Joan, Núria, Olga, Peng, Santi - the masters of sea! This has been a wonderful experience thanks to your friendship and you are all amazing people with beautiful minds!

I want to give a big, big thank you to the French teachers at UPF, with whom I have collaborated happily during those years, and who have made Study 2 a reality: Élisabeth Miche, Guilhem Naro, Lydia Fernández, merci beaucoup de m'avoir fait confiance et de m'avoir soutenue!

I also think about the community of researchers from UPF and beyond that I have met at some point during these years, at seminars, conferences and workshops. Thank you for your interest, your suggestions and for sharing your knowledge! Special thanks to Clément François, Joan-Carles Mora, Núria Esteve-Gibert, and Wolfram Hinzen, for their helpful comments on previous versions of this work. Thanks a lot to the verbotonal experts Charlotte Alazard-Guiu, Michel Billières, and Lorraine Baqué for their precious collaboration and their insights on the verbotonal method.

I have felt privileged to be offered a grant by the Department of Translation and Language Sciences at UPF. So, thank you to the DTCL and UPF, from the bottom of my heart, because, you know, this was my only option for a grant and without this financial support, there would have been no thesis. In addition, it has been a

pleasure and an enriching experience to teach our undergraduate and graduate students and I have been able to continue doing my job thanks to you.

Santi, you have been my guide during my first steps as a doctoral student, and I learned so much from you, thank you for your patience!

Yuan, Xiaotong, Peng, I am grateful to have been able to accompany you in your research. Thank you for allowing me! You do an awesome job. I really appreciate you guys!

Ïo, Júlia (and Sara), the GrEP is like good wine, it is getting better and better as time goes by! I'll be watching your next steps.

Ingrid, Marussia, Patrick, you have supported me and bore with me during all those years. I love you, friends!

Pedrito, Nurieta, Guillemet, ya sabéis que sois mis favoritos. Gracias a vosotros disfruto de cada día, y tengo tanta suerte. Os quiero.

As the tongue speaketh to the ear,
so the gesture speaketh to the eye”
Bacon, 1891

“[..] understanding the mind/brain means studying it in the body, and understanding the embodied mind means studying it in the world; and this is simply because the mind is in the body and the world. If SLA studies is a cognitive science—or seriously desires to become one—shouldn’t it follow suit?”

Atkinson, 2010

Abstract

Prosodic features of language such as prominence, melody, and rhythm, are frequently embodied by hand movements in face-to-face communication. However, little is known on the role of embodied techniques encoding the melodic and rhythmic features of speech on the phonological learning of a foreign language. The main goal of this thesis is to unveil the benefits of using a prosody-based, multisensory approach (visual, auditory, and kinesthetic) to support not only the learning of such prosodic features but also the overall pronunciation of a foreign language.

Three training studies with a pre- and posttest design have been included in the thesis, which assess the role of multisensory training through the perception and production of visuospatial hand gestures and percussive hand movements in the acquisition of prosodic features and general pronunciation of a foreign language and with a variety of populations and proficiency levels. The first study shows that training Mandarin Chinese tones with pitch gestures (that is, visuospatial hand gestures representing pitch movement) favors the recognition and the recall of novel words with these tones by Catalan naïve learners more than training without pitch gestures. The second study shows that training Catalan intermediate learners of French with phrase-level prosodic gestures (that is, a type of visuospatial hand gesture embodying

intonation, rhythm, and phrasing at the sentence level) helps them improve their accentedness and production of suprasegmental features in a discourse reading task more than training without phrase-level prosodic gestures. Finally, the third study shows that visually and acoustically highlighting the syllabic structure and rhythmic properties of French words with hand-clapping during training helps Catalan naïve learners of French improve their accentedness and final lengthening measures more than training without hand-clapping. Together, these findings expand our knowledge on how embodied multisensory techniques highlighting prosodic features can support phonological learning and underline the need to use reliable practical and embodied techniques for pronunciation instruction.

Resum

En la parla espontània, és freqüent que els trets prosòdics del llenguatge, com ara la prominència, la melodia i el ritme, s'expressin a través dels moviments de les mans. Tot i això, tenim poc coneixement sobre el valor d'emprar aquests moviments de les mans que representen els trets melòdics i rítmics de la parla per a millorar l'aprenentatge fonològic d'una llengua estrangera. L'objectiu principal d'aquesta tesi doctoral és analitzar els avantatges d'utilitzar un enfocament multisensorial (visual, auditiu i cinestèsic) basat en la prosòdia per millorar no només a l'aprenentatge d'aquestes característiques prosòdiques en una llengua estrangera, sinó també la seva pronunciació global.

Aquesta tesi doctoral inclou tres estudis experimentals amb un disseny pre- i post-test que avaluen l'efectivitat d'un entrenament multisensorial en l'adquisició dels trets prosòdics i de la pronunciació d'una llengua estrangera. Concretament, s'estudia l'efecte de la percepció i la producció de gestos i de moviments manuals percussius en una varietat de poblacions i de nivells de competència. El primer estudi demostra que un entrenament amb gestos manuals visuoespacials que representen els moviments melòdics dels tons del xinès mandarí afavoreix el reconeixement i la memorització de paraules noves que contenen aquests tons per part d'aprenents catalans, comparat amb un entrenament sense

gestos melòdics. El segon estudi mostra com la participació en un entrenament amb gestos prosòdics a nivell de frase (és a dir, un tipus de gest manual que visibilitza l'entonació, el ritme i el fraseig a nivell de frase) ajuda estudiants catalans amb un nivell intermedi de francès a millorar el seu accent en francès i la seva pronunciació dels trets suprasegmentals en una tasca de lectura. Finalment, el tercer estudi demostra que ressaltar visualment i acústicament l'estructura sil·làbica i les propietats rítmiques de paraules en francès picant de mans ajuda els nens catalans a millorar el seu accent estranger i a pronunciar l'allargament final del francès més adequadament que un entrenament sense picar de mans. En conjunt, aquests resultats amplien el nostre coneixement sobre les tècniques multisensorials i corporeïtzades, destaquen la importància de la visibilització de les característiques prosodiques d'una llengua per a estimular l'aprenentatge fonològic d'una llengua estrangera i subratllen la necessitat d'utilitzar tècniques corporeïtzades en l'ensenyament de la pronúncia d'una llengua estrangera .

Resumen

En el habla espontánea, es frecuente que los rasgos prosódicos del lenguaje, como la prominencia, la melodía y el ritmo, se expresan a través de los movimientos de las manos. Sin embargo, tenemos poco conocimiento sobre el valor de utilizar estos movimientos de las manos que representan los rasgos melódicos y rítmicos del habla con el fin de mejorar el aprendizaje fonológico de una lengua extranjera. El objetivo principal de esta tesis doctoral es analizar las ventajas de utilizar un enfoque multisensorial (visual, auditivo y kinestésico) basado en la prosodia para mejorar no solo el aprendizaje de estas características prosódicas, sino también su pronunciación global.

Esta tesis doctoral incluye tres estudios experimentales con un diseño pre- y post-test que evalúan el valor de un entrenamiento en la adquisición de los rasgos prosódicos y de pronunciación de una lengua extranjera. Concretamente, se estudia el efecto de la percepción y producción de gestos y de movimientos percusivos de las manos con una variedad de poblaciones y niveles de competencia. El primer estudio demuestra que un entrenamiento con gestos manuales visuoespaciales que representan los movimientos melódicos de los tonos del chino mandarín favorece el reconocimiento y la memorización de palabras nuevas con estos tonos por parte de participantes catalanohablantes, en comparación

con un entrenamiento sin gestos de tonalidad. El segundo estudio muestra cómo la participación en un entrenamiento con gestos prosódicos a nivel de frase (es decir, un tipo de gesto manual visuoespacial que visualiza la entonación, el ritmo y el fraseo) ayuda a estudiantes catalanohablantes con un nivel intermedio de francés a mejorar su acento francés y su pronunciación de los rasgos suprasegmentales en una tarea de lectura. Finalmente, el tercer estudio demuestra que resaltar visualmente y acústicamente la estructura silábica y las propiedades rítmicas de palabras en francés haciendo palmadas ayuda a los niños catalanohablantes a mejorar su acento extranjero y a pronunciar el alargamiento final del francés más adecuadamente que un entrenamiento sin hacer palmadas. En conjunto, los resultados anteriores amplían nuestro conocimiento sobre las técnicas multisensoriales y corporeizadas, además destacan la importancia de dar visibilidad a las características prosódicas de una lengua para estimular el aprendizaje fonológico de una lengua extranjera y finalmente subrayan la necesidad de incorporar técnicas corporeizadas en la enseñanza de la pronunciación de una lengua extranjera.

Table of contents

1

CHAPTER 1: GENERAL INTRODUCTION

Prosody in language, whose etymology in classic Greek refers to the song that accompanies a musical instrument (προσῳδία: *pros* (πρός) "together" + *oide* (ᾠδή) "song") has often been referred to as "the music of speech". In music, notes are combined to form beats and phrases and in turn, each of these elements determine the timing, pitch and volume of the musical paragraphs. Similarly, speech prosody stems from the combination of phonemes organized into larger units, such as syllables, prosodic words, and intonational units, which are responsible for the timing, pitch, and intensity characteristics of the discourse. In this work we adopt the view that (a) prosodic patterns are an essential building block of foreign language; and (b) prosodic patterns can be experienced in the same way as we often experience music, not only with our ears but also by moving our whole body, our head, or our hands: we perceive the music and we move along the rhythm and the melody. Similarly, *embodying* prosody with our hands should enhance our perception and production patterns of the phonology of a foreign language. In

the present work, we will use the term ‘embodiment’ as an extension of embodied cognition, i.e. how the perception and production of body movements influence knowledge and learning. The aim of the present dissertation is therefore to assess the impact of “prosodic embodiment” through hand visuospatial gesture and percussive movements, on the phonological learning of a foreign language.

The present dissertation includes three training studies with a pre-and posttest design which involve interdisciplinary research in the areas of embodied cognition, gesture studies, phonetics and phonology, and second language acquisition. As a consequence, this dissertation relies on more than one theoretical framework that will be assessed in the present introduction. First, the theory of Embodied Cognition and the premises for embodied learning are presented, with a focus on the particular role of hand movements and gestures. Second, an explanation follows on the general framework for foreign language phonological learning and how embodied methods favoring multisensory techniques have been either applied in the classroom or empirically tested within a perception-production paradigm, with a focus on prosody. Finally, the scope of and the main goals of the thesis are explained in light of the reviewed literature.

1.1 Embodied cognition, gesture, and their benefits for learning

1.1.1 The Embodied Cognition hypothesis

Traditionally, cognitive science views cognition - the ability to acquire knowledge and develop understanding - as an abstract information process in the mind that manages the brain's modal systems for perception (e.g., vision, audition), action (e.g., movement, proprioception), and introspection (e.g., mental states, affect). One of the most widely accepted frameworks, e.g. the computational theory of mind, views the human mind as an information processing system, a computational system that is physically implemented by neural activity in the brain (e.g., McCulloch & Pitts, 1943; Piccinini & Bahar, 2013; Rescorla, 2020). This theory posits that input (e.g. the mental representations or symbols of a stimuli) are fed into a processing unit and based on a finite set of rules, an output is produced, i.e. cognition. Such a model assumes that knowledge resides in a semantic memory system that is separated from the perceptual and motor systems. Conceptual representations are solely abstract and symbolic computations, and cannot contain information in the sensory and motor system. Therefore, any motor activity related to a representation would have to go through some sort of 'interface'.

From the perspective of embodied cognition, however, there is no such separation. With an early influence of the phenomenologist philosophical tradition (e.g., Merleau-Ponty, 1945), the different theories gathered under the umbrella term of Embodied Cognition represent first and foremost a criticism of Cartesian dualism, according to which the mind is entirely distinct from the body and can be successfully explained and understood without reference to the body or to its processes. On the contrary, Embodied Cognition theories postulate that cognition is not exclusively centralized in the brain but also dependent on the motor and sensorial systems and the physical interaction between the body and the environment (e.g., Barsalou, 2008, 2010; Fincher-Kiefer, 2019; Gallagher, 2005; Lakoff & Johnson, 1999; Shapiro, 2019). According to the Embodied Cognition hypothesis, however, “conceptual processing *already* is sensory and motor processing” (Mahon & Caramazza, 2008, p. 60).

In opposition to the view that representation has to go through an "interface" (see above), followers of a strong view of the Embodied Cognition hypothesis claim that there is no interface between a concept and the sensory/motor system and the process of concept retrieval would simultaneously trigger the process of retrieving sensory and motor information. However, because of its restrictive domain of application (manipulable objects or concrete actions) and the lack of neuroscientific evidence to support it (e.g., Gennari, 2012; Meteyard et al., 2012), this extreme vision of embodiment

has been challenged by more conciliatory descriptions of the theory (e.g., Meteyard et al., 2012, see Farina, 2021 for a recent comprehensive review of approaches within the embodied cognition paradigm). One influential alternative was proposed by Mahon and Caramazza (2008) and is named the g*rounding by interaction hypothesis*, according to which sensory and motor information provide an enhanced, richer version of conceptual representations. According to this view of embodiment, for the same representation, there is a level of abstract conceptualization, which can stand alone and does not need motor and sensory information, and there is also a level of grounded conceptualization, gathered from diverse sensory experiences. What we know about the world is the result of the interaction between both levels:

> "The activation of the sensory and motor systems during conceptual processing serves to ground 'abstract' and 'symbolic' representations in the rich sensory and motor content that mediates our physical interaction with the world" (Mahon & Caramazza, 2008, p. 68).

Common to all Embodied Cognition approaches is the idea that sensorimotor interactions are critical for both the development and the maintenance of cognitive capacities (Engel et al., 2013). A central notion in the Embodied Cognition paradigm is the process

of reenactment, or simulation, of perceptual, motor, and introspective states acquired during previous experience in contact with the world, body, and mind (e.g., Barsalou, 2008; Decety & Grezes, 2006; Foglia & Wilson, 2013; Goldman, 2006). Barsalou (2008) described the process of reenactment as follows:

> "As an experience occurs (e.g., easing into a chair), the brain captures states across the modalities and integrates them with a multimodal representation stored in memory (e.g., how a chair looks and feels, the action of sitting, introspections of comfort and relaxation). Later, when knowledge is needed to represent a category (e.g., chair), multimodal representations captured during experiences with its instances are reactivated to simulate how the brain represents perception, action, and introspection associated with it." (p. 618)

Crucially, reenactment is performed in the cortex areas related to motor actions (Gallese & Lakoff, 2005) and follows two steps: first, an online cognitive process which involves the perception and memorization of an experience, and second, an offline cognitive process which involves the activation of the experience at a later time.

Neurally, the reenactment principle can be explained with reference to the properties of the mirror neuron system (MNS; Gallese,

2005). This group of neurons respond both to action observation and to action execution and is activated upon watching another person perform some behavior (e.g., Rizzolatti et al., 1996; Rizzolatti, 2005). Moreover, Fu & Franz (2014) found that the MSN directly encodes viewer perspective during embodied human actions, suggesting that action observation automatically evokes internal imagery representations of the same action (Calvo-Merino et al., 2006). The MNS may therefore play important functional roles in understanding the actions produced by others and their intentions, and it is assumed to form the basis of the human capability to learn through imitation (e.g., Rizzolatti & Craighero 2004; Gallese et al., 2004). Furthermore, it has been suggested that the MNS is the basic neural mechanism from which language developed (Rizzolatti & Arbib, 1998).

Early evidence for the Embodied Cognition paradigm - i.e. evidence that the sensorimotor interactions participates in cognition - comes from theories and experiments involving action language. In the classic book *Metaphors we live by*, Lakoff & Johnson (1980) proposed their *conceptual metaphor theory*, according to which abstract concepts can be expressed through metaphorical expressions based on bodily experiences and actions such as 'you are running out of time' or 'argument is war'. Later, different types of studies started to unveil the link between language and bodily actions. Glenberg & Kaschak (2002) looked at the *action-sentence compatibility effect.* They asked participants to judge the

grammaticality of sentences implying a toward or away movement from the body (‘Close the drawer’, ‘Liz told you the story’) by responding on a device that required moving toward or away from their actual body. Results showed faster reaction times when the movements in the sentence and in the response were matching in terms of direction. Crucially, the effect was observed for sentences describing the transfer of both concrete and abstract concepts (see also Glenberg et al., 2008, for similar results). Myung et al. (2006) also found that participants made faster decisions about a target word (‘piano’) when a related word in terms of manipulation knowledge was presented as a prime (‘typewriter’) compared to an unrelated priming word (‘bucket’). Rieser et al. (1994) found that linguistic tasks related to spatial orientation are facilitated by the mental representation of movement both in children and adults. Descriptions of spatial associations between a character and an object (“After doing a few warm-up exercises, *he put on his sweatshirt* and went jogging”) were comprehended faster than those of spatial dissociations (“After doing a few warm-up exercises, *he took off his sweatshirt* and went jogging”) (Glenberg et al. 1987) and words with high ‘body-object interaction’ ratings (Siakaluk et al. 2008) or related to manipulable objects (Rueschemeyer et al., 2010) were recognized faster, providing further evidence of the role of embodiment on lexical-semantic processing.

Crucially, neurophysiological studies have endorsed the link between action and language. Studies showed that when uttering action words, the motor and premotor areas of the brain are activated (e.g. Hauk et al., 2004; James & Maouene, 2009; Pullvermüller & Fadiga, 2010; Pullvermüller, 2013), even when processing non-literal action language (e.g. Yang & Shu, 2016). Pulvermüller et al. (2005) applied transcranial magnetic stimulation to motor areas while asking participants to make lexical decisions on action words related to arm or leg movements and found faster reaction time when the brain area corresponding to the limb involved in the action word was stimulated. Conversely, when performing sensorimotor actions, brain areas for language are activated (e.g. Desai et al., 2010). Gentilucci & Dalla Volta (2008) reviewed behavioral and neuroimaging evidence showing bi-directional influence between arm movements and speech, and documented the existence of the same motor system for both modalities (see also Willems & Hagoort, 2007, for a review of the neuroscientific evidence on the relationship between language, gesture, and action). All these findings show that embodied cognition approaches provide a convincing conceptualization and explanation of some language processing patterns.

In addition to mental experience of reenactment, embodied cognition has also assessed the cognitive effects of body movement itself. Because humans have limited information-processing abilities, they exploit the environment to offload cognitive

demands. For example, Glenberg and Robertson (1999) showed that participants who were allowed to indexically link written instructions to objects in the environment during a learning phase performed better in a compass-and-map task than subjects who were not. Additionally, Risko and Gilbert (2016) observed that cognitive demands are also sent "onto the body": for example, in order to see a rotated picture, one may prefer to tilt the head to normalize the orientation instead of performing a mental rotation. Wilson (2002) claimed that cognitive off-loading is not restricted to spatial tasks and cited all the learning and reasoning strategies used in mathematics that involve external devices. Wilson (2002) also emphasized that off-loading "need not be deliberate and formalized, but can be seen in such universal and automatic behaviors as gesturing while speaking" (p. 629, see section 1.3 for empirical evidence on the off-loading effect of gestures). For example, iconic gestures have been shown to lighten a speaker's cognitive load both in the presence and the absence of the depicted item (e.g., Ping & Goldin-Meadow, 2010).

Another important concept related to body and articulatory movement is that of human imitation. Nocaudie (2019, p. 35) proposed a general definition of imitation as the - either voluntary or unconscious - reproduction of part or whole of a behavior after being perceived in another subject, so that it is possible to perceive the reproduction by the imitator as resembling the production of the model. Donald (1993) proposed that humans possess a mimetic

skill or mimesis, resting on the ability to produce conscious, non-linguistic representational acts by imitation, and that mimesis may be the one of the first cognitive abilities in the human species. Research has confirmed that human beings are extremely talented at imitation (e.g., Brass & Heyes, 2005; Carpenter & Call, 2009; Chartrand & Bargh, 1999) and can unconsciously and very accurately imitate the verbal and non-verbal behaviors of conversational partners (for reviews, see Heyes, 2011; Lakin et al., 2003; Pardo et al., 2017). In the context of phonology, according to Giles et al. (1991), individuals adapt to each other's behaviors in terms of a wide range of linguistic, prosodic, and nonverbal features to accomodate to peers and facilitate communication (see also McCafferty, 2008, for a review on gesture mimesis and language learning). Phonetic convergence seems to be triggered by an unconscious mimetic behavior to deliberately develop, amplify and regulate phonetic variation (e.g., Delvaux et al., 2004; Miller et al., 2010; Pardo, 2006; Nielsen, 2011; see Coles-Harris, 2017, for a review on phonetic convergence).

As seen above, body movements and the imitation of body movements constitute important features of embodied cognition. Some early studies have explored the role of motoric enactment on memory (e.g., Cohen, 1981; Saltz & Donnenwerth-Nolan, 1981). Cohen (1981) compared production and observation testing participants on their ability to recall actions following training under three conditions: Participants either performed the actions,

observed the experimenter performing the same actions, or simply heard and read the descriptions for these actions. He found that participants remembered actions better when these were performed either by themselves or by the instructor than when the actions were simply described verbally. Saltz & Donnenwerth-Nolan (1981) showed that motoric enactment is effective in sentence recall because it leads to the storage of some type of motoric trace or image. Notwithstanding, Engelkamp et al. (1994) showed that self-performed tasks led to superior memory performance in recognition tasks for longer lists of items (24–48 items) but not for shorter lists (12 items). It is important to note that Embodied Cognition does not clearly posit that doing an action would benefit more than mere observation of an action. Action observation also leads to the formation of motor memories in the primary motor cortex, which is considered a likely physiological step in motor learning (Stefan et al., 2005).

Multisensory integration designates the processes in the brain that allow us to take information we receive from the world through our five senses (sight, sound, touch, smell, self-motion and taste) and integrate it in our nervous system, organize it, and respond to it appropriately (e.g. Camponogara & Volcic, 2021; Stein & Meredith, 1993; Stein et al., 2009). The result of multisensory integration is the coherent representation of the world, creating meaningful perceptual experience, and leading to coherent adaptive behavior (e.g., Lewkowicz & Ghazanfar, 2009). For instance, we

can perceive the spatial information of an object (e.g., its length, height and size) by looking at it or touching it. Studies have shown that patterns of multisensory integration develop progressively across the life span. While audio-visual integration emerges late in the first year of life, between 8 and 10 months (Neil et al., 2006), haptic-visual integration only reach adult-like integration measures from eight years old onward (Gori et al., 2008, see Burr & Gori, 2012 for a review). Interestingly, Nardini et al. (2010) highlighted that even if children present lower multisensory integration rates than adults, they process sensory information from different sources separately faster than adults. Recently, Greenfield et al. (2017) tested four to eleven year-old children and found evidence that haptic-visual integration is refined with age in terms of both time and space. Multisensory integration also explores how different sensory modalities interact with one another and alter each other's processing, as demonstrated by multisensory illusions such as the McGurk effect (McGurk & MacDonald, 1976), the rubber-hand illusion (Botvinick & Cohen, 1998) or the body-transfer illusion (Petkova & Ehrsson, 2008). An early-observed effect of multisensory integration is decreasing reaction times when stimuli are presented in multiple simultaneous senses rather than when the same stimuli are presented in isolation (e.g., Forster et al., 2002; Hershenson, 1962; Hughes et al., 1994). Other studies have suggested that training with one modality may improve another, in particular, visual cues have proven helpful to

assist auditory speech processing (e.g., Atligan et al., 2018; Atilgan & Bizley, 2021; Helfer & Freyman, 2005).

In this dissertation, we take on board the perspective of embodied cognition and multisensory integration with the objective of enhancing phonological learning. In other words, our training paradigms will call upon three different senses (auditory, visual, and kinesthetic) and will also involve reenactment, motoric activity and their combination in an imitation paradigm. In the following section, we review studies on the benefits of embodied cognition for learning and in particular in the specific area of language learning.

1.1.2 Benefits of embodied cognition for learning

Decades ago, development psychologist Jean Piaget argued that sensorimotor experiences were essential for infant cognitive development (Piaget, 1952). There is a general consensus that infants are embodied learners: they use their senses and their body to gather information about their surrounding world (e.g., Laakso, 2011). However, whereas Piaget suggested that this would only apply at an early age, other authors have proposed that sensorimotor interactions with the environment continue to be important for language processing and increased conceptual understanding throughout children's cognitive and physical development (e.g., Gibbs, 2006, Thelen et al., 2001) and that these embodied experiences become more refined and flexible over time (e.g., Antonucci & Alt, 2011; Kontra et al., 2012). In a review article, Wellsby and Pexman (2014) emphasized the importance of sensorimotor experience in development and presented studies showing the role of embodied experiences for children's development of concepts and word learning, as well as language processing. Regarding word learning, their review highlighted that specific kinds of embodied experiences may be useful for learning different classes of words (nouns, verbs, and adjectives). Once infants are able to manipulate objects, they are able to use their senses to gather information about them and understand their functions (Smith, 2013) and map labels, i.e. words, onto representations based on these manipulations (Scofield et al.,

2009). All the sensory experiences with the world are later reenacted and influence infants' categorization decision of novel objects (e.g., Smith, 2005, Smith et al., 2007).

Importantly, Wellsby and Pexman (2014) suggest that it may be necessary for the sensorimotor information obtained through interaction to be directly related to the information learned to trigger learning. In other words, the embodied experience needs to be appropriate and relevant to the material to be learned (see also Kiefer and Trumpp, 2012). As an example, Glenberg et al. (2004) showed that manipulating toy objects referred to in a text or simulating these actions (imagined manipulation) both helped second-grade children understand and better memorize elements of the text compared to multiple readings. Regarding language processing, Wellsby and Pexman's review (2014) further gathered evidence on the beneficial effects of embodied training on early reading comprehension by poor readers (e.g. Marley et al., 2010) and during children's language processing during offline tasks (e.g., Engelen et al., 2011).

Beyond the effects of embodiment on natural cognitive development, embodied cognition is claimed to have special relevance for education (e.g., Ionescu & Vasc 2014; Macedonia, 2019, Shapiro & Stolz, 2019). Empirical research about embodied cognition and learning has primarily focused on how increasing the student's own motor involvement in a lesson boosts learning

outcomes (e.g., Bahnmueller et al., 2014; Smith et al., 2014). Research in classroom teaching methodologies documents increased performance and better concentration when active learning and the use of communicative gestures are involved (e.g., Craig & Amernic, 2006). Strong evidence of the benefits of an embodied approach in educational contexts has been found in particular for learning mathematics (e.g., Abrahamson & Sánchez-García, 2016; Hutto et al., 2015; Nathan & Walkington, 2017; Newcombe & Weisberg, 2017; Núñez et al., 1999; Pouw et al., 2014). In two review articles, Kiefer and Trumpp (2012) and Madan and Singhal (2012) underlined the benefits of embodied cognition through actions, gesture and physical exercise for memory tasks, as well as reading and writing tasks.

Physical activity naturally pertains to the possible application of embodiment in education and its beneficial effects have been reviewed extensively. This line of research has predominantly observed the physiological changes induced by single or multiple bouts of physical activity and their effect on cognitive functioning (e.g., Donnelly et al., 2016, for a review). Sports are claimed to have beneficial effects on cognition by facilitating learning and memory (e.g., Hillman et al., 2008; Liu-Ambrose et al., 2012). In their review article, Erickson et al. (2015) found that fitter and more active children showed a range of physiological benefits, performed better on tasks that require executive control and associative memory, and showed higher academic achievements.

There is solid evidence that physical activity positively correlates with cognitive performance, though modulated by the type of activity, level of intensity, duration of exercise, aspects of cognition, and learner characteristics (for reviews, see Barenberg et al., 2011; Y. K. Chang et al., 2012; Erickson et al., 2015; Fedewa & Ahn, 2011; Sibley & Etnier, 2003; Tomporowski et al., 2008). In general, the effect of physical activity on cognitive performance is greater for children in elementary and middle school. The largest effects were found on perceptual skills, followed by IQ, academic achievement, and math and verbal tests. Short bouts of exercise increase response speed and accuracy (Tomporowski, 2003), improve working memory capacity (Pontifex et al., 2009), and performance on free-recall tasks (Coles & Tomporowski, 2008).

An important application of embodied learning through self-performed body movement can be found in the field of music education (e.g. Juntunen, 2016; Romero Naranjo, 2013).The Dalcroze music pedagogy aims to develop abilities, such as sense of rhythm, finesse of hearing, and spontaneous expression that are vital to a competent musician (Juntunen, 2016). Jaques-Dalcroze (1920) sought an multisensory approach to music education that involves both the mind and the body of students learning to play musical instruments, in order to develop and improve the faculties that are used when engaging in music: the aural, visual, tactile, and muscular senses. All these senses are called upon through individual body movements and group activities, acting as a

physical metaphor for musical elements in order to learn musical concepts (Greenhead & Abron, 2015; Juntunen & Hyvönen, 2004). The exercises used in a Dalcroze-inspired classroom include the following categories of movement: functional (e.g., showing a pitch level with the hand), rhythmic, creative, dramatic, and dance (Abril, 2011).

> "Through movement of the whole body, music is felt, experienced, and expressed; reciprocally, the movements express what the participants hear, feel, understand, and know." (Juntunen, 2016, p. 142)

Different studies have reported a positive impact of Dalcroze exercises on the ability to recognize and respond to rhythmic patterns, demonstrate beat competency, and develop rhythm aptitudes among kindergarten and first- and second-grade children (Blesedell, 1991; Joseph, 1982; Rose, 1995). Crumpler (1982) found a significant improvement of first-grade children's melodic and pitch discrimination abilities after participating in Dalcroze exercises, whereas a control group that did not do such exercices did not show any improvement.

Orff, a direct disciple of Dalcroze, developed a specific method to teach music involving body percussion, the art of striking the body to produce various types of sounds, and created activities bringing together the spoken word with body percussion (Keetman & Orff,

1963). Interestingly, a basic implementation of body percussion, e.g. hand-clapping to songs, has been found to be beneficial from an educational perspective, both within and outside the classroom (e.g., Brodsky & Sulkin, 2011; Harwood, 1993; Marsh, 2008; Riddel, 1990). Hand-clapping to songs involves simultaneous seeing, hearing, and touching as well as motor experience executed by the arms, hands and palms. Interestingly, while hand-clapping to songs, the synchronisation of verbal and movement sequences demands the integration of language and motor production systems (Sulkin & Brodsky, 2007). Brodsky and Sulkin (2011) found that children who were more skillful at performing hand-clapping to songs were more efficient learners at school and performed better in hand/rhythm synchronization, verbal memory and handwriting.

In the realm of second language acquisition, one early application of embodied cognition comes from the Total Physical Response (TPR) method for word learning (Asher, 1969). In this method, instructors introduce new words by demonstrating their meanings using the body and subsequently prompt learners to repeat the same motions with their own bodies in response to words. Some evidence suggests that new words taught to beginning adult learners in classroom settings via the TPR method can be learned just as effectively as L1 words learned in naturalistic settings by children (Asher, 1972; Asher & Price, 1967). More recently, Mavilidi et al. (2015) explored the effects on memorization of enacting every day action words such as ‘fast’, ‘dance’, ‘soccer’

through whole-body movements (i.e., physical exercise related to the meaning of speech) and part-body movements (i.e., referential gestures). One hundred eleven preschool children learned 14 Italian words during a 4-week training program in one of four conditions: integrated physical exercise (related to the words), gesturing (enacting the actions indicated by the words while seated), conventional (verbally repeating the words while seated), and non-integrated exercise (unrelated to the learning task). They were tested for word recall during, directly after, and 6 weeks after training. Results indicated that children in the integrated physical exercise condition achieved the highest learning outcomes in terms of cued and free recall (see also Pesce et al., 2009).

Considerable research in the field of Conversation Analysis has documented how cognitive states are expressed in foreing language classroom interaction not only through speech but also via gaze, facial gesture, hand gesture, posture shift and the manipulation of documents and objects and how these embodied cognitive states participate in the management of peer interaction (e.g., Belhiah, 2009; Cekaite, 2009, 2015; Drew, 2006; Eskildsen & Wagner, 2013, 2015; Goodwin & Goodwin, 1986; Jakonen, 2020; Kääntä, 2015; Majlesi, 2015; Matsumoto & Dobs, 2017; Mori & Hasegawa, 2009). To give a few examples, Mori & Hasegawa (2009) showed how two students organized their actions during a word search activity by simultaneously using different semiotic resources and Jakonen (2020) suggested that teachers use their body in the

classroom as a pedagogical device after analysing teachers' movement trajectories and body positioning in Content and Language Integrated Learning (CLIL, teaching subjects such as science, history and geography through a foreign language). For example, the analysis showed that walking through the class allowed the teacher to monitor student individual and group progress during a task, to display availability and to invite students' interaction. Eskildsen & Wagner (2013) observed that the imitation of a speaker's gesture acts as a communicative resource to achieve and maintain understanding. Later, Eskildsen & Wagner (2015) analysed how gesture-speech combinations are created by second language learners to create a common understanding of new words and how they are reused at later occasions.

In a recent review article, Shapiro and Stolz (2019) argued for the necessity to empirically investigate the effects of "embodied education" and make all these findings available to teachers. The same authors pointed out an area of research within embodied cognition, i.e. the domain of gesture studies, that affords interesting applications for educational purposes. Consequently, the following section gives an overview of the field of gesture studies and the relevance of the multimodal communicative system constituted by speech and gesture for learning.

1.1.3 Gestures as a window onto embodied cognition

Moving away from the wider field of "non-verbal communication", since the last few decades, the field of Gesture Studies has focused on the close link between gestures and speech, supporting the idea that gestures are part of language itself and that gesture-speech units create meaning, reflecting people's thoughts during verbal communication and modulating the interaction between speakers (e.g., Goldin-Meadow, 2010, 2011; Goldin-Meadow & Wagner, 2005). The fact that speech cannot be stripped of the accompanying gestures without compromising the meaning or function of the message (e.g., Graziano & Gullberg, 2018) prompted gesture theorists to advocate the existence of a unique system between gesture and speech (e.g., McNeill, 1992; Kendon, 2004).

Kendon (1980, 1982) pioneered the field with a first attempt to comprehensively categorize different types of manual gestures used in communicative situations. Consequently, McNeill (1992) lined up these gestures on a continuum named "Kendon's continuum" to distinguish all the different types of manual expressions, from gesticulations (later called co-speech gestures by McNeill) to sign languages. Co-speech gestures occur together with speech and are situated at the left end of the continuum. This type of gesture is "global and holistic in its mode of expression, idiosyncratic in form and users are but marginally aware of their use of it" (Kendon, 2004, p. 104–105). According to McNeill (1992), co-speech

gestures (also named gesticulations or gestures) include all the spontaneous movements of the hands and arms that are simultaneously produced together with speech. Along the rest of the continuum, pantomimes depict objects or actions to narrate a story; and emblems are conventionalized signs created in accordance with the rules of a particular group of users (e.g., placing the thumb and index finger in contact to produce the OK sign in agreeing with someone; McNeill, 1992, p. 38). Finally, at the right end of the continuum, sign languages refer to a complete natural linguistic system used by a specific community with identical linguistic properties as spoken languages.

An important feature of co-speech gestures is that they convey a communicative intention (McNeill, 1992) and must be distinguished from adaptors and self-grooming movements, which are other types of spontaneous non-meaningful bodily movements such as a movement performed when the speaker scratches his/her chin, touches his/her hair or his/her clothes, etc. (Ekman & Friesen, 1969). In addition, co-speech gesture is the only point along the continuum where gestures convey meaning by combining properties that are unique to their respective category: gestures possesses global and synthetic properties (i.e., they contain meaning only as a whole entity and meanings are synthesized into one symbolic form), whereas speech possesses segmented and analytic properties (i.e. words are combined to create a sentence,

"distinct meanings are attached to distinct words"; McNeill, 1992, p. 19).

In a classification system that became the most widely adopted for the field of Gesture, McNeill (1992, 2005) distinguished four major dimensions of co-speech gestures, namely iconic, metaphoric, deictic (or pointing), and temporal-marking gestures (also called 'beat' gestures or 'beats'), depending on their form and referential or semiotic functions. Iconic gestures represent properties of an object, an action or a scene and display a close relationship to the semantic content of the speech they accompany. Metaphoric gestures are "like iconic gestures in that they are pictorial, but the pictorial content presents an abstract idea rather than a concrete object or event" (McNeill, 1992, p. 14). Deictic gestures, also named pointing gesture, are typically performed by pointing at something with a finger to connect some aspect of speech to an object or location in space (it can also be an "abstract pointing" when referring to something or someone who is absent, or a place or a moment in time). Finally, temporal-marking gestures are rapid and repetitive rhythmic movements of the arms, hands, fingers and are typically associated with prosodically prominent positions in natural discourse, remarking "the word or phrase they accompany as being significant […] for its discourse pragmatic content" (McNeill 1992, p. 15). McNeill defined beat gesture as a two-phases movement "up and down, or back and forth" (p. 15).

All four types of gestures possess discourse-pragmatic functions, however, iconic, metaphoric, and deictic gestures pertain to the group of referential gestures, i.e. they confey specific semantic information about a referent, whereas beat gestures are considered non-referential, i.e., they do not encode specific semantic content. Speakers naturally accompany their speech with beat gestures in a way that are able to extend the auditory prosody to the visual modality, helping them to structure their speech and emphasize relevant information (Prieto et al., 2018; Shattuck-Hufnagel & Ren, 2018). As pointed out by Shattuck-Huffnagel and Ren (2018), "the term 'beats' suggests a degree of rhythmic periodicity, invoking a conductor beating out the rhythm of an orchestral performance, and non-referential gestures have sometimes been defined in these terms, as e.g., beating out the rhythm of the speech" (p. 2). It is important to emphasize that most gestures can be characterized by several of the dimensions mentioned above (McNeill, 2005); for example, a gesture can be both pointing and metaphorical (when pointing to the future to the right on an imaginary horizontal temporal axis). Kendon (2017, pp. 167-168) proposed six pragmatic functions of gestures: *referential*, when they contribute to the referential of propositional meaning of speech; *operational*, when they are related to what is expressed verbally (confirming, denying or negating it); *modal*, when they express the speaker's point of view on what is being expressed verbally; *performative*, when they refer to the speech act being realized; *parsing*, when

they distinguish certain components of the discourse; and finally *interpersonal*, when they refer to the role of the speaker or the organization of the conversational sequence.

There is nowadays a general consensus that gesture and speech form an integrated communicative system and are coordinated both temporally (at the phonological level) and from a semantico-pragmatic perspective (e.g., Bernardis & Gentilucci, 2006; Clark, 1996; Goldin-Meadow, 2003; Kelly et al., 2010; Kendon, 1980, 2004; Levinson & Holler, 2014; McNeill, 1992, 2005; 2016; Özyürek & Kelly, 2007; Özyürek et al., 2007; Peeters et al., 2017;see Kelly et al., 2008; Wagner et al., 2014, for reviews). In the Growth Point theory, McNeill (2005) pointed out that gestures "synchronize with speech at the point where the speech and gesture coexpressively embody a single underlying meaning, a meaning that is the point of highest communicative dynamism at the moment of speaking." (p. 1) and suggested that gesture and speech develop from the same "growth point". McNeill (2016, p. 21) describes the growth point as the minimal unit of gesture-speech integration, containing both both imagistic (imagery is understood as a symbolic form encoded by the gesture and determined by meaning) and verbal content (linguistically encoded information). Hence, an utterance comprises both an imagistic and a linguistic side, based on the same communicative intention. McNeill (1992) enumerated three rules that govern speech and gesture synchronization: First, the semantic synchrony postulates

that speech and gestures can cover the same idea or concept unit at the same time in a redundant or complementary manner, creating a richer picture. The pragmatic synchrony rule posits that gesture and speech serve the same pragmatic purpose. Finally, the phonological synchrony rule predicts that the stroke phase of the gesture (i.e., the mandatory phase of the gesture, as it contains its meaning and effort; Kendon, 1980; McNeill, 1992, 2005) is temporally aligned with the phonological peak syllable of speech.

It has been proposed that co-speech gestures are temporally synchronised with speech prosody. Without distinguishing between gesture types, Kendon (1980: 210-211) proposed a hierarchy of gestural structures, from gestural units to gesture phrases, functioning in parallel to a hierarchy of prosodic units, from discourse to tone groups. Interestingly, Bolinger (1983) drew the parallel between the up and down of intonation contours and ascending and descending movements of the head and the body. Research has shown clear evidence of a tight temporal alignment between the prominent parts of gesture and prominent parts of speech (e.g., Danner et al., 2018; Esteve-Gibert et al., 2017; Esteve-Gibert & Prieto, 2013, 2014; Krahmer & Swerts, 2007; Leonard & Cummins, 2011; Loehr, 2004, 2007, 2012; McClave, 1998; Pouw & Dixon, 2018; Shattuck-Hufnagel & Ren, 2018; for reviews, see also Rusiewicz & Esteve-Gibert, 2018, and Wagner et al. 2014). Regarding specific types of co-speech gestures, deictic gestures (e.g., Esteve-Gibert & Prieto, 2013, 2014) and head and

eyebrow movements (e.g., Esteve-Gibert et al., 2017; Keating et al., 2003; Krahmer et al., 2002; Krahmer & Swerts, 2007) appear to have a prominence-lending function. In addition, Parrell and colleagues (2014) found that during a speaking-and-finger-tapping synchrony task, modulating either the duration of a syllable (speech modality) or the magnitude of the finger-movement (kinematic modality) both lead obligatorily to a temporal adaptation of the other modality. Regarding non-referential gestures, there is evidence that beat gestures and speech prosody are integrated early on in speech processing (e.g., Biau et al., 2016) and gesture and prosodic synchronization has been observed with beat gestures (e.g., Leonard & Cummins, 2010; Krivokapić, 2014; Krivokapić et al., 2017; Shattuck-Hufnagel & Ren, 2018). Interestingly, Pouw et al. (2020) found a direct, physical effect of simple arm movements on phonation: producing up-and-down movements of the arm, hand, and finger while phonating had a direct impact on pitch production, with higher F0 peaks, even when participants were instructed to resist such an effect on their phonation. Therefore, such dependency may partly explain the strong link between the production of gestures and prosodic prominence. However, the observed synchronization is not always perfect (Colletta, 2004; Rohrer et al., 2019). In their review article, Wagner et al. (2014) found that the start of the gesture tends to slightly precede the start of the associated speech, whatever the type of gesture, and that temporal coordination tends to be anchored in prosodic structure of

speech by aligning with stressed syllables and prosodic boundaries (see for example Ferré, 2010 for a study on spontaneous speech in French).

Building on the abovementioned Growth Point theory (McNeill, 1992, 2005), several cognitive models have detailed how gesture and speech interact (see Goldin-Meadow & Alibali, 2013, and Wagner et al., 2014, for reviews). The Lexical Retrieval hypothesis (e.g., Krauss et al., 2000) claims that gesture plays a facilitative role at a later point in the speech production process (during the formulation stage; see Levelt, 1989) to help speakers to access items in the mental lexicon. Empirical support for this hypothesis comes from studies revealing that gesturing (e.g., Beattie & Coughlan, 1999, Pine et al., 2007) or tapping (e.g., Ravizza, 2003), help retrieve words during a tip-of-the-tongue (TOT) state (i.e., when the speaker knows the target word but can not actually remember it at that moment) and also help speech production in bilingual speakers (e.g., Nicoladis, 2007). However, the fact that gestures occur significantly more during fluent speech compared to disfluent speech, and that gestures produced during disfluent speech display a pragmatic function rather than being related to lexical retrieval shows that gestures do not merely present a compensatory function (Graziano & Gullberg, 2018).

According to the Information Packaging hypothesis (Kita 2000), speech and gesture interact at an early stage of speech production,

during the conceptualization of the message ("preverbal message"; Levelt, 1989). More specifically, when gestures encode visuo-spatial representations they help the speaker to select, package and organize speech related to visuo-spatial information, in other words, to verbalize perceptual or motor knowledge. Evidence shows that people tend to gesture more when the conceptualization of information is more challenging (Alibali et al., 2000). Moreover, it has been shown that low verbal fluency is related to an increase of gesture production only with speakers with high spatial visualization skills (Hostetter & Alibali, 2007). Kita and Özyürek's (2003) Interface Hypothesis suggested that speech and gesture are generated by separate systems, and interact in a bidirectional fashion during speech conceptualization and formulation: Gestures are generated during the conceptualization stage from spatio-motoric representations of the referent (i.e., action and spatial information) and organize this spatio-motoric information into a suitable form for speaking, according to the linguistic possibilities and constraints of a specific language (e.g., Özçalışkan et al., 2016). A recent expanded version of this theory, the Gesture-for-Conceptualization Hypothesis (Kita et al., 2017), proposes that speakers can activate, manipulate, package, and explore spatio-motoric information both for speaking and thinking through the use of referential gestures.

In one of the most renowned gesture production models, Hostetter and Alibali's Gestures as Simulated Action framework (2008,

2019) situates gesture production within a larger embodied cognitive system and argues that gesture production stems from spatial representations and mental images. Gestures arise when speakers simulate actions and perceptual states as they think, which in turn activate the motor system. The authors argue that because manual and vocal systems are linked, both developmentally (e.g., Iverson & Thelen, 1999) and neurally (e.g., Rizzolatti et al., 1988), movements of the mouth and vocal articulators for speech production are coupled with movements of the hands and arms. Ping et al. (2014) showed that moving arms and hands interfered with a listener's ability to use information conveyed in a speaker's hand gestures, suggesting that understanding gesture relies, at least in part, on the listener's own motor system. Therefore, motor activation arising from simulated actions is more likely to be expressed overtly in gestures when the motor system is also engaged in producing speech. Hostetter and Alibali (2008) thus claim that through gestures, cognition becomes visible.

Similarly to embodied cognition, an important body of research has explored the potential gains of using gesture for educational purposes, including both first and second language learning. The following sections aim at offering an overview on these studies.

1.1.4 Benefits of gesture for learning

Gesture and speech develop together in infancy, playing an important role in language development (e.g., Capirci & Volterra, 2008; Colletta et al., 2015; Goldin-Meadow, 2007; Iverson & Goldin-Meadow, 2005). Gestures have been found to appear before language (e.g., Volterra et al., 1979; Liszkowski, 2008) and to pave the way for later linguistic development (e.g., Morford & Goldin-Meadow, 1992; Capirci et al., 1996, 2005; Butcher & Goldin-Meadow, 2000; Özçalışkan & Goldin-Meadow, 2005). In particular, there is recent evidence that prosody and gesture develop together (see Esteve-Gibert & Guellaï, 2018; Hübscher & Prieto, 2019, for a review). For children, gestures are generally attributed a facilitating function, for example facilitating access to the lexicon (e.g., Pine et al., 2007). Gestures may also be considered as predictors of linguistic abilities: Recently, Vilà-Giménez et al. (2021) showed that the use of spontaneous beat gestures produced by 14- to 58-month-old children during their interaction with caregivers (parents and educators) predicts their ability to perform better structured narratives by the age of five.

Representational gestures, such as iconic and metaphoric gestures (i.e. gestures representing concrete or abstract information, respectively, see section 1.1.3), have been shown to be helpful in many different ways for learning, both from the perspective of the interlocutor and the speaker. Taking the interlocutor / gesture

perceiver stance, mathematical lessons with gestures are shown to promote deeper reasoning, synthesis, and information retention than lessons that do not feature gestures (e.g., see Goldin-Meadow, 2018; Goldin-Meadow & Alibali, 2013 for reviews). De Ruiter (2017) claimed that iconic gestures provide additional visual redundant information helping listeners to better perceive and understand speech, thereby enhancing communication. Sullivan (2018) argued that instructor movement and use of representational gesture stimulates the mental imitation by activating the mirror neurons, and leads to improved student academic outcomes. Regarding child development, studies have highlighted the positive role of teachers' gestures in the learning processes (e.g., Goldin-Meadow et al., 1999; Valenzano et al., 2003). There is evidence that representational gestures facilitate math lesson understanding (e.g., Congdon et al., 2017; Ping & Goldin-Meadow, 2008) and benefit the comprehension of complex syntactic and/or semantic structures (e.g., McGregor et al., 2009; Theakston et al., 2014). Interestingly, other work has also focused on the role of prosody in syntax and syntax learning through the lense of embodied interaction and gesture (e.g. Kreiner & Eviatar, 2014; Matsumoto & Dobs, 2017). Many studies have also reported benefits of observing iconic gestures for narrative comprehension, in both adults and children (Dargue & Sweller, 2018a, 2018b, 2020a, 2020b; Dargue et al., 2019, Macoun & Sweller, 2016).

From the gesturer point of view, there is solid evidence that self-performing gesture boosts problem-solving strategies, for example in mathematical tasks (e.g., Broaders et al., 2007; Cook et al., 2008; Goldin-Meadow et al., 2009; Novack et al., 2014) or in spatial thinking tasks (e.g., Alibali & Kita, 2010; Alibali et al., 2011). Regarding the effects of gestures on memory, both adults and children may benefit from gesturing. Goldin-Meadow et al. (2001) found that participants who were allowed to gesture during a dual task (memorizing letters or words while explaining math problems) could remember more items than those who did not and suggested that, by reducing cognitive load on the explanation task, gesturing allowed participants to allocate more resources to the memory task (see also Cook et al., 2012; Wagner et al., 2004). Cook et al. (2012) further compared the effects of producing meaningful hand gestures vs. producing meaningless hand movements and no gesture and found that participants could recall significantly more items when producing meaningful hand gestures. Furthermore, Ping & Goldin-Meadow (2010) found that being allowed to produce gestures helped children recall more words than not being allowed to do so, regardless of the presence or absence of the reference objects. Producing gestures has also been shown to significantly enhance creativity and the development of new ideas (e.g., Beilock & Goldin-Meadow, 2010; Kirk & Lewis, 2017).

> "Gesturing does not merely reflect thought: Gesture changes thought by introducing action into one's mental representations. Gesture forces people to think with their hands." (Beilock & Goldin-Meadow, 2010, p. 1609)

Furthermore, Kita (2000) showed that gesture performance facilitates the selection and organization of visuospatial information (e.g., to describe a set of actions or a range of objects) into units that are congruent with the sequential order of the speech. Kita et al. (2017) associated gestures with the speech planning process and posited that representational gestures facilitate speakers' conceptualization and consequently speech production. In addition, Krauss et al. (2000) suggested that gestures can help speakers retrieve words in the mental lexicon during speech production (see Beattie & Coughlan, 1999; Nicoladis, 2007; Pine et al., 2007, Ravizza, 2003, for empirical evidence). Focusing on the role of referential gestures, Hostetter (2011) analyzed 63 studies and described six ways in which referential gestures may boost memory, comprehension, and learning: (i) by being better adapted at conveying spatial information than speech, (ii) by giving additional information that is not in speech, (iii) by having positive effects on the speaker's speech production, (iv) by presenting information that is redundant with speech, affording listeners additional cues to glean meaning, (v) by capturing a listener's attention, and (vi) by boosting a positive rapport between speaker and listener.

Finally, there is some evidence that learners get better results at different memory and cognitive tasks when producing hand gestures rather than when only observing them (e.g., Cherdieu et al., 2017; Frick-Horbury, 2002; Goldin-Meadow, 2014; Goldin-Meadow et al., 2009; Goldin-Meadow et al., 2014; for a review of the effects of enactment and gestures on memory recall, see Madan & Singhal, 2012). Neurophysiological evidence also shows that self-performing a gesture when learning verbal information favors the formation of sensorimotor networks that contribute to the representation and the storage of words in a native (Masumoto et al., 2006) or in a foreign language (Macedonia et al., 2011). Regarding narrative skills, gesture production during a retelling task seems to help more than gesture observation for children (Cameron & Xu, 2011) but not for adults (Dargue & Sweller, 2020b). Notwithstanding, there is also evidence that producing gestures may add cognitive load on learners with lower skills or proficiency, when the task is too difficult (e.g. Post et al., 2013).

Non-referential hand gestures, such as beat gestures, i.e. hand movements which typically associate with prosodically prominent positions in speech but do not encode specific semantic content (see section 1.1.3), have also been shown to have a positive effect on adults' and children's ability to recall information (e.g., Austin & Sweller, 2014; Igualada et al., 2017; Kushch & Prieto, 2016; Llanes-Coromina et al., 2018; So et al., 2012). Mixed results have

been obtained regarding the role of observing beat gestures for narrative comprehension by children, either positive (Llanes-Coromina et al., 2018) or negative (Macoun & Sweller, 2016). However, in a training study by Vilà-Giménez et al. (2019) found that listening to stories and observing a narrator produce beat gestures favored narrative discourse performance in children. In addition, neurophysiological studies have revealed the positive influence of beat gestures in speech perception and comprehension (e.g., Biau & Soto-Faraco, 2013; Dimitrova et al., 2016; Hubbard et al., 2009), including the processing of syntactic (Holle et al., 2012) and semantic information (L. Wang & Chu, 2013). From these studies, it can be concluded that beat gestures play the role of an attention-catcher, leading to the activation of language-related brain areas.

Regarding the production of non-referential gestures, Lucero et al. (2014) found that participants who were asked to produce beat gestures while speaking were faster to utter target words compared to participants who were asked to produce iconic gestures and no gesture. Regarding children, Vilà-Giménez and Prieto (2020) found that listening to and watching storytellers who were producing gestures and additionally being encouraged to produce beat gestures when retelling the story helped children obtain better narrative performance scores than those who only observed the beat gestures. In a recent systematic review, Vilà-Giménez and Prieto (2021) confirmed the beneficial role of non-referential gestures in

terms of information recall, narrative comprehension and precursor of narrative abilities for children.

All in all, observing and producing referential and non-referential gestures has been shown to promote a series of cognitive and linguistic benefits and facilitate learning, both in children and adults. The following section (a) reviews the studies that have described and classified teachers' and learners' gestures as part of their linguistic conceptualization and expression in the foreign language classroom, and (b) explores the effects of embodied learning for the acquisition of a foreign language, in particular vocabulary recall, about which most research has been carried out.

1.1.5 Embodiment and gesture in foreign language learning

Studies dealing with the role of body movements and gestures in the foreign language classroom have first described the use of gestures by learners during their interactions (see Gullberg & McCafferty, 2008, for a review). Early work has looked at the effects of gesture rate, showing a greater frequency of gestures by second language learners than in the mother tongue (Jungheim, 1995; Kita, 1993; Nobe, 1993, cited in Gullberg, 1998, p. 77). Nobe (1993) in particular found an increase in the use of iconic, metaphorical, and beat gestures when speaking a foreign language. Gullberg (1998) observed that, when speaking in a second language, learners do not replace speech with gestures, but rather use them in coordination with speech. She also observed that concrete metaphorical and deictic gestures (referring to the immediate environment) were widely used to signal or remedy a problem with lexical knowledge/recall. She also noted that learners with lower language proficiency relied less on this strategy and favored abstract deictic gestures to overcome problems related to grammar or narrative skills. McCafferty (1998) described how learners overcome communication difficulties and cognitive difficulties using gestures during storytelling. In the same vein, Van Compernolle & Williams (2011) described how learners used gestures when speaking in a second language during the completion

of a shared task that required reasoning. In a recent study, Graziano and Gullberg (2018) found that adult L2 learners were more likely to produce referential gestures and ongoing gestures (unfinished or frozen) during disfluent speech in the L2 compared to disfluent speech in the L1 and that the function of these gestures was more pragmatic in nature (signaling a problem in communication related to lexical retrieval). Importantly, the authors observed that gestures were overwhelmingly more present during fluent speech than during disfluent speech.

From the teachers' perspective, Tellier (2008a) suggests that an informed second language teacher should consciously use her hands as a teaching tool, just like her voice. Co-speech gestures are oftentimes used by instructors to reinforce the meaning of oral explanations, to clarify the meaning of new words, or to establish cohesion in the speaking turn (Beliah 2013). Moreover, the messages transmitted by the teachers' non-verbal behaviors, including gestures, may have a significant impact on students in the foreign language classroom, not only linguistically but also because they transmit impressions, emotions and regulate social interactions and hierarchical relationships between teacher and learners (e.g., Chamberlin-Quinlisk, 2008; see also Allen, 2000, for a broader description of teachers' non-verbal communication in the language classroom). Sime (2006) proposed that teachers' gestures have the following three functions in the foreign language classroom: a cognitive function that helps the learning process; an emotional

function that allows the teacher to express their emotions and state of mind; and an organizational function that helps manage the classroom. Lazaraton (2004) analyzed the gestures of a teacher while explaining new vocabulary and observed a prolific use of gestures. She suggested that gestures are an essential component of communication in the classroom and that they play an important role in understanding vocabulary (see also Smotrova & Lantolf, 2013). Smotrova (2014) observed that the multimodal interaction between the language teacher and the learner, involving speech and gestures, had a positive effect on comprehension and on the learning of grammar, vocabulary and pronunciation. Furthermore, W. Wang & Loewen (2016) observed the presence of a variety of iconic, metaphorical, deictic and beat gestures used by teachers during explicit corrective feedback (see also Seo, 2021, on the gestures used for corrective feedback for lexical errors).

Co-speech gestures have a variety of pedagogical functions in the classroom. Tellier (2006, 2008a) proposed a classification of the gestures used by teachers in the foreign language classroom according to their function and named them *pedagogical* gestures. Pedagogical gestures include arm and hand movements, head movements, and facial expressions which are produced by the teacher with the intention of facilitating access to the knowledge which is presented orally. Based on the different roles of the teacher defined by Dabène (1984), Tellier (2006, 2008a) described three main pedagogical functions of gesture, namely to organize, to

evaluate, and to inform. While organizational gestures refer to classroom management, interaction and participation, evaluation gestures allow the teacher to congratulate, approve or point out an error, and information gestures refer to some specific information on any linguistic element during instruction. Within the latter, three categories are considered: gestures encoding grammatical information (for example, using a deictic gestures to indicate a verbal tense on an invisible chronological axis, or drawing two semi-circles facing each other with the index fingers to indicate that word order should be inverted), gestures encoding lexical information (mostly gestures illustrating a concrete or abstract referent) and gestures encoding phonological and phonetic information (for example, a rising hand movement to illustrate rising intonation in questions, or placing the fingers together to form a round shape while pronouncing the sound [o]). The information gestures in Tellier's classification (2008a) could be considered as referential gestures within McNeill's (1992) gesture classification proposal, and this is why the term has been adopted in literature testing this type of gestures on new vocabulary acquisition (see below).

A handful of experimental studies have looked at the effect of watching or imitating referential gestures on the memorisation of vocabulary in a foreign language (see Macedonia, 2014; Macedonia & von Kriegstein, 2012, for reviews). In her seminal study, Allen (1995) trained 112 American university students to learn ten French

expressions either by reproducing or watching representational gestures, while another group learned the expressions by repeating them orally. The results showed that the students who performed the gesture when learning the new expressions recalled the meaning of more items than the other groups. After two months, both the groups that performed and watched the gestures showed significantly less decay than the group who learned without gestures. However, this study did not assess how many expressions learners had remembered, but rather how many expressions they could translate. Tellier (2008b) and Porter (2012) found similar results with children learning new words either by performing iconic gestures or by watching illustrative pictures of these words. For example, Tellier (2008b) asked 20 young French children to learn eight English words (house, swim, cry, snake, book, rabbit, scissors, and finger). Four of the items were associated with a picture while the other four items were illustrated by a gesture produced by an instructor that the children saw in a video and then enacted themselves. The results showed that the enacted items were memorized better than items enriched visually by means of pictures.

In a series of studies, Macedonia and colleagues found that producing referential gestures while learning new words not only helped participants to remember words with concrete and abstract meanings but also facilitated the retrieval of these words when creating new sentences (Macedonia et al., 2011; Macedonia &

Knösche, 2011). Furthermore, Macedonia and Klimesch (2014) conducted a fourteen-month longitudinal classroom study with an artificial language. They trained university students to learn 36 words (nine nouns, nine adjectives, nine verbs, and nine prepositions). For 18 items, participants only listened to the word and read it. For the other 18 items, participants were additionally instructed to perform the gestures proposed by the experimenter. Vocabulary learning was assessed through cued native-to-foreign translation tests at five time points. The results showed that enacting iconic gestures significantly enhanced vocabulary learning in the long run.

De Nooijer et al. (2013) examined the facilitative role of producing referential gestures during the learning phase (when new information is encoded) and during the lexical task (when the information is retrieved). They asked 115 9-to-10 year- old Dutch children to learn three different categories of novel verbs in their L1: object-manipulation verbs (implying manual activation, e.g., to chisel), locomotion verbs (implying leg activation; e.g., to stride), and abstract verbs (e.g., to dismiss) in one of four conditions: no gesture imitation, gesture imitation during encoding (i.e. during the storage of words in memory), gesture imitation during recall, and gesture imitation during both encoding and recall. Participants were tested on vocabulary recall immediately after the training session and one week later. Results showed that only gesture imitation of the object-manipulation verbs facilitated recall significantly more

compared to no gesture imitation, and only during encoding or recall, but not for both. The authors explained their results by highlighting the strong link between language and gesture, which favored the memorization of verbs implying hand movements.

Looking now at gesture observation, Morett (2014) further explored the role of referential gestures on three interrelated cognitive processes subordinate to word learning in a foreign language: communication (i.e. the quality and efficacy of the interactions involving the target language and co-speech gestures), encoding, and recall. Fifty-two naïve participants learned 20 Hungarian words either by observing referential gestures or no gesture and then, they themselves had to teach the meanings of the words to interlocutors who were also unfamiliar with Hungarian, either by producing gestures or without gestures. All participants were then tested for their recall of the target words. The results showed that gesture facilitated all three cognitive processes and that gesture production was more effective than gesture viewing. Interestingly, with a similar design, Morett (2018) found that spontaneous referential gestures impact even more word recall than non-spontaneous lexical information gestures.

Kelly et al. (2009) further assessed the role of iconic gesture congruence by training 28 adult English naive learners of Japanese to learn Japanese verbs by observing an instructor in one of four conditions: speech, repeated speech, speech and congruent iconic

gesture, and speech and incongruent iconic gesture. They found that the group of participants exposed to speech with congruent gestures recalled the largest number of verbs while the group exposed to speech with incongruent gestures recalled the smaller number of verbs. These results indicate that the gestures that encode the semantic meaning of the target word favor memorization, while incongruent gestures may hinder recall. In the same study, Kelly et al. (2009) measured event-related potentials (ERPs) to explore the neural correlates of gesture processing that are involved in semantic memory (the N400 and the Late Positive Complex/LPC). The N400 is thought to reflect the activation of the semantic memory system during on-line language comprehension and be involved with long-term semantic memory processes. Reduced N400 can be interpreted as a reflection of the ease of effort with which people can integrate a word into some previous discourse or memory structure (e.g., Kutas & Ferdermeier, 2000). The LPC (also known as P600) reflects recall of information in long-term memory and an enhanced LPC is claimed to occur when words which are deeply encoded in long-term memory are recalled (Rugg & Curran, 2007) and when *imagistic* words are retrieved from long-term memory (Klaver et al., 2005). Results of Kelly et al. (2009) showed that words encoded with gestures produced larger LPC responses compared to words encoded without gesture, suggesting that gestures enhance imagistic long-term memory traces of words in the brain. However, null differences were found

in the N400 component. The authors suggest that "gesture does not facilitate memory for newly learned words by making them superficially familiar in an automatic fashion (the N400), but rather they may help only in later stages when people specifically identify and recall - perhaps in some sort of imagistic fashion - particular source items from memory (the LPC)" (Kelly et al., 2009, p.329).

In contrast with the abovementioned findings, a couple of studies with a within-subject design did not find any particular beneficial effect of observing or performing iconic gestures. Rowe et al. (2013) taught 62 four-year-old children novel words depicting familiar objects in an artificial language. While some objects were presented orally with their translation in English ("In Max's language, a *mip* is a book"), other words were presented with a matching picture, and others with a matching iconic gesture. Children were tested for free recall and word-meaning recall immediately after training and one week later. Results did not show any difference between the types of word presentation. With a similar design, Krönke et al. (2013) asked 11 adult German native speakers to learn novel words depicting manipulable objects in an artificial language presented in one of five conditions: no gesture, congruent iconic gesture observation, incongruent grooming gesture observation, congruent iconic gesture production, and incongruent grooming gesture production. No significant difference was found between the conditions in free and cued recall tests. However, through neuroimaging, Krönke et al. (2013) found that

actively performing congruent iconic gestures yielded larger activation in cortex areas involved in semantic processing, suggesting deeper semantic encoding of novel words. Interestingly, Kelly and Lee (2012) found that observing gestures favors vocabulary learning in a foreign language only when the phonetic demands are not very high.

Gestures may also be associated with corrective feedback. Nakatsukasa (2016) examined the referential gestures of the teacher associated with immediate feedback on a concrete linguistic structure (prepositions for places) and found that, in the long term, the immediate gesture feedback condition made it possible to improve the learners' oral production (including the target prepositions) more than non-gesture feedback condition.

The use of another type of representational gesture (i.e., deictic gesture) has also been shown to be beneficial for foreign word learning. Gullberg et al. (2010) and Gullberg et al. (2012) examined the impact of deictic gestures on word recognition in an unfamiliar language by presenting to 41 Dutch participants continuous speech in Mandarin Chinese in a weather report. The frequency of appearance of the target words was manipulated to appear either frequently (8 times), or infrequently (2 times) during the seven-minute weather report. Gestural highlights were performed naturally by the weather report presenter in the form of deictic gestures linked spatially and temporally to the referential content

(six weather icons presented on the weather chart). Results showed that participants were quickly able to recognize disyllabic words appearing eight times in continuous speech and found a significant effect of item frequency and word length, but not of gestural highlighting. However, in a sound-to-picture matching task, an interaction between frequency and gestural highlighting was found: disyllabic items which were both frequent (occurring eight times) and accompanied by a deictic gesture obtained significantly more correct scores than other items.

The effects of beat gestures (simple rhythmic gestures used to convey emphasis) for first or second language word recall have also been analyzed, with some mixed findings. So et al. (2012) examined the impact of viewing representational and beat gestures on children's and adults' memory for new verbs. While children's memory was enhanced only by viewing representational gestures, adults' memory benefited equally from viewing both representational and beat gestures. Kushch et al. (2018) taught Russian words to naïve Catalan learners either with prosodic prominence (L+H* pitch pattern), visual prominence (beat gesture), both prosodic and visual prominence, or neither of them. Results revealed that participants memorized significantly more the words presented with a combination of gesture and prosodic prominence than in the other three conditions.

Interestingly, some studies showed that the effects of gesture on word memorization can be modulated by other factors. First, the type of gesture: Levantinou and Navarretta (2015) found that the observation of beat gestures, unlike that of iconic gestures, impaired the memorization of novel words. Dargue and Sweller (2020a) confirmed that speech comprehension may benefit more from iconic gestures than by other types of gestures. Second, Rohrer et al. (2020) showed that natural, repetitive use of beat gestures may well have a negative effect on foreign language learners' recall memory and comprehension. Finally, lower language proficiency may reduce the strength of the effect of referential gestures (Ibañez et al., 2010; see also Drijvers & Özyürek, 2018; Drijvers et al., 2019, for a comparison between native speakers and language learners regarding the role of iconic gestures in clear and degraded speech comprehension).

All in all, the evidence provided in this section has demonstrated the positive role of embodiment and gesture on first and second language learning processes, especially in terms of response time, memory, language comprehension and language production. Empirical evidence has been provided for the beneficial role of using gestures in language classroom interactions and to promote vocabulary acquisition. Importantly, vocabulary acquisition has been the main focus of investigation and little work has been devoted to test the predictions of the Embodied Cognition paradigm on foreign language acquisition in terms of phonological learning

(i.e. the perception and production of the segmental and suprasegmental features of a language). In section 1.3, I review the studies that have already started to explore the role of using hand gestures mimicking phonological features for phonological acquisition in a foreign language. One of the goals of the present dissertation will be to test whether a variety of embodied teaching techniques can boost the acquisition of pronunciation in a foreign language. In the following section, I review the theoretical background for phonological acquisition in a foreign language, as well as the literature on pronunciation training techniques, in particular prosodic training.

1.2 Phonological learning in a foreign language

When learning a foreign language, learners must not only acquire a lexicon and understand how to combine these words to form meaningful sentences, but also need to learn the phonology of the language. The phonological component of the target language consists of, for one part, all the phonemes that are combined to form the words, also called segmental features (e.g., vowels and consonants), and for another part, the modulations of the voice in terms of stress, rhythm and intonation while producing both words and sentences at the discourse level, also called suprasegmental (or prosodic) features (e.g., Rogers, 2000). Language learners must be able to perceive and correctly discriminate the sounds and the intonation of the target language to achieve successful comprehension and at the same time must also be able to pronounce words and sentences sufficiently well to attain comprehensible speech. Learners generally experience difficulties in both the perception and the pronunciation of sounds of the target language. The most influential models aiming at explaining these difficulties are perception-based and focus on the acquisition of segmental information. In general, they are based on the assumption that learners rely on the phonological restrictions and categories of their native languages (L1) when learning to perceive and produce foreign sounds.

1.2.1 Models of phonological acquisition

The first well-known perception-based model is Flege's Speech Learning Model (SLM; e.g., Flege, 1988, 1995, 2002; Flege & Bohn, 2021; Flege and Liu, 2001). All versions of this proposal claim that "the mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning" (Flege, 1995, p. 239). According to SLM, language-specific aspects of speech sounds are specified in long-term memory representations also called phonetic categories. SLM predicts that brand new sounds are easier to learn than similar sounds, which learners have to distinguish from their L1 phonetic inventory in order to create a new mental representation. To form new sound categories, the model considers two important factors which facilitate sound acquisition (Flege and Liu, 2001): (a) the quality of experience (i.e. the exemplars must be salient enough to make the learners aware of the phonetic differences between the L1 and the L2) and (b) the quantity of experience (i.e. frequent encounters with the exemplars will enhance the probability of L2 sound perception).

In a first version of the theory, accurate perception of foreign sounds was assumed to be necessary in order to be able to produce these sounds. Learners should first learn to audibly perceive the differences between foreign speech sounds and their native language in order to improve their pronunciation of isolated sounds

and later, the production of these sounds would "eventually correspond" to the properties specified in the corresponding phonetic category (Flege, 1995, p. 239). However, the most recent update on the model hypothesizes that L2 phonemic perception and production "coevolve without precedence" (Flege & Bohn, 2021, p. 29). To support this claim, Flege & Bohn (2021) highlight that the correlations between perception and production observed in previous studies do not demonstrate causality but may rather show a bi-directional connections (e.g. Flege, 1999; Baker & Trofimovich, 2006; Kim & Clayards, 2019). Evidence for this bi-directionality comes from a variety of behavioral experiments (e.g. Chao et al., 2019; Perkell, Guenther et al., 2004; Perkell, Matthies et al., 2004, see, Flege & Bohn, p. 29-31 for a review). Moreover, neurophysiological investigations have shown that the regulation of motor and sensory processes used in speech production and perception is localized in "partly overlapping, heavily interconnected brain areas" (Reiterer et al., 2013, p. 9) and that brain areas specialized for speech production are active during speech perception, and vice versa (Guenther et al., 1998).

A second influential model, Best's Perceptual Assimilation Model (PAM; e.g., Best, 1994, 1995) and its extension to L2 learning (PAM-L2; e.g., Best and Tyler, 2007) is based on empirical evidence coming from the phonological development of the first language, e.g. that the human perceptual system gradually becomes attuned to L1-specific sounds and thus becomes progressively

worse at discerning sounds that are not part of the L1. The mechanism behind this process is based on phonetic-articulatory mapping: “The listener directly perceives the articulatory gestures of the speaker and, through perceptual learning, comes to detect higher-order articulatory invariants in speech stimuli” (Best & Tyler, 2007, p. 25) without the need for mental representation of phonetic categories. Language learners filter L2 speech sounds based on the manner and place of articulation of speakers' vocal tract gestures and categorize them along a gradient. Depending on the similarity to the L1, learners will be able to discern the level of contrastive phonetic detail in nonnative speech input to a varying degree. Whereas SLM focuses on individual phonetic categories, PAM focuses on phonological contrasts. According to PAM, learners adjust the L1 category to the new similar sound or assimilate the new sound to two L1 categories. However, they encounter more difficulty in creating new categories for completely novel sounds. All in all PAM-L2 is a perception-based theory and does not make any particular prediction about production patterns.

Kuhl’s Native Language Magnet Theory (NLM-e, e.g., Kuhl, 1991, 1993; Kuhl & Iverson, 1995, Kuhl et al., 2008) holds that phonetic prototypes, i.e. the central and most representative instances of phonological categories, act as perception magnets. They attract the sounds belonging to the same category and hinder native speakers from perceiving acoustic differences between prototypes and phonetically similar sounds. This model stems from the study of

developmental data and postulates that there is an evolution from infants's flexible phonological knowledge which is potentially able to discriminate any sounds from any language, to more rigid adult knowledge where phonological categories are stable enough to avoid being affected by short exposition to a foreign language.

The Second Language Linguistic Perception Model (L2LP; e.g., Escudero, 2005; van Leussen & Escudero, 2015; Elvin & Escudero, 2019; Yazawa et al., 2020) is a computational model aimed at providing a comprehensive platform to explain L2 acquisition, perception, and lexicalization. More specifically, the model proposes to use the tenets of Optimality Theory to predict phonological development, with similar premises as Best's PAM: learners initially perceive the foreign sounds that match their optimal L1 perception. With respect to the development of a foreign language phonology, it posits that learners will either need to create new perceptual mappings and categories, or else adjust any existing mappings through the same learning mechanisms that operate in L1 acquisition. Finally, the model's hypotheses of separate perception grammars and language activation predict that learners will achieve optimal perception in the foreign language while preserving their optimal L1 perception.

Although these models generally focus on segmental comparisons between L1 and L2 sound systems (that is, they work at the segmental level), they can also be applied to the acquisition of

suprasegmental features. There is evidence that language learners tend to adopt L1 prosodic patterns, for example for intonation (e.g. Gabriel & Kireva, 2014; Gut & Pillai, 2014; He et al., 2012; Ortega-Llebaria & Colantini, 2014; Trofimovich & Baker, 2006; Ulbrich, 2013; Verdugo, 2002), word stress (van Maastricht et al., 2016a; van Maastricht, Krahmer et al., 2019), rhythm and fluency correlates (e.g. Gabriel & Kireva, 2014; Trofimovich & Baker, 2006) or stress placement (e.g. Nava & Zubizarreta, 2010). A theory for L2 intonation learning (LILt) has been proposed by Mennen (2015), motivated by the fact that transfers from the L1 are frequently observed in non-native intonation production even at high levels of proficiency (see Mennen, 2004, 2007 for an overview). According to LILt, L1 and L2 intonation can be compared along the following dimensions: the inventory of structural phonological elements (such as pitch accents, accentual phrases, prosodic words, boundary phenomena), the phonetic implementation of these elements (for example, how pitch accents are lined up with the segments of utterances, what their relative height is, or what their shape or slope is), their semantic or pragmatic function and their frequency of use. The model was based on the results of cross-linguistic comparisons of intonation which were analyzed using the autosegmental-metrical (AM) framework for the description of intonational phonology (Pierrehumbert, 1980; see also Jun, 2005; Ladd, 2008 for overviews). LILt claims that, in the same way as segmental

learning, there is a perceptual basis that explains the difficulties faced by learners when attempting to produce L2 intonation. Corroborating this idea, a handful of studies have suggested that intonational cues that are not present in or differ from the L1 are more difficult to perceive by learners (e.g., Gili Fivela, 2012; Liang and Van Heuven, 2007; Nibert, 2006; Trimble, 2013) and that age and age of arrival may have an influence on the learning outcomes (e.g., Huang and Jun, 2011; Mennen, 2004).

Regarding the perception of rhythm, early studies considered rhythm as isochrony of speech intervals (Abercrombie, 1967, pp. 97–99). Languages were classified into stress-timed (e.g., German, English, in which intervals between stressed syllables were thought to be of equal durations), syllable-timed (e.g., Romance languages, in which syllables were thought to be of equal durations), and mora-timed (e.g., Japanese, which exhibit even morae). However, attempts to find isochrony in any of the timing dimensions of speech rhythm or to support the claim that languages are divided into rhythmic classes based on periodicity have been unsuccessful (e.g., Roach, 1982; Pammies Bertran, 1999; Dauer, 1983). Nonetheless, empirical research showed that adults and babies can discriminate unfamiliar languages with contrastive rhythms, and cannot distinguish between the timing patterns of rhythmically similar languages (Ramus et al., 1999; Ramus and Mehler, 1999). The *Attentional Bounce Hypothesis* states that attention is oriented to syllables which are expected to be stressed (Pitt & Samuel,

1990). It claims that the position of these stressed syllables can be predicted on the basis of the metrical patterns in one's language, and that this is reflected by quicker phoneme detection at attended syllables (Pitt & Samuel, 1990; Rothermich & Kotz, 2011). A number of measures called *rhythm metrics* have been proposed to capture systematicity in patterns of durational variability in the speech stream (see Loukina et al., 2011 for a complete list and overview of rhythm measures). Studies using rhythm metrics to compare languages which are supposedly contrastive in speech rhythm have been able to capture tangible differences between these languages (e.g., Grabe & Low, 2002; Prieto et al., 2012; White & Mattys, 2007; Nolan & Asu, 2009, among others) as well as inform on L2 learners' acquisition of rhythm (e.g., A. Li & Post, 2014; Ordin & Polyanskaya, 2014, 2015; Stockmal et al., 2005). Roughly, the latter studies showed that deviances from the target rhythmic patterns reduced progressively with increasing proficiency, regardless of the native language of the learner.

Important to the development of SLA in general and phonological acquisition in particular is the notion of L2 *phonological awareness*. In the first language acquisition literature, phonological awareness has been defined as the metalinguistic ability to segment and manipulate phonological structure and has been mainly investigated in relation to reading skills (e.g., Carroll et al., 2003). In second language acquisition research, however, phonological awareness refers to the ability to create metalinguistic knowledge

on the phonology of the target language. In his *Noticing Hypothesis*, Schmidt (1990, 2001) elaborated three stages of L2 phonological awareness. The first level is unconscious *perception* (also called *detection* in Tomlin & Villa, 1994), which is considered not sufficient to activate learning. The second level is *noticing*, that is, the focal attention during which the learner becomes aware of some form (feature or aspect) and subsequently stores it in long-term memory, and which is necessary as an initial stage of learning (Schmidt, 1990, 1994; see also Robinson, 1995). Finally, the highest level of awareness is *understanding*, i.e. analyzing, organizing and restructuring the noticed material in long-term memory, involving the recognition of a general principle, rule or a pattern in the learnt material (Schmidt, 1992). Such an awareness continuum has found support in empirical studies (e.g., Bell, 2009; Martínez-Fernández, 2008; Rosa & Leow, 2004; Rosa & O'Neill, 1999). Importantly, there is evidence from studies exploring the learning of artificial grammars that phonological awareness is necessary for the successful learning of grammar rules, i.e. mere unconscious exposure is not sufficient to engage implicit learning (e.g., Hama & Leow, 2010; Rebuschat et al., 2013; Rebuschat & Williams, 2012). It is important to note the difference between implicit *learning* as described here, and implicit *instruction* (see section 1.2.4), where attention is drawn toward a specific form, but without giving any metalinguistic explanation. In a nutshell, according to Schmidt (1990, 2001), noticing is a necessary and

sufficient condition for learning, and more noticing leads to more learning, which can be achieved through instruction. Raising learners' L2 phonological awareness comes therefore as a strong motivation for explicit pronunciation instruction practices. In general, studies which have tested pronunciation skills and language awareness, usually with the help of questionnaires and student's written reports have reported a positive relationship between language awareness and the pronunciation of specific L2 target features (e.g., Alves & Magro, 2011; Couper, 2011; Ramírez Verdugo, 2006; Saito, 2013a, 2013b, 2015) as well as overall L2 pronunciation (Kennedy & Trofimovich, 2010; Kennedy et al., 2014; Saito, 2012; Wrembel, 2005; see also sections 1.2.4 and 1.2.5). As a consequence, several instructional approaches aiming at increasing learners' awareness of the target language have been proposed, such as *processing instruction* (VanPatten, 1996, 2002), *consciousness-raising* (Sharwood Smith, 1981), and *focus on form* (e.g., Long, 1991).

In foreign language classrooms, the amount and quality of exposure to the target language is relatively limited (e.g., Larson-Hall, 2008) and therefore, learners' gains on pronunciation largely depends on the type of instruction (e.g., Norris & Ortega, 2000; Saito & Hanzawa, 2018), the amount of classroom instruction (e.g., Saito & Hanzawa, 2016), and the amount of extra-curricular L2 learning (Muñoz, 2014). While these aspects are highly related to social and contextual factors (e.g., Toth and Moranski, 2018), Dörnyei (2009)

emphasized the importance of learners' individual cognitive abilities, as well as motivation and emotion. Regarding individual cognitive differences, previous research has stressed the importance of foreign language learning aptitude, i.e. a set of abilities that enhance foreign language learning, such as sound discrimination ability, phonemic coding ability, and memorization ability (e.g., Baker-Smemoe & Haslam, 2013; Safronova, 2016; Saito, 2017; Saito & Hanzawa, 2016; Saito, Suzukida, et al., 2019). Furthermore, recent research in domain-general auditory processing suggests that learners' ability to process basic auditory information (e.g., frequency, intensity, and duration) is linked to successful L2 learning (e.g., Kachlicka et al., 2019; Saito et al., 2020; Zheng et al., 2020). The particular role of working memory has also been stressed as essential in L2 processing in general (e.g., Rankin, 2017; Reichle et al., 2016) and there is some evidence that working memory is positively correlated with phonological learning (e.g., Aliaga-Garcia et al., 2011; Darcy et al., 2015, Kondo, 2012). Musical aptitude, defined as a set of perceptive skills regarding various aspects of music such as pitch, tone, and rhythm, has been found to be a crucial predictive factor affecting L2 pronunciation (e.g., M. Li & DeKeyser, 2017; Moyer, 2014; Piske et al., 2001; see Chobert & Besson, 2013; Milovanov & Tervaniemi, 2011 for reviews). Studies have shown that better musical pitch and rhythm sensitivity predict better perception of non-native sounds (Kempe et al., 2015) as well as better L2 pronunciation (e.g. Kempe et al.,

2015; Milovanov et al., 2010; Richter, 2018; Slevc and Miyake, 2006). With respect to suprasegmental features, M. Li and DeKeyser (2017) found that good pitch perception abilities positively influenced the accuracy of perception and production of Mandarin words with contrasting tone patterns in English-speaking, naïve learners of English. In addition, Saito, Sun et al. (2019) showed that rhythmic perception abilities significantly predicted speech rate performance by Chinese learners of English.

1.2.2 The challenge of teaching the phonological system of a foreign language

On the other side of the coin, for foreign language teachers, teaching how to recognize and how to pronounce the sounds of the new language has never ceased to be a thorny pedagogical issue. One the one hand, teachers are not confident in what to teach, when to teach it, and how to teach it, and they often put the blame on a lack of training (e.g. Darcy et al., 2012; Foote et al., 2012; MacDonald, 2002). On the other hand, they must handle learners who are not equal, because of unfavorable personal or environmental situations, intrinsic motivation, anxiety, learning style, and individual cognitive differences (see Suzukida, 2021, for a review). Despite this, in the last decades a substantial amount of research has shown that pronunciation instruction is beneficial for learners.

From the language classroom's perspective, before the 1960s, teaching pronunciation generally meant drilling activities with the objective of reaching native-like pronunciation, mistakes were corrected immediately and native speaker pronunciation was the model to attain. Pronunciation instruction as a whole was considered ineffective to help learners achieve communicative competence (e.g., Purcell & Suter, 1980) and the role of comprehensible input was favored over explicit instruction in the classroom (Krashen, 1981). Later, with the development of the

communicative approach to language teaching and learning, the emphasis on individual sounds and repetitive, out-of-context activities did not suit well to the pedagogical shift to communicative activities. Moreover, pronunciation activities were deemed too difficult to implement during class (e.g., MacDonald, 2002).

In the last decades, both researchers and practitioners have realized that a good pronunciation could enhance communicative skills. The strongly established communicative language teaching and learning model seems to move towards a pedagogical framework integrating focus on form - grammar, lexis, and pronunciation - with more general communicative skills (e.g., Burgess, 1994), with the goal to achieve successful communication instead of native-like proficiency and by focusing on intelligibility (e.g., Derwing et al., 1998; Levis, 2005; Morley, 1991; Prator, 1971). In fact, it has been demonstrated that effective communication is impossible when learners' pronunciation falls below a certain threshold level, even when their vocabulary and grammar are excellent (Derwing & Munro, 2015; Levis, 2018).

Recent review articles and meta-analyses have confirmed the crucial role of pronunciation instruction to improve language learners' pronunciation (J. Lee et al., 2015; Saito & Plonsky, 2019; Sakai & Moorman, 2018; Thomson & Derwing, 2015). J. Lee et al. (2015), and Sakai and Moorman (2018) underscored the fact that

the diverse tested instruction paradigms in previous studies mostly yielded medium effect sizes, with visible improvements mainly on controlled tasks such as repetition or reading tasks and higher effect sizes in laboratory settings. In line with this, nowadays foreign language teachers tend to regard pronunciation as an important aspect of language to be mastered by learners in order to achieve successful communication (e.g. Nagle et al., 2018). Burgess & Spencer (2000) listed a series of difficult aspects of pronunciation teaching: the selection of the features, the ordering of the selected features, the type(s) of discourse in which to practice pronunciation, the choice of the methods, and the amount of detail to go into at different stages. Teachers are reported to find it easier to teach segmental features (Saito, 2014), although deciding whether to focus on segmentals or on suprasegmentals, and to what extent, is also a common issue (e.g., Derwing et al., 1998; Jenner, 1989; Zielinski, 2008). Also, there is little evidence indicating at what proficiency level a pronunciation activity is appropriate (but see Gilbert, 2001a, b; Jenner, 1989; Murphy, 1991). Instructors have reported relying on their own intuitions when explaining pronunciation (e.g., Levis, 2005). The lack of adequate language teacher training in pronunciation may result in teachers' lack of knowledge and confidence and do not favor a central role of pronunciation instruction, which is often relegated to the sidelines or even ignored (e.g., Derwing, 2010). Furthermore, language textbooks, which are often the focal point for teaching practices and

in-class activities, may not always present an effective solution to help students with pronunciation (Derwing et al., 2012). In addition, improvements after pronunciation instruction are not easily and rapidly visible. For example, after successfully practicing a given pronunciation feature in a controlled exercise, the improvement may not transfer immediately to spontaneous speech, during which attention is generally focused on meaning (e.g., Bowen, 1972).

In general, the challenge posed to the teacher is to find a way to help learners enhance their language skills by taking into account the constraints placed upon their L2 phonological system (see section 2.1). A recent study by Tstunemoto et al. (2020) demonstrated that more experienced language teachers, in terms of both general and pronunciation instruction, tend to be more skeptical about how easy and how efficient it is to teach L2 pronunciation, while less experienced teachers are more positive about it. The authors suggest that teacher training should shift toward communicative-oriented dimensions of L2 speech and provide teachers with pedagogical skills to target these dimensions. Moreover, in recent literature a series of studies have also tried to determine the best way to assess learners' speech improvements. In the following section, we review the three main measures of non-native pronunciation evaluation, namely fluency, comprehensibility, and accentness.

1.2.3 Perceptual assessment of oral proficiency: perceived fluency, comprehensibility, and accentedness judgments

Accentedness, comprehensibility and intelligibility, and fluency are the most commonly used measures to perceptually assess oral proficiency of foreign language learners which are generally evaluated by native speakers of the target language (e.g., Munro & Derwing, 1995; Saito et al., 2017). In a review article, Saito and Plonsky (2019) pointed out that in order to effectively assess the effect of pronunciation instruction, pronunciation proficiency measures should be standardized and comprehensively analyzed by taking into account both global (fluency, comprehensibility, and accentedness) and specific constructs (segmental and suprasegmental features), and by contrasting human impressionistic judgments with acoustic analyses. The present section focuses on the three overall constructs of pronunciation, namely accentedness, comprehensibility, and fluency, as these have been the most frequently-used measures in pronunciation training studies. However, it is important to mention that other perspectives on pronunciation assessment have been proposed (Isaacs & Trofimovich, 2016), involving the relationship between pronunciation and other language skills such as listening and writing, and the role of factors influencing intelligibility and comprehensibility judgments, such as cognitive abilities (e.g., cognitive control) and sociocultural factors (e.g., native speaker

status, language variation).

Accentedness is generally defined as the perceived distance between L2 speaker's speech and that of a native speaker's (Trofimovich & Isaacs, 2012). Accentedness has been primarily related to pronunciation accuracy measures in terms of segmental and suprasegmental errors (Saito et al., 2016). Regarding the assessment of accentedness, previous studies have shown that both experienced raters who are specialists in linguistics or teaching and novice raters with no linguistics background produced similar judgments when evaluating accentedness (e.g. Isaacs & Thomson, 2013). While some studies suggest that suprasegmental features weigh heavily in the perception of foreign accentedness (e.g., Anderson-Hsieh et al., 1992; de Mareüil & Vieru-Dimulescu, 2006; Kang, 2010; Kang et al., 2010; Polyanskaya et al., 2017; Trofimovich & Baker, 2006; van Maastricht et al., 2016b, van Maastricht et al., 2020), by contrast other studies consider segmental accuracy to be an important cue for native judgements of accentedness (e.g., Saito et al., 2016, 2017; Trofimovich & Isaacs, 2012). Regarding the debate on what to teach in priority, recent meta-analytic studies and reviews suggest that both suprasegmental and segmental features should be trained in pronunciation instruction (Lee et al., 2015) and that teachers should take advantage of the strong interactions between the two (X. Wang, 2020, Zielinski, 2015).

Comprehensibility can be defined as the ease of understanding of the meaning of what is uttered (Derwing & Munro, 2009). It is partly affected by grammar, lexis and discourse complexity (e.g., Isaacs & Trofimovich, 2012; Saito et al., 2016; Trofimovich & Isaacs, 2012) and partly by pronunciation components, including a range of suprasegmental features (Anderson-Hsieh & Koehler, 1988; Crowther et al., 2015; Field, 2005; Isaacs & Trofimovich, 2012; Saito et al., 2016; Trofimovich & Isaacs, 2012; van Maastricht et al., 2016b; van Maastricht et al., 2020; Warren et al., 2009) to segments with high functional load (Munro & Derwing, 2006; Suzukida & Saito, 2021). Comprehensibility is tightly linked to the concept of intelligibility, which is the recognition of the components of an utterance, regardless of what is meant. While a comprehensibility measure is generally obtained thanks to listeners' judgments, intelligibility is often elicited with the written transcription of L2 speakers' oral productions (e.g., Munro & Derwing, 1995).

Fluency generally refers to a set of measurable temporal aspects of speech (e.g., De Jong et al., 2013; Segalowitz, 2010) and encompasses the observable notions of breakdown, repairs, pausing and speech rate (e.g., Tavakoli & Skehan, 2005). Ideally, one of the goals of language learners is to produce"speech at the tempo of native speakers, unimpeded by silent pauses and hesitations, filled pauses, self-corrections, repetitions, false starts and the like" (Lennon, 1990, p. 390). One frequently adopted measure of speech

rate is articulation rate, i.e. the pace at which speech segments are produced and in which all pauses are excluded from the calculation. These nuances in speech can also be captured through *perceived fluency*, i.e. the impression that native listeners have of the fluency of a certain speech sample. According to Lennon (1990, p. 391), fluency "is an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently." Studies on the relationship between objective fluency measures (such as speech rate and pausing) and subjective judgements on L2 speech samples have demonstrated a strong correlation between the two (e.g., Cucchiarini et al., 2002; Derwing et al., 2004; Lennon, 1990; Riggenbach, 1991; Rossiter, 2009). Interestingly, other studies have linked fluency ratings to measures that were not related to temporal or breakdown and repair aspects of speech, such as overall pronunciation and grammar proficiency, as well as vocabulary size, and age of arrival (e.g., Freed, 1995; Kormos and Dénes, 2004; Rossiter, 2009, Trofimovich & Baker, 2006).

In a recent study, Suzuki and Kormos (2019) found that fluency and comprehensibility were strongly correlated (see also Crowther et al., 2016; Isaacs & Trofimovich, 2012; Saito et al., 2017), and that both constructs were associated with grammatical accuracy and pronunciation. In addition, they found that comprehensibility was best predicted by articulation rate (speed fluency), whereas perceived fluency was most strongly associated with the frequency

of mid-clause pauses (breakdown fluency). Interestingly, Kang (2010) analyzed speech of 11 learners of English for measures of speech rate, pauses, stress, and pitch range and the same samples were evaluated by native speakers. The results revealed that accentedness ratings were best predicted by pitch range and word stress measures whereas comprehensibility scores were mostly associated with speaking rates.

Although aiming for a perfect native-like accent may not represent a feasible goal, accentedness measures remain a very salient feature of L2 speech, strongly affected by the pronunciation of both phonemes and prosody. Empirical evidence suggests that improving speech comprehensibility is perhaps a more realistic learning goal even for late learners, as even accented and disfluent speech can be understandable (e.g., Saito et al., 2016). Despite this, in research involving the perceptual evaluations of short strings of speech (on a single phoneme, syllable, or a word), it may be more adequate to adopt the criterium of accentedness, asking raters to evaluate in terms of nativelikeness by comparing learners' oral production to a model. Altogether, accentedness measures deserve to remain an important measurement together with comprehensibility and fluency in order to obtain a comprehensive evaluation of L2 speech.

In addition to global measures of pronunciation, acoustic analyses represent an objective, complementary measurement method for

pronunciation assessment (Saito & Plonsky, 2019). Saito and Plonsky (2019) proposed to include the acoustic analyses of fundamental frequencies, formants, intensity, articulation rate and pause ratio of learners' speech samples in the assessment of pronunciation teaching effectiveness on both segmental and suprasegmental features. The measures described above, both global and acoustic, have been frequently but heterogeneously used by researchers who have seeked to evaluate pronunciation after some instruction in studies with a training design. These studies are reviewed in the following sections.

1.2.4 Types of pronunciation training

Celce-Murcia et al. (2010) stated two approaches of pronunciation teaching, both based on perceptive models: first, the "intuitive-imitative approach" - or *implicit* method - which suggests that learners can improve pronunciation by listening and imitating a model, usually the teacher (or also audio recordings), and second, the "analytic-linguistic approach" - or *explicit* method - which encourages the use of some tools and techniques such as phonetic alphabet and transcriptions to learn about the phonology of the target language. There is ample empirical evidence that explicit phonetic instruction facilitates various dimensions of pronunciation development in a second language (for reviews, see J. Lee et al., 2015; Saito & Plonsky, 2019; Thomson & Derwing, 2015). Other studies have shown that methods that rely on implicit techniques can also trigger significant phonetic learning at the segmental and word levels (e.g., Wanrooij et al., 2013; Escudero & Williams, 2014; Ong et al., 2017; Tuninetti et al., 2020). Interestingly, Ong et al. (2015) investigated the effect of the distributional learning of difficult Thai tones with English-speaking naïve participants and found that learning only took place when participants were encouraged to pay attention during learning, that is, participants were instructed to indicate whenever they heard a 'beep', which forced participants to pay attention to each sound heard during the training phase.

As mentioned in section 1.2.1, Schmidt (1990, 2001) claimed that noticing, a term coined by the author, is necessary for the correct development of foreign language acquisition, and grants access to awareness and subsequent learning. Several instructional techniques promoting learners' attention to form-related features in L2 input have been proposed to teach pronunciation including explicit explanation (Derwing et al., 1998), recasts (Lyster, 1998), metalinguistic feedback (Hardison, 2004), and input practice (Bradlow et al., 1997). In terms of classroom practice, perception training should include a key role for explicit instruction where "learners' attention must be explicitly drawn to the differences in the L2 and the L1 via form-focused instruction, and errors in the learners' L2 production would benefit from explicit corrective feedback" (B. Lee et al., 2020, p. 3). Saito & Plonsky's (2019) synthesis of the literature on pronunciation instruction shows that, in this particular domain, form-focused instruction has been designed to help learners grasp the perceptual similarities and dissimilarities between L2 sounds and their L1 counterparts.

As a correlate of focus-on-form, an important technique used during focus on form pronunciation instruction is speech imitation. It is an established fact that repetitive practice enhances speed and efficiency in performing cognitive tasks (e.g. Schneider & Chein, 2003) and repetition is also necessary for grammar and lexical learning in a foreign language (e.g., Gass et al., 1999; Jensen & Vinther, 2003). Intensive perception training in which learners are

exposed to multiple repeated instances of L2 sounds leads to improvements in L2 phonetic perception and production of difficult segments (e.g., Bradlow et al., 1997; Lord, 2005; Saito, 2015). In an auditory word-priming experiment with 60 learners of Spanish, Trofimovich & Gatbonton (2006) showed the beneficial role of repetitive practice coupled with explicit focus on form related to the phonological properties of the words (i.e. participants were asked to judge how *clear* the words sounded). Results showed that repetition induced faster response times in general, and additional focus-on-form triggered even faster responses, in particular for learners with lower pronunciation skills, showing an effect of repetition and focus-on-form on language processing. The same beneficial effect of repetition during auditory priming had been observed by Jung et al. (2017) for the learning of lexical stress by 57 Korean learners of English. The results of this study showed that auditory priming improved the production of lexical stress in a reading aloud task. In an example of focus-on-form classroom application, Lord (2005) examined the improvement of English-speaking learners of Spanish after an advanced-level phonetics course. Pronunciation instruction included explicit articulatory instruction, oral practice, transcription, and student use of Praat speech analysis software. Results showed improvements in the production of voiceless stops, diphthongs, and fricatives over the course of the semester. Recently, more evidence is supporting that a technique called high variability phonetic training (HVPT)

using multiple voices rather than one voice - hence the term variability - help enhance listeners' ability to perceive non-native sounds (e.g., Aliaga-Garcia, 2017; Cebrian & Carlet, 2014; see Barriuso & Hayes-Harb, 2018, Thomson, 2018 for reviews).

Within a communicative approach, meaningful content and form-focused instruction can be integrated: teachers can draw learners' attention to problematic features and increase their frequency or salience to encourage awareness and learning (Lyster, 2007). For example, Gatbonton and Segalowitz (1988, 2005) proposed a framework called ACCESS (Automatization in Communicative Contexts of Essential Speech Segments), in which the learners engaged in genuine communicative activities and exchanged useful and needed information, but which also required the repetition of meaningful utterances containing the target segmental or suprasegmental features. An ACCESS lesson consists of three phrases, starting with the phase o*f Creative Automatization* during which learners are engaged in a meaningful communicative activity where they actually use the target forms repeatedly by interacting with each other. In the second phase, the *Language Consolidation* Phase, learners complete a series of activities to raise phonological awareness on the target forms, identify them and practice their production with feedback. Finally, the *Free Communication* phase allows learners to reuse the target structures in a different meaningful communicative activity (see Trofimovich

& Gatbonton, 2006, for an example lesson plan to teach the intonation of Yes/No and Information questions in English).

Another approach, the *Task-Based Pronunciation Teaching* (TBPT) is based on *task-based instruction*, a type of communicative methodology in which learners are engaged in real-world tasks, fostering meaning-oriented communication and interaction. In this approach, learners have to rely on their own resources to complete the communicative task and need to attend to linguistic forms at the same time (Ellis, 2009). Following Robinson's Cognition Hypothesis (2001, 2005), the more complex tasks will promote more interaction, attention to form, and uptake of information from the input, and therefore will foster more accurate L2 language. A handful of studies have found TBPT effective for the pronunciation of difficult segmental contrasts (Mora-Plaza et al., 2018; Solon et al., 2017) and suprasegmental features (Jung et al., 2017; McKinnon, 2017), also confirming the role of task complexity. In addition, in a recent classroom application of TBPT, Gordon (2021) taught a variety of suprasegmental features to three groups using low, intermediate or high task complexity. Results showed a significant improvement in comprehensibility for the learners who followed the TBPT with the highest complexity. By contrast, no improvement in fluency and accentedness was obtained after training.

Summarizing, a recent article by Colantoni et al. (2021) proposed a set of five principles for the teaching of pronunciation based on experimental evidence. The first principle proposes that, on the assumption that perception leads production (e.g., Flege, 1995; Escudero, 2009; Goodin-Mayeda, 2019), initial stages of pronunciation instruction should involve perception-based activities. Second, the authors recommend that initial instruction should incorporate prosodic features such as rhythm and intonation and not focus on segments alone (as in de la Mota, 2019). Third, even with lower proficiency learners, practice should be incorporated in a communicative context (e.g., Mora & Levkina, 2017). The fourth principle is that focus should be made on features with a higher functional load (e.g., Brown, 1988; Munro and Derwing, 2006; Dupoux et al., 2008). Finally, features that do not impede intelligibility should be left for later instruction.

A recent review by X. Wang (2020) suggests that researchers should take a holistic perspective on the acquisition of L2 segmental and suprasegmental features, as both suprasegmental and segmental features are tightly related, and that training on larger speech chunks may also facilitate segmental accuracy (see also Zielinski, 2015, for similar considerations). Interestingly, confirming Colantoni et al.'s second principle, McAndrew's (2019) meta-analytic review showed that pronunciation instruction focusing on the learning of suprasegmentals leads to large learning effects, even if instruction sessions last only a few hours. Thomson

and Derwing's (2015) review article stated that while 53 percent of the studies included in their analysis investigated segmental training, 23 percent focused on suprasegmentals and 24 percent dealt with both, usually in combined lessons but occasionally as separate comparison groups. However, little is known about the effectiveness of the specific techniques used in the prosodic-based instruction paradigms included in the reviewed studies. In this regard, J. Lee et al. (2015) noted that a more thorough description and empirical assessment of the training activities would be needed in written reports. In the following section, we will look in more detail at the results of prosody-based pronunciation training studies and review the techniques which focus on the prosody of the L2.

1.2.5 Prosodic pronunciation training

As mentioned in the previous section, focusing and raising awareness of pronunciation in general seems beneficial for the development of L2 learners' pronunciation. Regarding prosodic training specifically, only a few studies have looked at the effects of an explicit focus-on-form approach to prosody teaching and learning on global and specific measures of pronunciation. Saito and Saito (2017) examined the effects of suprasegmental training on prosodic and comprehensibility outcomes in Japanese beginner learners of English. Ten students received three hours of suprasegmental instruction over six weeks, while ten other students followed meaning-oriented instruction without any focus on suprasegmentals. Results showed significant gains in overall comprehensibility, word stress, rhythm, and intonation in a reading-aloud task for both trained and untrained lexical contexts for the experimental group only. In particular, learners produced longer and clearer stressed vowels, reduced vowels in unstressed syllables, and used appropriate intonation patterns for yes/no and wh-questions.

A couple of classroom-based studies have directly compared prosodic pronunciation instruction to segmental pronunciation training and found some advantage of training prosody over segments. In a three-week pronunciation training study, Gordon et al. (2013) compared explicit suprasegmental instruction, explicit

segmental instruction, and no explicit instruction with learners of English and found that only the explicit group trained on suprasegmentals significantly improved comprehensibility scores from pretest to posttest in a sentence repetition task. Recently, R. Zhang and Yuan (2020) compared the effects of segmental and suprasegmental pronunciation instruction with Chinese learners of English. During an 18-week training period, 30 learners followed segmental training, the same number of learners followed suprasegmental training while a third group received instruction without reference to pronunciation. The results showed that both the segmental and suprasegmental groups improved their pronunciation significantly, in terms of comprehensibility in a sentence-reading task. However, in a spontaneous speech task, only the suprasegmental group made significant progress in comprehensibility and maintained these gains at a delayed posttest.

As a classroom application of prosodic training, Missaglia (1999, 2008) developed the *Contrastive Prosody Method*. The aim of this method was to impede and correct specific prosodic errors and fossilized features in the L2 such as intonation contours and word and sentence stresses through awareness training of the different prosodic variants of speech acts in the target language. Missaglia (1999) compared the *Contrastive Prosody Method* to segmental pronunciation training with 20 Italian beginner learners of German in a 20-hour pronunciation course over 10 weeks. Ratings of a reading-aloud task by five German native speakers showed that the

group that was trained with the Contrastive Prosody Method obtained significantly better improvement between pre- and posttest in terms of global pronunciation and both segmental and suprasegmental accuracy.

Together with repetition (see section 1.2.4), imitation is one of the most popular methods to teach pronunciation in the classroom, where the teacher generally provides a model to copy and repeat after (e.g., Celce-Murcia, 2001). Computer-assisted learning, based on the development of speech analysis technology, has shown interesting possibilities for learning L2 suprasegmental features by allowing learners to obtain immediate visual feedback on their oral production. Such computer programmes allow learners to compare the visual representation of target pitch contours produced by native speakers to their own, notice the differences between the two and try to match it with the native model by repeating the input (e.g., Olson, 2014). A series of early studies reported that learners following such a method improved their pronunciation (deBot, 1983; Weltens & deBot, 1984a, 1984b). Further beneficial effects were found in terms of accentedness (Hardison, 2004), global oral proficiency (Gorjian et al., 2013), intonation (Hincks & Edlund, 2009; Ramirez Verdugo, 2006), and the accuracy of stress patterns (Schwab & Goldman, 2018; Tanner & Landon, 2009).

Interestingly, some studies have found that the use of musical activities highlighting the rhythmic and melodic properties of the

foreign language are helpful in improving global measures of pronunciation. For example, Derwing et al. (1998) compared segmental training, suprasegmental training and no pronunciation training with 48 intermediate learners of English. Results showed that participants in the suprasegmental training group, who were asked to focus on rhythmic features by tapping out the beats, counting the syllables and finding the stressed syllables in musical chants obtained higher improvements in comprehensibility and fluency in spontaneous speech at posttest, compared to participants who practiced the identification, discrimination and pronunciation of individual sound contrasts, and participants who did not follow any pronunciation training. Fischler (2009) created a method based on rap songs called *Rap on Stress* to work on speech rhythm with young advanced learners of English and found improvement in stress placement after four weeks of training. Students also may focus on melodic features by singing or listening to songs (e.g. Good et al., 2015; Ludke, 2016, 2018). Combining computer-assisted techniques and music-based techniques, W. Wang et al. (2016) tested the effects of a computer application that automatically generated a percussive beat corresponding to the rhythm of English sentences. Twenty Chinese learners of English were asked to pronounce 15 English sentences before hearing the rhythmic cue. They could practice reading the sentences as many times as they liked before recording them using their own voice. Then, they practiced as many times as they liked repeating the

same sentence alongside the rhythmic cues and recorded their voice a second time. Participants' accentedness was evaluated by 10 English native speakers and results showed that rhythmic priming particularly benefited beginner learners with the lowest ratings at pretest.

In sum, prosodic training strategies that highlight the rhythmic and melodic features of the target language have been shown to play a positive role in improving learners' global pronunciation and suprasegmental features. However, teaching a different prosodic system still remains a challenge, and practical techniques to teach prosody need to be proposed and tested. In the following section, we get to join both theories of embodiment and phonological learning by reviewing experimental and classroom techniques dedicated to the improvement of pronunciation through the use of embodied techniques such as the use of visuospatial hand gestures and percussive hand movements that highlight prosodic features.

1.3 Embodied pronunciation learning: prosody in movement

Visually representing prosodic features of speech by means of hand gesture is not uncommon in the foreign language classroom. Smotrova (2017) conducted an observational study where she collected video recordings from two 50-minute classes focused on reading instruction by the teacher of a beginner ESL university classroom. These lessons included increasing awareness about suprasegmental features of English such as word stress and syllabification. The findings indicated that the teacher employed a mixture of preplanned and spontaneous gestures to teach suprasegmental features, which were then picked up and imitated by students in their learning process. To help learners with words' syllabification, the teacher "marked the syllables with her body by slightly nodding her head and tapping the fingers of her left hand with her right hand" (Smotrova, 2017, p. 69). Smotrova (2017) also observed that the teacher helped learners "see" word stress placement by producing an entire upward movement of her body when pronouncing the stressed syllable with higher voice intensity. Finally, while working on the pronunciation of proverbs, the teacher "makes the rhythm of the proverb visible by moving her hands upward and downward in a rotating motion. The teacher complements it with a slight movement of her whole body in rhythm with the stressed syllables, creating an impression of

dancing on the spot (Smotrova, 2017, p.77). Other observational studies have highlighted the use by teachers of other body movements marking prosodic features. For example, ascending and descending horizontal hand gestures mimic the melody of the phrase (Tellier, 2008a) while horizontal hand movements or lateral body movements can represent vowel duration (Hudson, 2011). In addition, the use of beat gestures (i.e., rapid and repetitive rhythmic movements of the arms, hands, fingers typically associated with prosodically prominent positions in natural discourse) or hand-clapping can help highlight the rhythm of speech, divide the speech into syllables, or mark prominent and stress positions in speech (Chan, 2018; Baker, 2014; Hudson, 2011).

One method for teaching foreign language pronunciation that makes an essential use of hand gestures representing prosodic features of speech is the verbotonal method. The verbonal system was first developed by Guberina (1956) as a technique to enhance speech production for patients with hearing pathologies. Later on, he further adapted and extended this technique to foreign language learners (Guberina, 1961; Renard, 1979), under the premise that these learners were suffering a similar "deafness" to L2 sounds also called phonological 'sieve' (Troubetzkoy, 1964). According to the phonological 'sieve' hypothesis, a phonological system is the result of the organization of a limited number of perceived sounds that are necessary for communication and are specific to each language. By emphasizing the concept of the phonological sieve and positing that

humans analyze sounds using the benchmark of the sounds of our mother tongue, the verbotonal method assumes that perception precedes production in learning pronunciation, in conformity with the major theories of phonological acquisition (PAM; e.g. Best, 1994; SLM; e.g. Flege, 1995; L2LP; e.g. Escudero, 2005; see section 1.2.1).

Crucially, the verbotonal method underlines the importance of prosody and encourages the teaching of suprasegmental features through body movement and gesture from the early stages of language learning, and with oral rather than written input (Billières, 2002; Intravaia, 2000). One of the fundamental ideas of the verbotonal method is that "any sound is the result of a movement" (Billières 2002, p. 43) and that micro-articulation, i.e. articulators' gestures and organs used to produce sounds, and macro-articulation, i.e. body movements, are interdependent. Depending on the type of error produced by the learner, three levels of teacher feedback are activated simultaneously (e.g. Billières, 2002; Klein, 2010; Renard, 1979): prosodic correction, and phoneme correction through nuanced pronunciation and facilitative consonantal context. The teacher's corrective feedback of consonants and vowels is based on the notions of *tension* (i.e. the energy that is necessary to make speech sound) and of *clarity* (i.e. a *clear* sound implies high-pitched frequency, e.g., [i], a *dark* sound implies low-pitched frequency, e.g., [u]). Body movements can help tense or relax the phoneme thanks to the connection between

micro- and macro-motricity. For instance, the tension of a consonant (e.g., [p], [t], [k]) can be increased by moving up the head and clenching the fists, and it can be decreased by moving your body towards the ground, lowering the head and arms. The technique of facilitative consonantal contexts consists of correcting the pronunciation of vowels perceived as too tensed or too relaxed by changing the preceding consonant in the syllable: clearer vowels (e.g., [y]) will be obtained by switching a dark consonant (e.g., [b]), with a clearer consonant (e.g., [t], [s]).

From his classroom practice with Japanese learners of French, Klein (2010, p. 51) reported the usefulness of a vast range of gestures and movements: using a slow horizontal hand gesture to indicate that a sound should be relaxed, clenching fists to tense a sound, using referential gestures to associate each phonological element of a minimal pair of words to their corresponding meanings, touching the nose when producing nasal sounds, lowering the head to produce darker sounds and rising the head to produce clearer sounds, marking the syllables with beat gestures, using a rising hand movement when uttering a question and a falling hand movement at the end of assertions. Curiously, despite the existence of actual classroom practice and the availability of teacher training programs using verbotonal method, to date, there is no published description or inventory of the gestural techniques employed in this method, especially regarding prosodic correction. In addition, very few experimental studies have tested the

effectiveness of the method (e.g,, Alazard-Guiu, 2014; see section 1.3.1.3) and to our knowledge no empirical investigations have assessed the potential beneficial role of gesture within this approach.

Even though the abovementioned pronunciation teaching techniques have been said to return good results, they are essentially based on practical trials and observation in the classroom and not on experimental research. A recent line of research has started to empirically test the effects of specific hand gestures visually encoding a set of phonological features of the target language. According to the specific features depicted, visuospatial hand gestures can be further classified into: (a) pitch gestures (a term coined by Morett & Chang, 2015), which are gestures mimicking F0 movements; (b) durational hand gestures, which are gestures showing phonemic contrast in duration; (c) hand articulatory features which cue one articulatory property of a phoneme, such as the aspiration in aspirated consonants; and (d) phrase-level prosodic gestures which depict both rhythmic and prosodic features at the phrase level. In addition, percussive hand movements like hand-clapping visually (as well as aurally) encode the rhythmic structure of speech.

In the present dissertation, we will deliberately focus on the potential benefits of the use of visuospatial hand gestures and percussive hand movements like hand-clapping representing

prosodic features for pronunciation instruction. Even though there is some recent empirical evidence on the benefits of hand articulatory gestures on the acquisition of segments (e.g., Amand & Touhami, 2016; Hoetjes et al., 2019; P. Li et al., 2021; Xi et al., 2020), more work is needed to systematically assess the value of embodying prosodic features in the context of pronunciation instruction. The following subsections review the studies that have been conducted on this topic.

1.3.1 Visuospatial hand gestures representing prosodic features

a) Tonal contrasts

In the so-called tonal languages like Mandarin Chinese, pitch variation (i.e., a change in the fundamental frequency) at the syllable level leads to a distinction in meaning between words that are segmentally identical (Xu, 1994). The lexical tonal contrasts are particularly difficult to acquire for speakers of non-tonal languages (e.g., Kiriloff, 1969). Despite this intrinsic difficulty, there is evidence that speakers of both tonal and non-tonal languages can be trained with success in both the perception and production of L2 tonal systems (e.g., Francis et al., 2008; Hao, 2012; M. Li & DeKeyser, 2017, among many others). In addition, as mentioned in section 1.2.1, language learners frequently face difficulties in learning intonational patterns of an L2 because they tend to transfer the intonational patterns of their L1 to the L2, both in perception (e.g., Cruz-Ferreira, 1989; He et al., 2012; Ortega-Llebaria et al., 2015) and production (e.g.,; Ortega-Llebaria & Colantoni, 2014; see Mennen, 2015, for a review). Learning intonational patterns may be even more challenging for speakers of tonal languages (e.g. Cortés-Moreno, 1997). The studies reviewed below all look at the effect of pitch gestures on the learning of L2 intonational and tonal contrasts. The term pitch gesture was first coined by Morett & Chang (2015) and refers to a type of visuospatial hand gesture in

which upward and downward hand movements mimic melodic up and down pitch movements of tonal contrasts. These hand gestures are based on a spatial conceptual metaphor in which the position of the hands high in space represents high-frequency pitch and the position of the hands low in space represents low-frequency pitch.

A handful of studies have demonstrated the close cognitive link between spatial height and speech "tonal height". Casasanto et al. (2003) showed lines 'growing' vertically (bottom to top) and horizontally (left to right) to two groups of participants respectively while listening to sounds with different increasing pitch modulations. Participants were asked to reproduce the sounds they listened to after each trial. The results showed that vertical displacement strongly modulated participants' estimation of acoustic pitch, whereas horizontal displacement did not, confirming that the metaphoric relationship between pitch and height is not only linguistic but also conceptual. Interestingly, this spatial conceptual metaphor of pitch is present in 4-month old infants (Dolscheid et al., 2012), indicating that this visuospatial-acoustic dependency might be language-independent. In addition, there is evidence that speakers intuitively associate metaphorical gestures encoding spatial height with musical pitch (Cassidy, 1993; Connell et al., 2013; Forsythe & Kelly, 1989). Connell et al. (2013) further investigated the role of visual movement in the perception of pitch. Participants were asked to judge whether a target note produced by a singer was the same as or different from a preceding note. Some

of the notes were presented with the corresponding downward or upward pitch gestures, while others were accompanied by contradictory spatial information, for example, a high pitch with a falling hand gesture. The results showed that pitch discrimination was significantly biased by the spatial movements produced in gesture, such that downward gestures induced perceptions that were lower in pitch than they really were, and upward gestures induced perceptions of higher pitch, supporting the "shared representation" hypothesis of height in both pitch and space.

These behavioral dependencies have been further supported by neuroscience research. In several studies, musical pitch processing activated primary visual areas (e.g., Degerman et al., 2006; Dolscheid et al., 2014; Foster & Zatorre, 2010), suggesting that representations underlying musical pitch may be visuospatial in nature. In an fMRI experiment, Dolsheid et al. (2014) investigated if pitch representations overlap with unimodal (visual or tactile) or multimodal (visual + tactile) spatial representations in three different sessions, visual, tactile, and auditory. In the visual block, while in the scanner, participants were asked to compare two visual stimuli presented on a screen and indicate as accurately and fast as possible whether both stimuli were the same or different with respect to either shape (circle and square) or position (high and low). In the tactile block, the experimenter presented the stimulus by touching the palm of the hand of the participants with two wooden artefacts following the same dimensions (high vs. low,

circle vs. square). Finally, in the auditory block, participants listened to two consecutive auditory stimuli and were asked to judge if they were similar or different in terms of tone (pitch) or instrument (timber). Results of whole brain and ROI (regions of interest) analyses revealed unimodal activation of visual areas during musical pitch judgments, suggesting that judgments of musical pitch depend in part on visual areas that are involved in spatial height processing and supporting the spatial metaphor for pitch. Crucially, by comparing shapes that differ in spatial height to those that remain at a constant position (control), the authors found activations in primary visual cortex, an area shown previously to be sensitive to changes in spatial position (e.g., Bosking et al., 2002). No evidence of multimodal activation was found, but the authors suggest that overlap may happen in more complex pitch and space judgment tasks. Other studies have shown that brain regions involved in multimodal processing are also involved in pitch memory (Rinne et al., 2009) pitch production (Peck et al., 2009), pitch identification (Schwenzer & Mathiak, 2011) and pitch transformation (Zatorre et al., 2010).

Pitch gestures have been found to be helpful both for the learning of non-native tonal contrasts in a tonal language patterns and pitch contours in an intonational language. Regarding specific pitch contours, Kelly et al. (2017) explored the effect of pitch gestures (vs. no gesture or incongruent gesture) on the perception of intonation features of Japanese by 57 English-speaking participants

without any previous knowledge of Japanese. Unlike English, in Japanese the acoustic patterns throughout declarative and question sentences remain identical and only the intonation of the final syllable changes. Results showed that observing pitch gestures signaling a question with an upward hand movement and an affirmative sentence with a downward movement on the final syllable of a sentence helped learners identify significantly better the intonational contrast compared to incongruent gestures and no gesture. Regarding pitch contour pronunciation, Yuan et al. (2019) taught 64 Chinese beginner learners of Spanish a selection of intonation patterns (statements, yes/no questions, and requests) with or without pitch gestures representing nuclear intonation contours. The results showed that observing the instructor performing a pitch gesture on the target intonational contours while uttering the sentence significantly improved the participants' pronunciation of such contours at posttest compared to the pronunciation by participants after observing the instructor simply uttering the sentences.

A number of studies have demonstrated that observing or performing pitch gestures significantly improves the production of L2 lexical tones. The first study to examine the effects of training with the production of pitch gestures by learners themselves was carried out in a classroom setting by Chen (2013). In a between-subject experiment, 40 learners of Chinese from different countries learned the target Chinese lexical tones with or without

pitch gestures. The authors observed better tonal production accuracy and wider pitch range when producing Mandarin Chinese words together with gestures, as well as higher discrimination scores in the group who were exposed to and produced themselves the gestures, regardless of their tonal or nontonal language background. However, in this study, many aspects such as the participants' background, the materials and the procedure were not controlled. Interestingly, Zheng et al. (2018) only found a limited effect of pitch gesture production during the pronunciation of tones. They trained 24 English naïve learners of Mandarin Chinese with Chinese monosyllabic words in one of three conditions: speech only, head nods or pitch gesture. Participants' oral production during training was acoustically analysed in terms of F0 and results showed no difference between the groups, except for the falling tone F4, with better F0 accuracy in the gesture group, and for the dipping tone F3, with better F0 accuracy in the head nod group.

Regarding perception outcomes, Morett and Chang (2015) taught 57 English speakers to learn 12 minimal pairs of Mandarin words differing in tone in three different conditions: imitating pitch gestures depicting the specific contours of each Mandarin tone, imitating iconic gestures depicting the meaning of the Mandarin words, and no gesture. Their results showed that imitating pitch gestures facilitated the discrimination between the meanings of Mandarin words differing in tone while the use of iconic gestures and no gesture did not. These findings provide evidence that

participants map visuospatial information conveyed by the pitch gestures on their representations of lexical tones and associate these phonological representations with the semantic representation of referents. These findings support the view that phonological representations of words may be activated prior to conceptual representations of referents (e.g., Poss et al., 2008; Van Donselaar et al., 2005). However, surprisingly, no difference was found between the three groups in terms of tone identification. However, Hannah et al. (2017) found a positive effect of pitch gesture production on Mandarin lexical tone perception. They asked 25 English speakers to listen to and to identify monosyllabic words with the four tones embedded in noise and presented with congruent and incongruent, facial-only and facial-gestural information mimicking the melodic movements of the lexical tones. Results showed that participants could more accurately identify tones with congruent auditory and facial or facial-gestural information. In addition, the facial-gestural/congruent condition obtained significantly better scores in tone identification than participants in the facial-only/congruent condition, showing the additional beneficial effect of gesture. In contrast, Zheng et al. (2018) only found a limited effect of pitch gesture observation when simultaneously imitating the lexical tones, with a significant improvement in identification only on the falling tone. Crucially, Zhen et al. (2019) examined the role of pitch gestures on the perception of lexical tones by controlling a set of parameters:

congruency of hand gesture movement with pitch contour, modality of perceiving or reproducing the gestures, and spatial orientation of the movement, either horizontal or vertical. They found that gesture observation and production equally benefited the perception of lexical tones compared to speech only, as long as they were congruent with the pitch contours. In addition, when performed horizontally, performing the hand gestures was found significantly more helpful than perceiving them.

In sum, previous studies on pitch gestures have started to show their beneficial role for tonal and intonational perception and production in a foreign language. However, more research is needed regarding the effect of pitch gesture on lexical tone perception and still little is known between the potential effects of observing versus producing pitch gestures during embodied training. The main goal of the first study of this dissertation (Study 1) will therefore address both questions.

b) Vowel durational contrasts

In a similar fashion to tonal contrasts, in some languages such as Japanese or Finnish, the variation in duration of a vowel (short vs. long) in otherwise identical words signals a distinction in meaning (Odden, 2011). In other languages, such as English, Swedish, or Cantonese, vowel length contrasts involve both vowel duration and vowel quality (Odden, 2011). However, many languages do not use duration as a cue to distinguish vowel categories and for this

reason, vowel durational contrasts are considered a difficult feature to acquire for L2 learners (e.g., W. Chang, 2018; Luo et al., 2019, McAllister et al., 1999).

Durational gestures are visuospatial hand gestures matching the length of the corresponding sound (typically vowels) or group of sounds (typicallys syllables). Evidence from sound processing experiments show that durational contrasts in speech may be adequately represented by horizontal hand movements. Research suggests that the perception of the abstract concept of time, and by extension temporal duration, is grounded to sensorimotor experiences related to the domain of space. According to the spatial metaphor account, people employ spatial metaphors in thinking or talking about time such that they use their concrete spatial experience to support their understanding of abstract time processing (Boroditsky, 2000; Gibbs, 2006; Lakoff & Johnson, 1980, 1999). The temporal relation of two events can be expressed metaphorically as a relation between two locations in space (e.g., tomorrow is ahead of yesterday) or as the distance from a spatial location representing the onset of the duration and a spatial location representing the offset of the duration. In a nonlinguistic experiment, Casasanto and Boroditsky (2008) reported evidence on the relationship between horizontal visuospatial movements (i.e. spatial displacement) and the phonological representation of duration (i.e. temporal duration). They showed ‘growing’ horizontal lines of varying lengths representing different durations

to nine participants and then asked them to estimate via mouse clicks either the length of the line or its duration of presentation. To estimate displacement, subjects clicked the mouse once on the center of the X, moved the mouse to the right in a straight line, and clicked the mouse a second time to indicate that they had moved a distance equal to the maximum displacement of the stimulus. To estimate duration, subjects clicked the mouse once on the center of the hourglass icon, waited the appropriate amount of time, and clicked again in the same spot. Results showed that information about spatial length influenced judgments of temporal duration. Cai and Connell (2012) confirmed the link between time and space and the determinant role of perception by examining the interaction between time and space as a function of the haptic sensory modality. Twenty-six participants estimated the length of a stick while listening to a note during a specific amount of time in two conditions: haptic-only (i.e., tactile and proprioceptive) or haptic-visual perception. As in Casasanto and Boroditsky (2008), participants attended to both the spatial length and temporal duration and then reproduced either length or duration. When visual and haptic modalities were acting together, the perception of spatial duration strongly affected their perception of temporal duration, corroborating the findings by Casasanto and Boroditsky (2008). However, when participants could only touch the stick but not see it, time perception was not affected. The authors suggested a two-way interdependence between time and space, mediated by

the sharp acuity of the visual modality. Crucially, in a follow-up study, Cai et al. (2013) found that short and long horizontal hand gestures accompanying the emission of a musical note significantly modulated participants' estimation of temporal duration.

In languages like Japanese, the duration of syllables and more particularly of vowels is a crucial contrastive element. Studies focusing on the effects of using durational gestures cueing vowel length contrasts in Japanese for phonological learning have yielded mixed results. First, Hirata and Kelly (2010) reported that observing a short vertical chopping movement (similar to a beat gesture) during the production of short vowels and a long horizontal hand sweep gesture for long vowels during training did not help English naïve learners of Japanese to better perceive the vowel durational contrasts (replicated in Kelly et al., 2017). Later, Hirata et al. (2014) compared the effects of the same gestures, named syllable gestures (one beat for short vowels and a horizontal hand sweep for long vowels), with mora gestures (two beats). Results showed that the observation of syllable gestures (as opposed to mora gestures) facilitated the perception of the durational contrast in a balanced manner both in word-initial and word-final positions as well as at both fast and slow speech rates. In a follow-up study, Kelly et al. (2014) trained 88 English speakers to learn Japanese bisyllabic words by either observing or producing syllable gestures and mora gestures and did not find any difference between them in terms of auditory learning in four different

conditions: syllable gesture observe, syllable gesture produce, mora gesture observe, mora gesture produce. Finally, using the ERP data collected in Kelly et al. (2014), Kelly & Hirata (2017) examined the neural correlates of these four conditions and again, did not find any difference between conditions. The authors concluded that hand gestures only had a limited effect on the perception of Japanese durational contrasts. However, in a recent training study, P. Li et al. (2021) slightly changed the gesture configuration employed in the experimental design by using only horizontal hand gestures and testing both perception and production effects. They examined the effects of a long vs. short horizontal sweeping gesture mimicking vowel length contrasts on the learning of Japanese minimal pairs of words by 50 Catalan naïve learners of Japanese. While no advantage was found for the perception of the contrast, results showed a positive effect of this gesture on the pronunciation of the words. Participants in the gesture group obtained a greater improvement in pronunciation and target-like vowel durations at posttest in a word imitation task, compared to the group who did not train with gestures.

c) Phrasal rhythm and melody

Rhythm is a speech property related to the temporal organization of sounds in terms of grouping (e.g. Jun, 2005) and emerges from phonological properties such as syllable structure, phonotactics, and prosodic contrasts at the lexical and postlexical levels

(Astésano, 2001). According to Kohler (2009, p. 41), rhythm is "the production, for a listener, of a regular recurrence of waxing and waning prominence profiles across syllable chains over time. Salient and less salient syllables form the metrical patterning of utterances and for a specific language, regular metrical structures allow for a degree of rhythmic predictability. For example, in French, a final stressed syllable marks the end of an intonational phrase (Di Cristo & Hirst, 1993). Evidence from first language acquisition shows the crucial role of rhythm perception in language development (e.g., Gordon et al., 2015; Johnson & Jusczyk, 2001; Morgan & Saffran, 1995; see Bharata et al., 2018; Thorson, 2018 for reviews) and language processing (e.g., Magne et al., 2007; Pitt & Samuel, 1990; Roncaglia-Denissen et al., 2013). As rhythm is language-specific and of utmost importance for language development and phonological processing, it is essential that pronunciation instruction takes into account the problems of foreign language learners when facing rhythmic differences across languages (see section 1.2.1).

Auditory priming studies by Cason and collaborators have shown that the phonological processing of speech by adult participants is enhanced by the temporal expectancy generated by a musical rhythmic prime (Cason & Schön, 2012; Cason, Astésano, et al., 2015). First, Cason and Schön (2012) presented French participants with matching and mismatching percussive rhythmic primes followed by nonwords respecting French phonotactics, and asked

them to state whether a target phoneme had been pronounced in the nonword. Behavioral measures in the form of reaction times showed that target phonemes were detected faster when positions matched the prime beat. Additionally, when a beat expectancy violation occurred, ERP measurements showed a larger-amplitude and longer latency response at P300. These findings were successfully reproduced in a follow-up study (Cason, Astésano, et al., 2015) with spoken sentences in French preceded by a prime musical meter to induce metrical expectancy about both stress patterns and the number of syllables. Additionally, in this study, a group of participants underwent a short audio-motor training session several times during the experiment (just before and halfway through each block) which consisted of vocally repeating the prime rhythm using different sounds to distinguish between strong and weak musical beats. The results revealed that the priming effect was enhanced by the audio-motor training. In an EEG (electroencephalography) study, Falk, Lanzilotti, et al. (2017) presented participants with sentences in French which were preceded by matching or non-matching musical rhythmic primes and observed that phase coupling, i.e. the synchronisation between auditory rhythm and neural oscillations, was enhanced by the rhythmic auditory input when the latter was coupled with accented syllables. Their findings support the hypothesis that explicit rhythmic cues that map onto speech metrical structure enhance temporal expectancy and facilitate the processing of upcoming

events in speech at predicted times (see also Falk & Dalla-Bella, 2016; Falk, Volpi-Moncorger, et al., 2017; Kotz & Gunter, 2015).

Rhythmical auditory priming may therefore help learners parse speech input through its prosodic structure and help identify the salient parts of speech. Importantly, rhythm and acoustic prominence in speech can be highlighted by visual and gestural features. Ghaemi and Rafi (2018) compared the effects of gesture, printed visual cues and auditory input on the learning of English word stress. In the three conditions, English words were printed largely on a piece of paper and the syllables were clearly specified by dots. In the first group, pronunciation and stress patterns of new words were taught aurally through the repetition of the words. In the second group, the stressed syllables were additionally printed in bold. Finally, in the third group, the stressed syllables were not only printed in bold, but also emphasized by the teacher's hand gesture. The hand gesture consisted in a forward, horizontal hand movement during unstressed syllables and upward movement during the stressed syllables. Although the three groups showed an improvement between pre- and posttest, training with gesture yielded a significantly larger improvement in the memorization of word stress patterns two weeks after training.

Beat gestures have been typically associated with prosodically prominent positions in speech and they have been shown to trigger stronger perceptions of prominence. Krahmer and Swerts (2007)

found that beat gestures have similar effects to pitch accentuation such that when these gestures are produced together with pitch accentuation on a given syllable, they lead to stronger perceived prominence. Interestingly, beat gestures may thus be useful to highlight foreign language rhythmic patterns. Gluhareva and Prieto (2017) trained 20 Catalan learners of English to watch and listen to native English instructors producing a set of discourse-embedded responses, either accompanying their speech with beat gestures on prosodically prominent segments or without gestures. When tested on the same context prompts, participants who were exposed to the beat gesture condition during training were rated as less accented than those who did not on a set of difficult items. These results showed that participants may have better perceived and consequently produced prominence patterns thanks to the beat gestures. In a follow-up study focusing on production, Kushch (2018) asked Catalan learners of English to either observe or imitate the discourse with beat gestures produced by instructors and found that producing beat gestures while imitating speech helped reduce accentedness at posttest more than observing beat gestures. To further explore the benefits of beat gestures, Llanes-Corominas et al. (2018) encouraged adolescent low-intermediate Catalan learners of English to intentionally produce beat gestures during an oral reading task and found that these participants obtained greater improvement in terms of accentedness, comprehensibility, and fluency in an oral reading task compared to participants who were

not instructed to move their hands. However, for lexical stress acquisition, results on the usefulness of beat gestures have been inconclusive.

Van Maastricht, Hoetjes, et al. (2019) taught Spanish lexical stress to 62 Dutch naïve learners of Spanish with cognate words embedded in short sentences. Participants followed a short audiovisual training session in one of three conditions: speech only, where the instructor did not move her hands; beat gesture, where the instructor produced a beat gesture while uttering the stressed syllable; and metaphoric gesture, where the the instructor produced a metaphoric gesture while uttering the stressed syllable. The metaphoric gesture represented the lengthening of the syllable, which is a clear correlate to lexical stress in Spanish (the instructor started with joined hands, then moved both hands to each side, then back together; see also section 1.3.1.2 on durational gestures). Participants were tested before and after training on a sentence reading task and the target words were extracted and their stress production was categorized as Learning (incorrect at pretest and correct at posttest), Always Able (correct both at pre- and posttest), Never able (incorrect both at pre- and posttest), or Unlearning (correct at pretest but not at posttest). Results did not show any difference between the three groups in terms of lexical stress production accuracy and no advantage was observed neither for the beat gesture nor for the metaphorical gesture.

As mentioned in section 1.2.5, one of the specificities of the verbotonal method is the importance given to hand gestures in the teaching of the pronunciation of the target language. Billières (2017) especially recommended a technique consisting in combining logatomes (i.e. the repetition of nonsense syllables instead of the actual words in a phrase) and hand gestures to teach phrasing, pitch movements and stress placement. According to Billières (2017), hand gestures can help understand the rhythmic and intonational structure of utterances by 'drawing' it in space. For example, to teach the prosody of declarative French sentence that is composed by two prosodic phrases, one of the teachers hands can start horizontally and move upward to indicate the melodic contour of the first prosodic phrase, mark a short pause, then move downwards to indicate the falling melodic contour of the second prosodic phrase, and finally indicate the final lengthening of the stressed syllable by lengthening the final downward movement. In the present dissertation, this type of gesture will be named *phrasal-level prosodic gesture*. Alazard et al. (2010) conducted an eight-week phonetic training course with 4 English-speaking beginner learners of French. Two learners worked their oral skills through the verbotonal method, mainly focusing on prosodic patterns by using phrasal-level prosodic gestures, while two other learners worked mainly on oral reading, text comprehension and creative writing with a communicative approach. Acoustic and perceptual analysis of learners' oral reading productions after

training showed a higher improvement in fluency in the group that followed the verbotonal method. Subsequently, Alazard (2013) compared the difference between the effects of the verbotonal method and the articulatory method, expliciting metalinguistic knowledge about the articulation of the sounds of the target language. For eight weeks, at the rate of two sessions per week, 20 English-speaking learners of French participated in a pronunciation course with one or the other method. Results showed better reading fluency in learners who followed the verbotonal method after three weeks of training, especially when the learner's level was lower. However, this advantage disappeared after eight weeks. According to the author, this could be due to the introduction of reading exercises during the sessions after three weeks. Finally, the same oral reading productions were further analyzed by focusing on the pronunciation of vowels (Alazard-Guiu et al., 2018), but these analyses did not reveal any difference between the two methods.

In general, the studies mentioned in this section have tried to assess the effects of the verbotonal approach as a whole by using specific hand and body movements that are documented in the verbotonal method (e.g., Renard. 2002). Yet, to our knowledge, no previous empirical investigation has assessed the effects of using specific types of hand gestures representing phrase-level rhythmic and melodic features on pronunciation. Hence, the third study (Study 3) of this dissertation will experimentally assess this issue.

1.3.2 Kinesthetic and tactile movements representing prosodic features

Based on evidence that language processing can be enhanced by multisensory integration (e.g., Atligan et al., 2018; Atilgan & Bizley, 2021; Helfer & Freyman, 2005), several studies have claimed that multisensory approaches would enhance language teaching and learning (see Minogue and Jones' (2006) systematic review of studies exploring first language acquisition). For foreign language phonological learning, a handful of empirical studies have shown that visual information can enhance auditory perception and the acquisition of novel speech sounds (e.g., Hardison 2003, 2005; Hazan et al., 2005, 2006; Hirata & Kelly, 2010; Inceoglu, 2016; Y. Li & Somlak, 2017). However, little is known about the potential effects of kinesthetic and tactile training on pronunciation (but see Esteve-Gibert et al., 2021; Ozakin et al., 2021, for recent studies on segmental learning). In this section, I review the teaching methodologies and studies that have encouraged kinesthetic activities for the learning of L2 prosody.

Odisho (2007) made the case in favor of teaching pronunciation based on a multisensory and multicognitive approach. In addition to using an aural modality, the author suggested complementing ear training with visualization and the tactile/kinesthetic experiences of sound production. For example, one of the techniques proposed to teach stress placement consists of self-tapping strong and weak

beats with the hand on the chest while pronouncing a sentence to feel the rhythm of the sentence. Another recommendation to work on stress is to walk around making short and long steps while uttering the sentence. One of the few pronunciation teaching methods that fully integrates the use of body and gestures is the one developed by Acton and colleagues (2013) for English, called the "haptic-integrated English pronunciation (EHIEP) framework". Acton et al. (2013; see also Burri & Baker, 2016, 2019; Burri et al., 2019) advocates that the imitation of the voice, body movements and facial expressions and the involvement of haptic (i.e. kinesthetic) techniques help noticing prosodic elements and promote their memorization and their integration into real exchanges. This haptic approach comprises a set of touching techniques (e.g., 'Butterfly', 'Touchinami','Tai Chi', 'Rhythm Fight Club') which involve asking learners either to 'self-touch' (i.e., to touch a part of their own bodies) or to touch a physical object, as well as body movements and gestures. In the 'Butterfly' technique, learners mark the rhythm of words by tapping one shoulder with one hand when uttering a stressed syllable and tapping the elbow when uttering an unstressed syllable, and in the 'Touchinami' technique, learners observe and perform sweeping hand movements and a systematic final touch of the opposite hand to mimic intonational patterns while uttering declarative statements and yes/no questions. In the 'Tai Chi' technique, learners hold a ball and stretch their arms to learn the stressed syllables, and in the

‘Rhythm Fight Club’, learners perform boxing-like movements to mimic syllable stress. In a recent qualitative study, Burri & Baker (2019) taught the haptic techniques to 15 teachers of English, who reported the haptic techniques to be highly engaging and beneficial, suggesting that the incorporation of touch, movement, and hand gestures can be of great interest to language teachers.

With a more artistic perspective, Haught and McCafferty (2008) proposed that interpreting roles within a theatre activity allow learners to imitate the prosody and the body movements of the teacher and improve L2 fluency. Similarly, Soulaine's (2013) study encouraged body movements and gestures inspired by dramatic expression and dance to improve stress and rhythm in French learners of English. Also based on theatre practice, Llorca (2001) offered videos with practical activities where body movements and gestures allow students to better perceive French and English prosody. She suggested that learners should observe how the modification of a gesture leads to the modification of the voice and to make them aware of the coordination between gesture and spontaneous speech.

Following an embodied method involving tactile information reminiscent of Acton’s EHIEP framework (see section 1.3.1), Hamada (2018) trained 58 Japanese learners of English during 15 group lessons to pronounce sentences with one of the following techniques: either 'haptic-shadowing', where learners were required

to produce light punches on each word and a more pronounced punch on the accented words of the sentence; or “IPA-shadowing”, in which learners could read a transcription of the sentence in the International Phonetic Alphabet (IPA), an internationally recognized set of phonetic symbols based on the principle of one-to-one correspondence between sounds and symbol. After training, both groups improved the comprehensibility and pronunciation of segmental features, but only the "haptic-shadowing" group, where sentence rhythm and stress were made salient, improved the pronunciation of suprasegmental features.

Yang (2016) tested the effects of integrating body movement to computer-assisted language learning with Chinese primary school children learning English. In the control group, participants merely repeated sentences after the teachers while in the experimental group, participants listened to the same sentences modified through a low pass-filter so as to remove all segmental information from the sentences. In that way, only prosodic information was available for participants to perceive. In addition, participants in the experimental group were encouraged to perform body movements like hand-clapping or walking along with the melody. Then, all the participants were able to record and compare their pronunciation to native models. Results showed that participants who followed embodied training improved pronunciation, comprehensibility, and fluency more than the control group.

In her dissertation, F. Zhang (2006) reported positive effects of activities inspired from the verbotonal approach and kinesthetic activities in the acquisition of Chinese prosody by 22 English-speaking learners of Chinese. The 'somatically-enhanced' approach proposed by the author included a session of body relaxation to reduce learners' anxiety and improve their receptiveness for learning (p. 150). Another important feature of the approach was hand-clapping to the rhythm of sentences (see below) combined to rhythmic displacement in a circle. Learners were also encouraged to use gestures and body movements associated with each of the four Chinese tones while humming (an alternative to logatomes) and producing sentences. Interestingly, the chosen gestures were not related to a visuospatial metaphor of intonation contours but were illustrating the degree of *tension* taking place in the vocal cords when pronouncing the lexical tones (pp. 158-160). As the production of tone 1 requires the vocal cords to stay tense during a certain amount of time, she proposed to push both hands upwards as though trying to touch the ceiling with fingers spread out and palms facing upwards. For tone 2, the vocal cords are at first neither tense nor lax, but then become tense rapidly, therefore, starting with a forward slumping of the shoulders or the head, learners tense up their arms and gradually push their hands up directly over their heads, using very tense hands with the fingers spread out and the palms facing upwards. For the low level Tone 3, learners adopted a relaxed, forward slumping of the

shoulders accompanied by a forward motion of the head similar to nodding to produce the sound. Finally, for Tone 4, the vocal cords suddenly tense and then gradually become more lax. Learners raised their hands up high and then relaxed their body by bending their head forward. Compared to training with a communicative approach alone, participants who were encouraged to use body movements during the learning process obtained higher intelligibility rating scores, higher mean F0 values, wider pitch range, and more accurate tonal patterns. In addition, the experimental group of students vocalised more by producing longer and more complex utterances and showed stronger motivation scores.

A type of rhythmic percussive hand movement that is starting to raise interest in pronunciation research is hand-clapping, an activity that lends itself very easily to the classroom context as it does not require any equipment and can be easily performed by learners of all ages. By hand-clapping to the rhythm of spoken words or sentences, learners are able to kinesthetically reproduce the prosodic structure of those target words or sentences. To our knowledge, only three studies have recently tested the effect of hand-clapping on pronunciation learning (see also B. Lee, 2020, on the effect of hand-clapping on comprehension).

In a two-week training study, B. Lee et al. (2020) compared the effects of prosodic training with hand-clapping to training with oral

repetition and training with explicit segmental instruction on the perception of L2 suprasegmental features by 111 Japanese learners of English. In the perception-based group, participants practiced the identification and counting of syllables with hand-clapping. First, the instructor would utter a word or a phrase and the learners would clap their hand on the syllables, indicating the position and relative strength of the syllables. When errors were made, the instructor gave corrective feedback either by repeating the target overemphasizing stress and syllable segmentation or by demonstrating a proper clapping rhythm. This activity was followed by an exercise engaging learners to perceive the differences between standard American English pronunciation and Japanese-accented (American) English pronunciation and heighten learners' metalinguistic awareness on this issue. In the production-based group, participants orally imitated the words and phrases pronounced by the instructor, and in the explicit pronunciation group, learners were given explicit descriptions of the phonemes and were trained to identify them. Learners were tested on both controlled and spontaneous production tasks before and after training and their pronunciation accuracy was rated perceptually by three evaluators. Results showed that despite the fact that all the groups improved, the perception-based group obtained significantly larger gains, in particular at the delayed posttest.

Iizuka et al. (2020) examined whether hand-clapping had an effect on the acquisition of Japanese long phonemes, specifically long vowels, moraic nasals, and geminates, by English learners of Japanese. Thirty-one beginner English-speaking learners of Japanese learned loanwords (Japanese words adopted from English) either with or without hand-clapping performed by the instructor and the learners themselves. At pre-, post- and delayed posttest, phoneme identification was tested in a dictation task while phoneme pronunciation was assessed through a picture elicitation task. Overall, findings indicated a positive impact of hand-clapping on receptive knowledge, but only a small impact on productive knowledge.

Finally, Y. Zhang et al. (2020a) investigated the benefits of hand-clapping to the rhythmic structure of words on pronunciation. During a short audiovisual training session, 50 Chinese adolescents learned a set of unknown French words by watching an image conveying their meaning and by repeating the words after an instructor in two between-subject conditions: while one group of participants only repeated the words, another group imitated not only the words but also the hand-clapping produced by the instructor. The participants were tested using an oral imitation task before and after training. Accentedness ratings revealed only a nearly-significant difference in improvement between the two groups. However, an acoustic analysis of the relative duration of the final stressed syllable in the target words showed a significant

improvement between pre- and posttest for the clapping condition only, showing that participants who performed hand-clapping while pronouncing the words during training lengthened the final syllable more appropriately than participants who were not trained to clap.

Overall, given the mixed results obtained in previous studies, further evidence is needed to assess the effects of hand-clapping on L2 pronunciation. To explain the lack of effect of hand-clapping on reducing accentedness, Y. Zhang et al. (2020a) suggested that learning the meaning and the pronunciation of the words at the same time may have resulted in cognitive overload and reduced the effects of clapping on pronunciation. In Study 2, we will explore the effects of training Catalan-speaking children with hand-clapping while they learn a set of French words, using a similar design as in Zhang et al. (2020a). However, crucially in order to allow participants to focus exclusively on pronunciation rather than word meaning, the target words will be Catalan-French cognates (words with similar phonological patterns and same meaning in both languages, e.g. French 'téléphone'– English 'telephone'). Importantly, while the transparency of lexical meaning offered by cognates can facilitate comprehension and word memorization, the similarity in phonological forms may enhance phonological transfer from their L1, thus penalizing pronunciation (Amengual, 2012; Flege, 1987; Goldrick et al., 2014; Mora & Nadeu, 2012).

1.4 Scope of the thesis, main goals, research questions, and hypotheses

The present dissertation focuses on the role of embodiment in the acquisition of phonological features in a foreign language. The main aim of the thesis is to empirically assess the potential benefits of embodied prosodic training with visuospatial hand gestures and percussive hand movements on the acquisition of a set of non-native phonological features, both at the perceptive and the productive levels. While the Embodied Cognition paradigm supports the benefits of visualizing and producing body movements on language comprehension and lexical processing (e.g. Glenberg & Kaschak, 2002; Glenberg et al., 2008; Myung et al, 2006 among others), less is known about possible benefits on phonological learning. Crucially, we adopt a multisensory approach (visual, auditory, and kinesthetic) that can be easily implemented in the classroom and which relies on visuospatial and kinesthetic prosodic training paradigms. The three studies in this dissertation will determine the effects of training students with visuospatial hand gestures and percussive hand movements (e.g., hand-clapping) in terms of quantitative pronunciation gains.

The present dissertation includes three training studies using the same between-subject, pre- and post test design which directly compares three different types of embodied prosodic teaching

techniques (i.e., pitch gestures, phrase-level prosodic gestures, and hand-clapping) to conventional listen-and-repeat techniques. The aim of the three training experiments is to improve on the following phonological features: (a) the perception of novel tonal patterns in minimal word pairs in a tonal language (Study 1); (b) the pronunciation of novel cognate words involving longer word-final syllables (Study 2); and (c) the pronunciation of sentences in a second language (Study 3). The main research questions for each of the three studies are the following:

(1) Does embodied training with pitch gestures improve perceptive phonological learning of tones at the syllabic level?
(2) Does embodied training with hand-clapping improve productive phonological learning at the word level?
(3) Does embodied training with phrase-level prosodic gestures improve productive phonological learning at the sentence level?

Following the L2 perceptual acquisition models (see section 1.2.1), the training paradigm followed in the three studies of the dissertation always includes the perception of a model, both in terms of observing and producing speech and hand movement, reinforcing the importance of perception and imitation. The type of activity used in the three training studies is an elicited imitation task (e.g., Vinther, 2002). In this type of task, the participant listens to a cue/training stimulus performed by a model speaker and immediately following the presentation of the stimulus, the

participant repeats the stimulus orally. Gallimore and Tharp (1981) evaluated the technique of the elicited imitation task and found that this task yields stable test–retest correlations over a period of years, that it is related to language behavior in natural settings, and that it reflects stages of language development, among other things. In the experimental groups in each of the three studies of the present dissertation, the concept of imitation was extended to the observation and the reproduction of a hand movement simultaneously with the oral stimulus in order to create an embodied training paradigm.

While Studies 1 and 2 used naive learners of the language, that is, participants who did not know the target language nor were actively learning the target language outside of the experimental setting, Study 3 used actual language learners in a classroom setting. Therefore, for each study, we ensured that the participant understood the meaning of the stimuli, either by providing the orthographic transcription (Study 1), by eliciting the meaning with an image (Study 2), or by providing the translation of the difficult words (Study 3). In order to focus on phonological learning, it was ensured that the length of the stimuli could be handled by the participants in terms of their working memory capacity. Hence, while naïve learners in Study 1 dealt with monosyllabic words, those in Study 2 were presented with transparent/cognate words. In Study 3, the vocabulary, grammatical difficulty and the length of

the phrases in the training phase were adequate for the proficiency of these specific learners.

Regarding the choice of measures to test phonological learning, phonological perception was evaluated by means of an identification task and a word-meaning association task (Study 1), and pronunciation was evaluated in terms of comprehensibility, fluency, and accentedness, as well as segmental and suprasegmental features in a dialogue reading task (Study 2). In Study 3, pronunciation was assessed through a cognate word imitation task. Participants' pronunciation was evaluated using two measures: by assessing accentedness as a global perceptual measure of pronunciation, and by acoustically measuring the relative duration of the prominent syllable to assess rhythmic patterns. Therefore, one of the goals of the dissertation was to evaluate the direct effect of embodied prosodic training on the learning of suprasegmental features (lexical tones identification in Study 1, suprasegmental accuracy in Study 2, and duration of the prominent syllable in Study 3) and on global assessments of perceived pronunciation.

In the upcoming chapters, the three studies that constitute the body of the dissertation are presented. A summary of each study is offered below:

- Chapter 2 (Study 1): Observing and producing pitch gestures facilitates the learning of mandarin chinese tones and words

This study investigated the role of observing and producing pitch gestures mimicking tonal movements over a syllable on the phonological learning of Mandarin Chinese tones in terms of tone perception and meaning retrieval of monosyllabic words contrasting only in tone. In a laboratory setting, a total of 106 Catalan adults with no previous knowledge of Chinese learned minimal pairs of Chinese monosyllabic words contrasting in lexical tones during a short training session either by observing the instructors' pitch gestures vs. observing no gesture (Experiment 1) or imitating instructors' pitch gestures while repeating the words aloud vs. observing pitch gestures silently (Experiment 2). We predicted that an embodied prosodic training involving observing or observing and producing pitch gestures would (a) enhance participants' identification of Mandarin Chinese tones and (b) improve the retrieval of word meaning when presented as minimal pairs of tonal contrast (Experiment 1). Furthermore, we hypothesized that producing the gestures would benefit participants more than observing them (Experiment 2). The results of the tone identification and word-meaning association tasks at pre- and posttest were assessed by means of binary accuracy scores (0 = not accurate / 1 = accurate).

- Chapter 3 (Study 2): Embodied prosodic training helps improve L2 pronunciation in an oral reading task

This study investigated the role of phrase-level prosodic hand gestures depicting speech rhythm and intonation during the oral repetition of logatomes (i.e., a series of identical nonsense CV syllables that maintain prosodic structure intact) on the pronunciation of sentence-level prosody in read speech. As part of their language course, seventy-five Catalan learners of French participated in three training sessions to improve their oral reading of short dialogues in one of three conditions: repeating sentences, repeating logatomes and sentences, and repeating prosodic gestures, logatomes and sentences. We hypothesized that embodied prosodic training with repeating both gestures and logatomes before repeating the sentences would help learners improve their oral-reading pronunciation of the trained dialogues and that the benefits of embodied prosodic training would also generalize to an untrained dialogue. Participants' oral production was evaluated by three native speakers of French on five Likert scales from 1 to 9 in terms of fluency, comprehensibility, accentedness, and accuracy of suprasegmental and segmental features.

- Chapter 4 (Study 3): Embodying rhythmic properties of a foreign language through hand-clapping helps children to better pronounce words

This study investigated the role of performing rhythmic hand-clapping on the syllabic structure of words on the phonological learning of cognate words in terms of pronunciation of the words and of the prominent syllable. In a laboratory setting, twenty-eight 7- to 8-year-old Catalan children with no previous knowledge of French learned cognate words in French (e.g. French *aspirateur*, Catalan *aspirador* 'vacuum cleaner') during a short training session either by imitating the instructor's native pronunciation of the words while clapping to the rhythmic structure of those words or only by repeating the words without seeing and imitating hand-clapping. We predicted that children who participated in the embodied rhythmic training condition would significantly improve their pronunciation of the target words more than children who participated in the Non-Clapping condition and only repeated the target words. Participants' oral productions were rated for accentedness by three French native speakers on a Likert scale from 1 (not accented at all) to 9 (very accented) and an acoustic analysis of the duration of word-final vowels was carried out.

All in all, the three studies included in this dissertation propose three different types of embodied training for phonological learning

that are couched in a multisensory approach, using a set of hand gestures and percussive movements (e.g., pitch gestures, phrase-level prosodic gestures, and hand-clapping). While Study 1 tests perception skills, Study 2 and Study 3 assess pronunciation gains through global perceptive measures and acoustic analyses. In addition, variation in the population under scrutiny allows us to assess the effects of an embodied training paradigm both for adults and children, and for naïve and true foreign language learners.

2

CHAPTER 2: OBSERVING AND PRODUCING PITCH GESTURES FACILITATES THE LEARNING OF MANDARIN CHINESE TONES AND WORDS

Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2019). Observing and Producing Pitch Gestures Facilitates the Learning of Mandarin Chinese Tones and Words. *Studies in Second Language Acquisition*, 41(1), 33-58. doi:10.1017/S0272263118000074

2.1 Introduction

Tonal languages like Mandarin Chinese, as opposed to intonational languages like English or Catalan, use pitch variations at the word level—that is, lexical tone contrasts—to distinguish meanings between otherwise segmentally identical words (Xu, 1994). For speakers of non tonal languages, acquiring these lexical tones has been shown to be particularly difficult (e.g., Kiriloff, 1969; Wang, Perfetti, & Liu, 2003b). Despite this intrinsic difficulty, there is evidence that speakers of both tonal and non tonal languages can be trained with success in both the perception and production of L2 tonal systems (e.g., Francis, Ciocca, Ma, & Fenn, 2008; Hao, 2012; Li & DeKeyser, 2017, among many others). Laboratory research has shown that learners of non tonal languages can be successfully trained to discriminate Mandarin tones by using short auditory tone training procedures consisting of paired combinations of tones both in perception (e.g., Wang, Spence, Jongman, & Sereno, 1999; Wang, Jongman, & Sereno, 2003a; Wong & Perrachione, 2007) and in production (Wang et al., 2003a). Very recently, Li and DeKeyser (2017) showed the importance of specificity of practice in the learning of tones, in the sense that training in perception or production led to progress in only that skill area, not both. They found that after a three-day training session, participants who learned 16 Mandarin tone words in the perception condition obtained better results in perception post-tasks, while participants

trained in the production condition obtained better results in production post-tasks.

In general, a challenge for educational research is to assess the procedures that can reinforce the teaching of a different prosodic system, such as the use of visualizers, gestures, or supporting transcription systems. In this respect, Liu et al. (2011) showed that having the support of visual illustrations depicting the acoustic shape of lexical tones (together with pinyin spelling of the spoken syllables) can help facilitate their acquisition. Research in gestures and second language acquisition has described the positive effects of observing iconic gestures on vocabulary learning (e.g., Kelly, McDevitt, & Esch, 2009, among others) as well as the positive effects of beat gestures on both L2 pronunciation learning and vocabulary acquisition (e.g., Gluhareva & Prieto, 2017; Kushch, Igualada, & Prieto, 2018, among others). However, little is known about the supportive use of gestures in learning pitch modulations in a second language, as well as potential differences between the benefits of perception and production practices. This study examines the role of pitch gestures, a specific type of metaphoric gesture that mimics melody in speech, in the learning of L2 tonal features, and focuses on the potential benefits of observing versus producing these gestures in the context of pronunciation learning.

2.1.1 Multimodal cues and lexical tone perception

Research in second language acquisition has shown that access to audiovisual information enhances nonnative speech perception in general (see Hardison, 2003, for a review). A series of studies have reported that when it comes to learning novel speech sounds, language learners benefit from training that includes both speech and mouth movements compared to just speech alone (e.g., Hardison, 2003; Hirata & Kelly, 2010; Wang, Behne, & Jiang, 2008). With respect to the learning of novel tonal categories, research has shown that having access to visual information about facial articulators has beneficial effects on tone perception for both tonal-language speakers in their native language (e.g., Burnham, Ciocca, & Stokes, 2001; Reid et al., 2015) and non tonal language speakers (e.g., Chen & Massaro, 2008; Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004; Reid et al., 2015; Smith & Burnham, 2012). For example, Reid et al. (2015) tested the role of visual information on the perception of Thai tones by native speakers of typologically diverse languages, namely three tonal languages (Thai, Cantonese, and Mandarin), a pitch-accented language (Swedish), and a non tonal language (English). The results of a tone discrimination test in audio only (AO), audiovisual (AV), and visual only (VO) conditions showed a significant increase in tone perception when auditory and visual (AV) information was displayed together. Similarly, eyebrow movements (Munhall et al., 2004) and the visible movements of the head, neck,

and mouth have been found to play a beneficial role in the perception of lexical tones (Chen & Massaro, 2008).

2.1.2 Gestures and L2 word learning

It is becoming increasingly clear that co-speech gestures (i.e., the hand, face, and body movements that we produce while we speak) are an integral aspect of our language faculty and form an integrated system with speech at both the phonological (i.e., temporal) and semantic-pragmatic levels (e.g., Bernardis & Gentilucci, 2006; Goldin-Meadow, 2003; Kendon, 2004; McNeill, 1992, 2005). Concerning co-speech hand gestures in particular, there is ample evidence of the cognitive benefits of their use in educational contexts (e.g., Cook, Mitchell, & Goldin-Meadow, 2008; Goldin- Meadow, Cook, & Mitchell, 2009). A growing body of experimental research in second language acquisition has shown that co-speech gestures can be used as an effective tool to help students improve their language skills (Gullberg, 2006, 2014; see Gullberg, deBot, & Volterra, 2008, for a review on gestures in L1 and L2 acquisition).

According to McNeill (1992), co-speech gestures comprise a broad category that includes iconic gestures, metaphoric gestures, deictic gestures, and beat gestures. Whereas iconic gestures use space to mimic concrete objects or actions (e.g., using one's hand to form a spherical shape to represent a ball), metaphorical gestures use space to represent something abstract (e.g., fingers forming a heart shape

to represent the idea of affection). Experimental and classroom research in the last few decades has stressed the benefits of observing (and producing) both iconic and metaphoric gestures for word recall in a first language and word learning in a second language. Kelly et al. (2009) reported that observing congruent iconic gestures was especially useful for learning novel words in comparison to observing the same content presented only in speech, or in speech associated with incongruent iconic gestures. In a study involving 20 French children (average age 5.5) learning English, Tellier (2008) asked them to learn eight common words (house, swim, cry, snake, book, rabbit, scissors,and finger). Four of the items were associated with a picture while the other four items were illustrated by a gesture that the children saw in a video and then enacted themselves. The results showed that the enacted items were memorized better than items enriched visually by means of pictures. In a recent study, Macedonia and Klimesch (2014) looked at the use of iconic and metaphoric gestures in the language classroom in a within-subject longitudinal study lasting 14 months. They trained university students to learn 36 words (nine nouns, nine adjectives, nine verbs, and nine prepositions) in an artificial language corpus. For 18 items, participants only listened to the word and read it. For the other 18 items, par- ticipants were additionally instructed to perform the gestures proposed by the experi- menter. Memory performance was assessed through cued native-to-foreign translation tests at five time points. The results

showed that enacting iconic gestures significantly enhanced vocabulary learning in the long run. Goldin-Meadow, Nusbaum, Kelly, and Wagner (2001) suggested that "gesturing may prime a speaker's access to a temporarily inaccessible lexical item and thus facilitate the processing of speech" (p. 521)—an idea consistent with the Lexical Retrieval Hypothesis proposed by Krauss, Chen, Gottesmen, and McNeill (2000; see also Krauss, Chen, & Chawla, 1996, for a review).

However, gestures need not be semantically related to words to boost word learning and recall. Studies investigating beat gestures (rhythmic hand gestures that are associated with prosodic prominence) have demonstrated that watching these gestures also favors information recall in adults (Kushch & Prieto, 2016; So, Sim Chen-Hui, & Low Wei- Shan, 2012) and children (Austin & Sweller, 2014; Igualada, Esteve-Gibert, & Prieto, 2017), as well as second language novel word memorization (Kushch et al., 2018).

2.1.3 Producing vs. perceiving gestures

Under the approach of embodied cognition, cognitive processes are conditioned by perceptual and motor modalities (Borghi & Caruana, 2015). In other words, any knowledge relies on the reactivation of external states (perception) and internal states (proprioception, emotion, and introspection) as well as bodily actions (simulation of the sensorimotor experience with the object or event to which they refer). Much research on this domain,

especially in neuroscience, has shown brain activation of motor and perception networks when participants were engaged in different tasks involving abilities such as memory, knowledge, language, or thought (see Barsalou, 2008, for a review). By highlighting the importance of appropriate sensory and motor interactions during learning for the efficient development of human cognition, embodied cognition has crucial implications for education (see Kiefer & Trumpp, 2012; Wellsby & Pexman, 2014, for reviews). We believe that hand gestures can be investigated from this perspective.

In general terms, the production of hand gestures by learners has been found to be more effective than merely observing them for a variety of memory and cognitive tasks (Goldin-Meadow, 2014; Goldin-Meadow et al., 2009; for a review of the effects of enactment and gestures on memory recall, see Madan& Singhal, 2012). Goldin-Meadow et al. (2009) investigated how children extract meaning from their own hand movements and showed that children who were required to produce correct gestures during a math lesson learned more than children that produced partially correct gestures, who in turn learned more than children that did not produce any gestures at all. Furthermore, recent neurophysiological evidence seems to show that self-performing a gesture when learning verbal information leads to the formation of sensorimotor networks that represent and store the words in either native (Masumoto et al., 2006) or foreign languages (Macedonia,

M¨uller, & Friederici, 2011). However, mere observation of an action without production also seems to lead to the formation of motor memories in the primary motor cortex (Stefan et al., 2005), which is considered a likely physiological step in motor learning. Stefan et al. (2005) contend that the possible engagement of the same neural mechanisms involved in both observation and imitation might explain the results of behavioral experiments on embodied learning. For example, Cohen (1981) tested participants on their ability to recall actions following training under three conditions: They either performed the actions, observed the experimenter performing the same actions, or simply heard and read the instructions for these actions. He found that participants remembered actions better when they were performed either by themselves or by the instructor than when the actions were simply described verbally. Notwith- standing, Engelkamp, Zimmer, Mohr, and Sellen (1994) showed that self-performed tasks led to superior memory performance in recognition tasks for longer lists of items (24–48 items) but not for shorter lists (12 items).

2.1.4 Gestures and L2 pronunciation teaching

Little is known about the potential benefits of using co-speech gestures in the domain of L2 pronunciation learning, and specifically the potential differences between the effectiveness of observing versus producing gestures in the L2 classroom. A handful of studies have focused on the potential benefits of

observing co-speech gestures for pronunciation learning, with contradictory results. For example, Hirata and Kelly (2010) carried out an experiment in which English learners were exposed to videos of Japanese speakers who were producing a type of rhythmic metaphoric gesture to illustrate the Japanese short and long vowel phoneme contrasts, namely using a vertical chopping movement or a long horizontal sweeping movement, respectively. Their results showed that observing lip movements during training significantly helped learners to perceive difficult phonemic contrasts while the observation of hand movements did not add any benefit. The authors thus speculated that the mere observation of hand movement gestures might not have any impact on the learning of durational segmental contrasts. Hirata, Kelly, Huang, and Manansala (2014) explored specifically whether similar types of metaphoric gestures can play a role in the auditory learning of Japanese length contrasts. For this purpose, they carried out an experiment in which English speakers were trained to learn Japanese bisyllabic words by either observing or producing gestures that coincided with either a short syllable (one quick hand flick), a long syllable (a long horizontal sweeping movement), or a mora (two quick hand flicks). Basing themselves on a previous study (Hirata & Kelly, 2010), they hypothesized that producing beat gestures rather than merely observing them would enhance auditory learning of both syllables and moras. Although training in all four conditions yielded improved posttest discrimination scores,

producing gestures seemed to convey no particular advantage relative to merely observing gestures in the overall amount ofimprovement, regardless of whether the gesture accompanied a syllable or a mora. All in all, the results reported by this line of work have shown that hand gestures do not make a difference when learning phonological contrasts like length contrasts in Japanese (but lips do).

By contrast, positive results of hand gestures have been documented for learning suprasegmental functions, for example, highlighting prosodic prominence of words within a sentence. A recent study with a pretest/posttest design by Gluhareva and Prieto (2017) found positive effects of observing beat gestures placed on prosodically prominent segments on pronunciation results in general. Catalan learners of English were shown rhythmic beat gestures (simple up-and-down or back-and-forth motions of the hands) that highlighted the relevant prosodic prominence positions in speech during pronunciation training. The instructor replicated naturally occurring co-speech gestures as much as possible, placing beat gestures on words that carried the most semantic and prosodic weight. After training, the participants who had observed the training with beat gestures significantly improved their accentedness ratings on a set of difficult items.

2.1.5 Pitch gestures

In this study we will focus on the effects of observing versus producing another type of co-speech gesture sometimes used by second language instructors, namely pitch gestures, on the learning of lexical tones in a second language. Pitch gestures (a term coined by Morett & Chang, 2015) are a type of metaphoric visuospatial gesture in which upward and downward hand movements mimic melodic high-frequency and low-frequency pitch movements. How can pitch gestures, frequently used in CSL (Chinese as a Second Language) classrooms, promote the learning of lexical tones? Experimental evidence has shown that pitch and space have a shared audio-spatial representation in our perceptual system. The metaphoric representation of pitch was first investigated by Casasanto, Phillips, and Boroditsky (2003) in a nonlinguistic psychophysical paradigm. Native subjects viewed lines "growing" vertically or horizontally on a computer screen while listening to varying pitches. For stimuli of the same frequency, lines that grew higher were estimated to be higher in pitch. Along these lines, Connell, Cai, and Holler (2013) asked participants to judge whether a target note produced by a singer in a video was the same as or different from a preceding note. Some of the notes were presented with the corresponding downward or upward pitch gestures, while others were accompanied by contradictory spatial information, for example, a high pitch with a falling gesture. The results showed that pitch discrimination was significantly biased by the spatial

move- ments produced in gesture, such that downward gestures induced perceptions that were lower in pitch than they really were, and upward gestures induced perceptions of higher pitch. More recently, Dolscheid, Willems, Hagoort, and Casasanto (2014) explored the link between pitch and space in the brain by means of an fMRI experiment in which participants were asked to judge whether stimuli were of the same height or shape in three different blocks: visual, tactile, and auditory. The authors measured the amount of activity in various parts of subjects' brain regions as they completed the tasks and found significant brain activity in the primary visual cortex, suggesting an overlap between pitch height and visuospatial height processing in this modality-specific (visual) brain area.

We therefore surmise that the strong cognitive links between the perception of pitch and visuospatial gestures can have an important application in the learning of melody in a second language.

2.1.6 Pitch gestures and the learning of tonal words and intonation patterns

Relatively little experimental work has been conducted thus far on the potential ben- eficial effects of pitch gestures on the learning of L2 tones and words in a tonal language. CSL teachers report that pitch gestures are commonly used in the classroom and that there may be variability in the gesture space used to allow more or less ample pitch movements, and in the articulators used to perform the

pitch gesture, which can vary from the whole arm to a simple head movement. However, in all these gestures the spatial metaphor to describe pitch certainly remains the same.

Two longitudinal studies by Jia and Wang (2013a, 2013b) showed a positive effect of teachers' pitch gestures on the perception and production of tones by elementary-level learners of Mandarin. In a longitudinal study, Chen (2013) showed that 40 learners perceiving and producing "tonal gestures" (as he labeled them) seemed to have sig- nificantly superior communicative skills and performed significantly better in tonal production with a higher frequency of accurate responses, regardless of their tonal or non-tonal background. Moreover, the learners displayed a wider pitch range when producing Mandarin words together with gesture. Nonetheless, Chen's study was a classroom training study with no experimental control of (a) the materials used in the training session, (b) the perception and production activities during training, and (c) the participants' language background.

To our knowledge, four recent experimental studies have been carried out on the potential benefits of pitch gestures on the learning of L2 tones and/or intonation, with positive results. Three of these studies dealt with the effects of observing pitch gestures. Hannah, Wang, Jongman, and Sereno (2016) looked at how pitch gestures affect nonnative Mandarin tone perception by testing 25 English speakers who listened to two monosyllabic words with the

four tones under four conditions: audio-facial/congruent, audio-facial/incongruent, audio-facial-gestural/congruent, and audio-facial-gestural/ incongruent. After each pair of words, participants had to immediately indicate whether they had heard a level, “mid-dipping,”“rising,” or “falling” tone. The authors found that participants in the audio-facial-gestural/congruent condition obtained sig- nificantly better scores in tone identification than participants in the audio-facial/ congruent condition. In the second of these studies, Kelly, Bailey, and Hirata (2017) explored the effect of two types of metaphoric gestures on the perception of length and intonation features of Japanese phonemic contrasts by 57 English-speaking participants that had no previous knowledge of Japanese. They found that when visuospatial gesture depicting intonation were congruent with the auditory stimuli, accuracy was significantly higher than the control no gesture condition. Moreover, when the gesture was incongruent, accuracy was significantly lower than the control condition. The third study, by Yuan, González-Fuente, Baills, and Prieto (2018), tested whether pitch gesture observation would help the learning of difficult Spanish intonation patterns by 64 Chinese basic-level learners. Half of the participants received intonation training without gestures while the other half received the same training with pitch gestures representing nuclear intonation contours. Results showed that observing pitch gestures during the learning phase improved learner’s production outcomes significantly more than training

without gestures. By contrast, rather than focusing on observing, the fourth experimental study (Morett & Chang, 2015) tested the potential benefits of producing pitch gestures on the learning of L2 tones. In a between-subjects experimental design, 57 English speakers were divided into three groups and then trained to learn the meaning of 12 minimal pairs in Chinese. They had to repeat aloud the 12 Chinese words and imitate the gestures they saw performed by an instructor in a video in three conditions. One group of subjects saw and mimicked pitch gestures depicting the lexical tone pitch contours while hearing the Mandarin tones; the second group saw and mimicked gestures conveying word meanings (semantic gestures); and the third group were taught without gestures. Then participants were tested on a Mandarin lexical tone identification task and a word-meaning asso- ciation task. The results showed that, in comparison with semantic gestures and no gestures, producing pitch gestures facilitated the learning of Mandarin words differing in lexical tone, but failed to enhance their lexical tone identification. These findings suggested that the visuospatial features of pitch gestures strengthen the relationship between English speakers' representations of Mandarin lexical tones and word mean- ings. However, the null results found in the lexical tone identification task challenge the belief that the production of pitch gestures can enhance lexical tone acquisition. Furthermore, because all participants in the gesture groups had to both observe and produce pitch gestures or semantic gestures

(depending on the group) one cannot disentangle the potential effects of observing versus producing gestures. Thus, an open question that was not addressed by any of these four studies is whether it is observing or producing pitch gestures that has the stronger impact on L2 phonological acquisition.

2.1.7 Goals and hypotheses

The present study represents the first attempt to compare the effects of observing versus producing pitch gestures on the initial learning of tones and lexical items in Mandarin Chinese. First, we aim to enrich the debate on embodied cognition by exploring the respective roles of observing and producing gestures. Second, on a more practical level, we would like to determine the most advantageous pedagogical approach for the teaching of lexical tones to beginning learners of Mandarin Chinese. The study comprises two complementary between-subjects experiments. While Experiment 1 investigates the effects of observing pitch gestures on learning tones and words in Mandarin Chinese, Experiment 2 investigates the effects of producing such gestures. In both experiments, subjects without any previous knowledge of a tonal language were randomly assigned to the Gesture (experimental) condition or the Non-Gesture (control) condition. Both experiments included two parts, an audiovisual perceptual tone training session with minimal pair combinations of the four Mandarin Chinese tones, and an audiovisual vocabulary training

session focused on monosyllabic Mandarin Chinese words differing only in lexical tone. While after the tone-learning session, participants were asked to complete a lexical tones identification task, after the vocabulary training session they were asked to complete a word-meaning recall task and a word-meaning association task. First, based on previous findings, we predicted that observing pitch gestures would produce greater benefits for tone and word learning than not observing them, and second, given the literature on enactment and embodied learning, we predicted that the benefits of producing pitch gestures would be greater than the benefits of just observing them.

2.2 Experiment 1

The main goal of Experiment 1 was to assess the effect of pitch gesture observation on the learning of Chinese tones and words. The experiment consisted of a between-subjects training procedure with newly learned Chinese tones and words.

2.2.1 Participants

A total of 49 undergraduate and graduate students (age: M 5 19.86, SD 5 1.44; 15 males, 34 females) were recruited at the Communication Campus at the Universitat Pompeu Fabra in Barcelona, Spain. All participants were native speakers of Catalan and considered Catalan to be their dominant language relative to Spanish (mean percentage of Catalan in total daily language use 5 72%, SD 5 .664). All were right-handed and reported no previous knowledge of Mandarin Chinese or any other tonal language. All had normal or corrected-to-normal vision and normal hearing. Participants were assigned to either the control No Gesture (NG) group or the experimental Gesture Observe (GO) group. In the NG condition, the instructors in the training video remained still and the participant remained still and silent while viewing the video. In the GO condition, the instructors in the training video performed gestures while teaching the tones and the participant remained still and silent while viewing the video. The groups were comparable in terms of the number of participants (24 in the NG group, 25 in the GO group), age (M5 19.88 in the NG group, M5 19.68 in the GO

group), gender distribution (71% female, 29% male in the NG group and 68% female, 32% male in in the GO group), the amount of Catalan spoken in daily use (M 5 72.8% in the NG group, M5 71.2% in the GO group), and results on a memory span test (M 5 5,88 words in both groups). Participants were informed that the experiment consisted of an introductory tutorial on Mandarin Chinese tones and words and that they would learn how to pronounce the tones and some vocabulary. They were therefore unaware of the real purpose of the study. They signed a written consent form and received 10 euros for their participation.

2.2.2 Materials

The experiment consisted of three consecutive phases, first a tone familiarization phase containing introductory information on Mandarin tones, then two consecutive training sessions, one focusing on tones and the other on vocabulary items, and finally the corresponding test tasks. As will be explained in the following subsections, audiovisual stimuli were prepared for use in the two training sessions and auditory items were pre-recorded for the tone identification and word-meaning recall and word-meaning association tasks.

a) Audiovisual materials for the tone familiarization phase

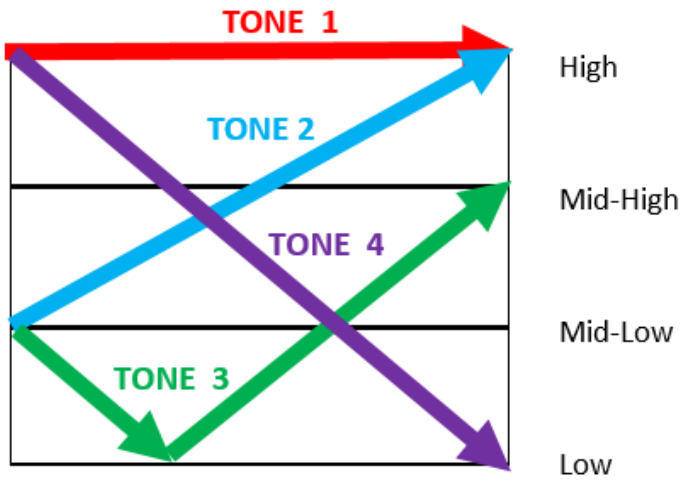
All the audiovisual materials for the three phases of the experiment were recorded by a male native speaker of Chinese and a female

bilingual Catalan-Chinese speaker. The video recordings were carried out at the experimental language research laboratory of the Universitat Pompeu Fabra's Department of Translation and Language Sciences using a PDM660 Marantz professional portable digital video recorder and a Rode NTG2 condenser microphone. The two instructors were recorded against a white background and the video clips for all the recordings showed the speaker's face and the upper half of their body so that participants could see all hand and face movements.

With narration in Catalan, the familiarization video first illustrated the four Mandarin tones both verbally and visually with the help of the 4-scale diagram shown in Figure 1 (adapted from Zhu, 2012). Mandarin Chinese distinguishes between four main lexical tones which are numbered according to their pitch contours: high flat-level (tone 1), rising (tone 2), low falling and rising (tone 3), and high-falling (tone 4) (Chao, 1968). For example, the syllable <ma> can have four different meanings according to the tone used: <ma>1 means mother, <ma>2 means hemp, <ma>3 means horse, and <ma>4 means scold. Two different videos were produced for the habituation phase, one for the GO condition, the other for the NG condition. Both lasted around 8 minutes. The monosyllabic words presented in the familiarization phase were all different from the words in the subsequent training phase, and they were accompanied in the video by subtitles showing their orthographic transcription (generally in pinyin) and tones.

Figure 1

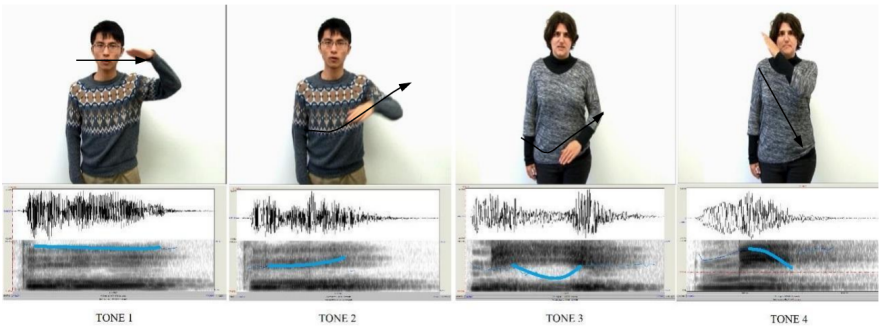
Diagram representing the four lexical tones in Mandarin Chinese



One instructor was a native Mandarin Chinese speaker and the other was an experienced CSL teacher for Catalan speakers. When performing the pitch gestures used in both the familiarization and training videos of the GO condition, the instructors used their right hand to gesture from left to right. They were also asked to produce the target words naturally while keeping their body and articulators like eyebrows, head, and neck totally still. Later the videos were digitally flipped to allow participants to observe the gestures from their left to their right. Figure 2 shows four stills from the videos illustrating the four target Mandarin tones (tones 1, 2, 3, 4) in the GO condition.

Figure 2

Screenshots illustrating the four target Mandarin tones in the Gesture Observe, with the corresponding sound waves and pitch tracks. The two left panels show the target syllable "puo" produced with tones 1 and 2 by the male speaker and the two rightmost panels show the target syllable "mi" produced with tones 3 and 4 by the female speaker.



Importantly, the two instructors were trained to use clear visuospatial hand gestures, making sure that the hand movements accurately mimicked the pitch variations and the natural duration corre- sponding to each lexical tone. To do this, we relied on the visual pitch line obtained in Praat (Boersma & Weenink, 2017) for each word in the stimulus recordings. For spatial consistency across renditions, the imaginary space for the hand movements was divided into four areas: the high tonal range corresponded to the face level, the mid-tonal range to the shoulder level, the mid-high frequency range to the chest level, and finally the mid- low

frequency range to the area of the hips. The duration of the tones, which can be a clue to determining what tone is being used, was left to the instincts of the instructors.

To guarantee that the speech characteristics in the NG and GO conditions would not differ, recordings of the same item in the two conditions were performed consecutively. Following González-Fuente, Escandell-Vidal, and Prieto (2015), mean pitch and duration cues were calculated for each speech file. Mean F0 was extracted from Praat for each item and computed in a Generalized LinearMixedModel (GLMM) test using IBM SPSS Statistics 23 (IBM Corporation, 2015) to determine whether there were significant differences in speech duration between the NG and GO conditions. PITCH was set as the fixed factor and SUBJECT and TONE were set as random factors. Results reveal no significant differences of mean pitch between the two conditions. MSD (mean syllable duration, in ms)was calculated by dividing the total duration of the target sentence by the number of syllables. A GLMM test was run with TIME set as the fixed factor and SUBJECT and TONE set as random factors. We found a significant difference of duration [$F(1, 66) = 5{,}134$, $p = .027$], with speech in the GO condition lasting significantly longer than speech in the NG condition ($M =$ 63 ms). Nevertheless, on the assumption that this difference was a consequence of the extra time required to produce the gesture in the GO condition, we decided not to modify the

recorded stimuli in any way to keep the stimuli as natural as possible.

b) Audiovisual materials for the tone training session

A total of 36 monosyllabic Mandarin words (18 minimal pairs differing only in tone) were chosen as stimuli for both tone training phases (see Table 1). Words were selected so that all minimal pair words shared the same phonological shape (except for tone) and the same grammatical category. There were a total of 5 pairs of verbs, 10 pairs of nouns, and 3 pairs of adjectives. All words conformed to the phonotactic restrictions of Catalan (Prieto, 2004) to avoid additional difficulty. The words were presented in orthographic form following the pinyin orthographic conventions, except when this would cause difficulty for Catalan speakers (the forms in brackets in Table 1 are the forms that participants were shown).

The 36 stimuli were recorded and presented in pairs to heighten contrast perception (Kelly, Hirata, Manansala, & Huang, 2014). In total, each of the four lexical tones was repeated nine times. The video recordings for the GO condition were produced with pitch gestures and the video recordings for the NG condition were produced without. After recording, the videos were edited using Adobe Premiere Pro CC 2015 software to produce six videos in which the six tonal contrasts (each composed of three pairs of stimuli) were put into sequences in different orders to avoid

primacy and recency effects (i.e., each video started and ended with different pairs of tonal stimuli).

Table 1

Pairs of stimuli for the tone training and vocabulary training sessions (18 pairs; 36 words)

Tonal contrast	Pinyin	English	Tonal contrast	Pinyin	English
1-2	bō bó [puo]	wave uncle	2-4	má mà	linen insult
	chī chí [txi]	eat pool		ná nà	take sodium
	fā fá	send raft		lí lì	pear chestnut
2-3	fú fǔ	fortune axe	1-4	tī tì [thi]	stairs shave
	bí bǐ [pi]	nose pen		pō pò [phuo]	slope spirit
	tá tǎ [tha]	battery tower		gē gè [ke]	song piece
1-3	tū tǔ[thu]	bald soil	3-4	mǐ mì	rice honey
	dī dǐ [ti]	taxi background		lǔ lù	prisoner deer
	chū chǔ [txu]	first storage		gǔ gù [ku]	drum hire

Note. When the orthographic form of the syllable presented to the participants differed from the pinyin orthography, the orthographic form is specified here within brackets.
In grey, the words selected for the vocabulary training.

c) Audiovisual materials for the vocabulary training session

A total of 12 targets were selected from the list of words in Table 1 (the minimal pairs selected appear in bold), which consisted of six minimal pairs of words differing only in their lexical tones. In each pair, Catalan translations of the two words were matched for mean log frequency per million words using NIM, an online corpus search tool that is useful for establishing word frequencies in Spanish, Catalan, or English (Guasch, Boada, Ferré, & Sánchez-Casas, 2013). The target minimal pairs were video-recorded in consecutive pairs following the same procedure described for the materials used for tone training. After the recordings, the pairs of stimuli were edited using Adobe Premiere Pro CC 2015 software. Items were repeated in randomized order within three blocks. In total, participants ended up seeing and hearing each vocabulary item (Catalan meaning 1 Chinese word) a total of three times. Six different videos containing the trials in different orders were created and distributed among the participants to avoid any primacy or recency effects.

d) Auditory materials for the test tasks (tone identification, word-meaning recall, and word-meaning association tasks)

For the tonal identification task, eight items (four pretrained: “mì,” “fǔ,” “txí,” “dī”; four new: “té,” “nù,” “lā",” “txě”) were chosen as real syllables or pseudo-syllables respecting Catalan phonotactic rules. Auditory materials were recorded by three native speakers of

Mandarin Chinese, two of them male and one female, who were not the instructors. The files were then uploaded on an online survey builder (https://www.surveygizmo.com) that automatically randomized the order of items.

For the vocabulary tests (word-meaning recall and word-meaning association tasks) the 12 items from the training session were used. The recordings featured a speaker of a different sex than in the training session to ensure that posttest performance reflected learners' ability to identify Mandarin lexical tones across word tokens rather than their recall of the specific token produced during the learning phase.

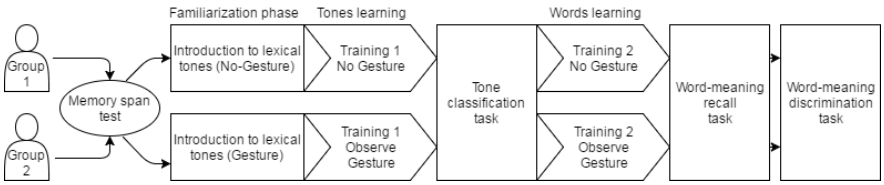
2.2.3 Procedure

Participants were tested individually in a quiet room. They were randomly assigned to one of the two between-subjects groups, 24 in the NG condition and 25 in the GO condition. Participants were asked to sit in front of a laptop computer equipped with earphones and mark their answer to the tone identification task on a sheet of paper next to the computer. First, a word memory span test (Bunting, Cowan, & Saults, 2006) adapted to the Catalan language was carried out to control for short-term working memory capacity. After completing the memory span task, participants in both conditions were instructed to remain silent and listen carefully to the audiovisual stimulus recordings as they played back on the computer. Participants in the experimental (GO) group were

additionally asked to pay attention to the gestures conveying the melodic movements. No feedback was provided at any point during the experimental tasks.

Figure 3

Experimental Procedure for Experiment 1



As mentioned previously, the experiment consisted of three phases (see Figure 3). In the familiarization phrase, participants were presented with a video consisting of a short introduction to the Chinese tones (8 min). After this, participants went on to view the tone training video (5 min), which was followed by a tone identification task (10–12 min). Finally, participants were shown the vocabulary training video (6 min), which was followed by two tasks, namely a word-meaning recall task and a word-meaning association task (15–20 min).

a) Tone training and tone identification task

After familiarization, participants were trained to discriminate between pairs of Mandarin Chinese lexical tones. The tone training video contained a total of 18 units which each consisted of pairs of

target tones (see Table 1). Within each unit, participants were exposed to the following sequence (see Figure 4): (a) the target pairs of tones to be discriminated; (b) the orthographic form of the pairs of Mandarin words together with their tone marks; and (c) the pair of video clips of these words as produced by the instructor.

Figure 4

Example of a trial sequence of the tone training video in the Gesture Observe condition involving tones 4 and 3 over the syllable "mi"

Immediately after viewing the training video, participants were asked to complete a tone identification task by listening to eight audio-only items. They were instructed to listen to the syllable and then write down what they had heard together with the correct tone mark. They could only listen to the syllable once. When they finished writing their answer, they had to go to the next screen to listen to the next syllable. The answers were afterward coded as 0 if the tone mark was incorrect or 1 if it was correct, regardless of the orthographic form of the word written by the participants.

b) Vocabulary training and word-meaning recall tasks

In the vocabulary training session, participants were asked to learn 12 words. The vocabulary training video contained a total of six units each containing minimal pair words, which were presented in three consecutive blocks with the stimuli in different orders. In total, they listened to the same stimuli three times. Within each unit, and for each of the word pairs, they were exposed to the following temporal sequence (see Figure 5): (a) the orthographic form of the Catalan word corresponding to the targetMandarinword to be learned; and (b) the video clip with the target word as produced by the instructor.

Figure 5

Example of a unit sequence during the vocabulary training session in the Gesture Observe condition with the minimal pair of Mandarin Chinese words bō "Cat. onada - Eng. wave'" and bò "Cat. oncle—Eng. uncle"

After they had viewed the training video, participants carried out a word-meaning recall task in which they were instructed to listen to the 12 target Mandarin Chinese words and translate each one of them into Catalan. They could only listen to each word once before writing down their answer and then going on to the next screen to

listen to the next word. Subsequently, they carried out a word-meaning association task involving the same 12 Mandarin words. Here they were shown the Catalan translations of the two words of a minimal pair but only heard one of the two and were asked to select the correct translation.

c) Statistical analysis

A total of 392 experimental responses were obtained (49 participants x 8 tone identification questions) for the tone identification task and a total of 588 responses were obtained (49 participants x 12 words) for each of the word-learning tasks. Statistical analysis of the results of the three tone and vocabulary tasks (e.g., the tone identification task, the word-meaning recall task, and the word-meaning association task) was carried out using IBM SPSS Statistics v. 24 (IBM Corporation, 2016) by means of three GLMM. Results of the memory span task revealed that all participants behaved within a normal range in short-term working memory capacity (M = 5.88 items remembered, SD = .712), and thus all of them were included in the analysis.

In each of the three models, ACCURACY of response was set as the dependent variable (two levels: Correct vs. Incorrect), which was modeled with a Binomial distribution and a Logit link. CONDITION (two levels: NG vs. GO) was set as a fixed factor. One random effects block was specified, in which we controlled for

subject intercept, with the type of tone defined as a random slope (covariance type: variance components).

2.2.4 Results

The GO group scored higher than the NG group in the three tasks (see Table 2) and for all four tones. Results of the three GLMM models revealed a significant main effect of CONDITION in the tone identification task, $F(1, 390) = 3.890$ ($\beta = .657$, $SE = .333$, $p = .049$, $\text{Exp}(\beta) = 1.929$), in the word-meaning recall task, $F(1, 586) = 4.789$ ($\beta = .683$, $SE = .312$, $p = .029$, $\text{Exp}(\beta) = 1.980$), and in the word-meaning association task, $F(1, 586) = 10.365$ ($\beta = 1.043$, $SE = .324$, $p = .001$, $\text{Exp}(\beta) = 2.834$), meaning that the GO experimental group significantly outperformed the NG control group in all three tasks. Calculating odd ratios (Exp(b), reported previously) is a reliable method to analyze effect sizes with logistic regressions. Odd ratios represent the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure (Szumilas, 2010). Odd ratios superior to 1 are associated with higher odds of outcome. In the three tasks, the GO condition received a much higher probability of obtaining more accurate values than the NG condition (specifically, compared to the NG control condition, the odds of obtaining correct answers is 1.929 higher in the GO condition in the tone identification task, 1.980 higher in the

word-meaning recall task, and 2.834 higher in the word-meaning association task).

Table 2

Means and standard deviations of accuracy (based on accuracy means per participant) for the three tasks in Experiment 1

	No Gesture		Gesture Observe	
	M	*SD*	*M*	*SD*
Tone identification	.70	.19	.80	.23
Word-meaning recall	.49	.22	.64	.25
Word-meaning association	.74	.19	.89	.12

In sum, results of the tone identification task show that observing pitch gestures significantly improved tonal perceptual learning in participants without any prior knowledge of Mandarin Chinese. Similarly, results from the two word-learning tasks demonstrate that a short vocabulary training session in which they observe pitch gestures may enhance L2 students' vocabulary learning in a tonal language like Chinese, at least at an initial stage of learning, and thus confirm the role of merely observing this specific type of gesture regarding the learning of words with tones.

2.3 Experiment 2

The main goal ofExperiment 2 was to assess the effect of pitch gesture production on the learning of Chinese tones and words. The experiment consisted of a between-subjects training procedure with newly learned Chinese tones and words.

2.3.1 Participants

Fifty-six undergraduate students (age *M* = 19.93 years, *SD* = 1.414; 9 males, 47 females) were recruited at the Universitat Pompeu Fabra in Barcelona, Spain. None of them had been subjects in Experiment 1. All were native speakers of Catalan and considered Catalan to be their dominant language relative to Spanish (mean percentage of Catalan in total daily language use = 68.4%, *SD* = .794). All of them reported no previous knowledge of Mandarin Chinese or any other tonal language.

In the control group (No Gesture Produce condition, henceforth NGP), the instructors in the training video performed gestures while teaching the tone words and the participant was instructed to repeat the tone words after the instructor but not to perform any hand movement. In the experimental group (Gesture Produce condition, henceforth GP), the instructors in the training video performed gestures while teaching the tone words and the participant was instructed to repeat the tone words after the instructor and at the same time mimic the gesture performed by the

instructors. The rationale for adding a control group where participants had to produce speech was that it required some form of active learning, which would be more accurately comparable to a condition where participants have to produce gestures. The groups were comparable in terms of the number of participants (28 in the NGP group, 28 in the GP group), age (M = 19.71 in the NGP group, M = 20.14 in the GP group), gender distribution (81% female, 19% male in the NGP group and 86% female, 19% male in the GP group), the amount of Catalan spoken in daily use (M = 67.8% in the NGP group, M = 68.6% in the GP group), and results on the memory span test (M = 5.54 words in the NGP group, M = 5.66 words in the GP group). They went through the same preliminary steps as in Experiment 1.

2.3.2 Materials

In Experiment 2, observing pitch gestures only was compared with observing and producing those gestures. Therefore, the video stimuli were the same for both conditions and identical to those used in the GO condition of Experiment 1 except that the instructions were different. Here, in both control group and experimental group, par- ticipants were instructed to repeat the Mandarin words they heard spoken by the instructor on the video. However, those in the GP condition were additionally instructed to mimic the pitch gestures illustrated by the instructor with their own right hand as they heard and repeated them (as in Kelly et al.,

2014). To allow them enough time to repeat the Mandarin word and produce the gesture, a 5-second black screen followed the modeling of each tone by the instructors in the video. In the two conditions, the training video was the same, but participants were asked to respond differently.

2.3.3 Procedure

As in Experiment 1, Experiment 2 consisted of three phases. In the initial familiarization phase, the experimenter initially informed the participants about the general procedure of the training session, after which they were presented with a short video introducing the Chinese tones(8min). Here, they were also familiarized with the pitch gestures by repeating two monosyllabic items for each tone, for a total of eight familiarization items. In the NGP condition, they were asked to repeat the word and pay attention to the gesture, while in the GP condition, they were asked to repeat the word and mimic the pitch gesture. There was no feedback on the pronunciation of the tones; however, at this stage, the experimenter could offer some feedback on the production of gesture if needed. Next, they viewed a tone training video (8 min), which was followed by a tone identification task (10–12 min). They then watched a vocabulary training video (9 min), which was followed by two tasks, a word- meaning recall task and a word-meaning association task (15–20 min). In the NGP condition, for each minimal pair they were first presented with the two Chinese

syllables in written form, and then heard the instructor produce the target syllable with both tones and the corresponding gestures. When the screen subsequently went black they had to repeat the syllable aloud only. Participants in the GP condition watched the same video as in the NGP condition and repeated the target syllables; additionally, however, they were asked to copy and perform the pitch gesture.

Accuracy of speech during the training was not measured and no feedback was provided during the training. However, the experimenter was present in the room and could thus make sure that the participants were performing the gestures/speech appropriately depending on the condition.

Statistical analysis

A total of 416 responses were obtained (26 participants x 2 conditions x 8 tone identification questions) for the tone identification task and a total of 624 responses were obtained (26 participants x 2 conditions x 12 words) for each of the word-learning tasks. Statistical analysis of those results (tone identification task, word-meaning recall task, and word-meaning association task) was carried out using IBM SPSS Statistics v. 24 (IBM Corporation, 2016) by means of three GLMMs. Results of the memory span tasks revealed that all subjects behaved within a normal range in short-term working memory capacity (M = 5.88

items remembered, *SD* = .712), and thus the experimental data from all of them were included in the analysis.

In each of the three models, ACCURACY of response was set as the dependent variable (two levels: Correct vs. Incorrect), which was modeled with a Binomial distribution and a Logit link. CONDITION (two levels: NGP vs. GP) was set as a fixed factor. One random effects block was specified, in which we controlled for subject intercept, with the type of tone defined as a random slope (covariance type: variance components).

2.3.4 Results

The GP group scored higher than the NGP group in the three tasks (see Table 3). The results of the three GLMM models revealed a significant main effect of CONDITION in the three models, namely in the tone identification task, $F(1, 446) = 4.550$ ($\beta = .769$, $SE = .331$, $p = .033$, $\text{Exp}(\beta) = 2.158$), in the word-meaning recall task, $F(1, 670) = 7.360$ ($\beta = .827$, $SE = .305$, $p = .007$, $\text{Exp}(\beta) = 2.287$), and in the word-meaning association task, $F(1, 670) = 4.237$ ($\beta = .535$, $SE = .260$, $p = .040$, $\text{Exp}(\beta) = 1.707$), indicating that the GP experimental group outperformed the NGP control group in the learning of both tones and words. In the three tasks, the GP condition received a much higher probability of obtaining more accurate values than the NGG condition (spe- cifically, compared to the NGG control condition, the odds of obtaining correct answers is 2.158 higher in the GP condition in the tone identification task,

2.287 higher in the word-meaning recall task, and 1.707 higher in the word-meaning association task).

Table 3

Means and standard deviations of accuracy (based on accuracy means per participant) for the three tasks of Experiment 2

	No Gesture Produce		Gesture Produce	
	M	*SD*	*M*	*SD*
Tone identification	.59	.21	.72	.25
Word-meaning recall	.40	.21	.57	.22
Word-meaning association	.74	.17	.82	.14

All in all, the results revealed that the group of participants who produced the pitch gestures performed significantly better in all three tasks, namely the tone-learning task and both word-learning tasks. Note that our results partially contrast with those obtained by Morett and Chang (2015), who did not find that producing pitch gestures significantly helped lexical tone identification compared to other types of gesture. However, results from the vocabulary tasks support Morett and Chang's (2015) results on the role of pitch gestures in vocabulary learning.

Comparing the effects of pitch gesture observation and pitch gesture production to further compare perception and production of gestures, we statistically compared the effects of passively observing pitch gestures with the effects of a more "enacted" training condition, that is, observing pitch gestures and additionally mimicking them while repeating the tonal words. Because the training procedures and tone perception and vocabulary tests were the same in every other respect across both experiments, we set out to perform a direct comparison between the GO condition from Experiment 1 and the GP condition from Experiment 2.

As before, we ran three GLMMs, one for each dependent variable, that is, the proportion of correct responses in the tone identification task, the word-meaning recall task, and the word-meaning association task. In each of the three models, ACCURACY of response was set as the dependent variable (two levels: Correct vs. Incorrect), which was modeled with a Binomial distribution and a Logit link. CONDITION (two levels: GO vs. GP) was set as a fixed factor. One random effects block was specified, in which we controlled for subject intercept, with the type of tone defined as a random slope (covariance type: covariance components). Results of the GLMM did not reveal any significant main effect of CONDITION in any of the tasks.

Given these results, it is necessary to explore why the benefit of producing pitch gestures seen in Experiment 2 is no longer visible

when data from the two experiments are compared. The main difference between the experimental conditions of Experiment 1 (GO) and the control condition of Experiment 2 (NGP) being the production of speech, we compared the scores in these conditions and found that the NGP group had sig- nificantly lower scores than the GO group in the tone identification task, $F(1, 422) = 14.724$ (β = - 1.236, SE = .322, p = .000, Exp(β) = 0.290), in the word-meaning recall task, $F(1, 634) = 10.604$ (β = -1.132, SE = .348, p = .001, Exp(β) = 0.322), and in the word-meaning association task, $F(1, 634) = 12.198$ (β = -1.035, SE = .296, p = .001, Exp(β) = 0.355). Therefore, it seems that repeating the tonal words while watching the gesture during the training had a negative outcome on scores in all the tasks.

2.4 Discussion and conclusion

The present study has added more evidence in favor of the use of pitch gestures to learn tones in a second language and, crucially, has assessed the potential differences between gesture perception and production in facilitating tone and word learning. The study comprised two experiments that examined whether the learning of Mandarin lexical tones and words would be enhanced by: (a) a short training session where participants merely observe pitch gestures (Experiment 1) or (b) a short training session where participants mimic pitch gestures (Experiment 2). The results demonstrated that both the observation and the production of pitch gestures showed a beneficial effect in subsequent tone-learning and word-learning test tasks in comparison with the control non gesture condition. Specifically, while the results of Experiment 1 demonstrated that a short training session involving observing pitch gestures enhanced the acquisition ofMandarin Chinese tones and words more than a comparable short training session without gestures, the results of Experiment 2 showed that a short training session in which subjects produced pitch gestures while repeating the words enhanced the acquisition of Mandarin Chinese tones and words more than just observing the gestures and repeating the words.

The results of our study add more evidence in favor of the benefits of pitch gestures for learning L2 tones and intonation (Hannah et

al., 2016; Kelly et al., 2017; Morett & Chang 2015; Yuan et al., 2018). Specifically, our results partially replicate and extend the findings by Morett and Chang (2015). Their experimental results showed that while the production of pitch gestures by participants facilitated the learning of words differing in lexical tones in Mandarin Chinese, they failed to enhance lexical tone identification performance. By contrast, our results showed an amplified effect of pitch gestures in that not only producing but also just observing pitch gestures triggers an enhancement of both tone identification and word-learning scores. These experimental results support the findings from Chen's (2013) longitudinal classroom study, where students who saw and used gestures were more accurate in answering their instructors' tonal queries than students taught with the traditional 5-scale tone chart (Chao, 1968), and the findings seen in Jia and Wang (2013a, 2013b).

In more general terms, these results add more evidence about the importance of using different types of supporting gestures for L2 instructional practices. As we have seen before, semantically related iconic gestures have also been found to enhance novel word acquisition (Kelly et al., 2009; Macedonia et al., 2011; Tellier, 2008; Thompson, 1995). However, pitch gestures do not convey semantic information per se. So why is it that they produce these beneficial effects? We believe that the metaphorical visuospatial properties of pitch gestures are visually encoding one of the essential phonological features of words in a tonal language,

namely their lexical tone. It is presumably the enrichment of these phonological properties through visual means that provides a positive supporting channel for the acquisition of novel words in tonal languages. Moreover, the benefits of pitch gestures for tone identification provide further evidence for theories claiming that pitch perception is fundamentally audio-spatial in nature (e.g., Cassidy, 1993; Connell et al., 2013; Dolscheid et al., 2014) as well as supporting the spatial conceptual metaphor of pitch (Casasanto et al., 2003).

In contrast with the positive results obtained in various studies on the role of pitch gestures on the acquisition of second language tones or intonation (Hannah et al., 2016; Kelly et al., 2017; Morett & Chang, 2015; Yuan et al., 2018), there is to date no clear view on how other types of metaphoric (and beat) gestures affect phonological learning. In contrast with the positive effects of pitch gestures for learning L2 tones and intonation, the results of studies targeting the effectiveness of what are called "length gestures" to learn duration contrasts in a second language are not so clear. In various studies, Kelly, Hirata, and colleagues (Hirata & Kelly, 2010; Hirata et al., 2014; Kelly & Lee, 2012; Kelly et al., 2017) have explored the role of two types of gestures that metaphorically map the duration of a vowel sound in Japanese duration contrasts without thus far detecting any positive effects. For example, Hirata and Kelly (2010) investigated the role of co-speech gesture perception in the auditory learning of Japanese vowel length

contrasts. In the study, participants were exposed to videos of Japanese speakers producing Japanese short and long vowels with and without hand gestures that were associated with vowel length. A short vertical chopping movement was used to mark short vowels and a long horizontal sweeping movement was used to mark long vowels. The results of the experiment showed that there was no noticeable benefit for participants when they learned vowel length by viewing videos showing length gestures as opposed to viewing videos that did not show such gestures. More recently, Kelly et al. (2017) suggested that it may be possible to safely narrow down the effective use of perhaps the utility of visuospatial gestures in pronunciation learning is limited to the use of pitch gestures for the learning of intonation patterns (but not thus excluding the use of various types of metaphoric gestures for the study of duration). There might be a set of several possible reasons that can explain the discrepancy between the results of the previously mentioned studies. First, as Kelly et al. (2017) noted, pitch gestures tend to have a stronger effect on learning L2 pitch differences than length/duration gestures on learning durational differences. Indeed, Kelly et al. (2017) explored the potential differences in the effect of length and pitch gestures on learning length and pitch phonological contrasts, respectively. In this study, English-speaking adult participants were exposed to videos with a trainer producing Japanese length contrasts and sentence-final intonation distinctions accompanied by congruent metaphoric

gestures, incongruent gestures, or no gestures. The results showed that for intonation contrasts, congruent metaphoric gestures (i.e., pitch gestures) had a positive effect, as identification was more accurate in comparison to other conditions. For the length contrast identification, however, similar results were not obtained, and no clear and consistent pattern emerged. In fact, the use of congruent metaphoric gestures seemed to make length contrast identification more difficult.

We would like to suggest that the type of metaphorical gestures used by Kelly, Hirata, and colleagues (Hirata & Kelly, 2010; Hirata et al., 2014; Kelly & Lee, 2012; Kelly et al., 2017) may have had an influence too. Specifically, the mora gestures used in the studies of Kelly, Hirata, and colleagues (e.g., the short vertical chopping movements) might have come across as "non-intuitive" to English speakers and thus did facilitate (or even hindered) their learning of durational information in the second language (see also the comments on the lack of effectiveness of length gestures in Kelly et al., 2017). The fact that other studies like Gluhareva and Prieto (2017) have found that observing other types of rhythmic gestures (e.g., beat gestures) has a positive effect on general pronunciation results leads us to suspect that perhaps the pitch gestures must seem natural to have positive results.

Another goal of the present study was to compare the effects of observing versus producing pitch gestures on learning Chinese

tones and words. Results from a variety of studies have suggested that the production of gestures by the learners is more effective than observing them alone in various learning contexts (Goldin-Meadow, 2014; Goldin-Meadow et al., 2009; Macedonia et al., 2011; Masumoto et al., 2006; Saltz & Donnenwerth-Nolan, 1981). Regarding pitch gestures specifically, Morett and Chang (2015) did explore their effect, but all the participants in their study had to perform pitch gestures, and thus the study could not disentangle the potential effects of observing versus producing gestures. In our data, a comparison of results from the GO group in Experiment 1 and the GP group from Experiment 2 revealed that training with mere observation and training with production of both speech and gesture had equally beneficial effects in both tone- and word-learning tasks. One explanation for this effect can be the specificity of practice effect explored by Li and De Keyser (2017). Their study provides strong evidence that tone-word perception and production skills each depend on the practice used to develop them. In our experiments, the tasks used to evaluate participants' acquisition of tones in Mandarin after training exclusively targeted perception, which may explain why the results obtained from the GO group were as good as those obtained from the GP group, and why the results from the NGP group were so low.

Another explanation could be related to the effects of using gesture on the speaker's cognitive load. Whereas some studies have suggested that gestures help reduce the cognitive load or processing

cost by conveying the same message through an additional modality (Goldin-Meadow, 2011; Wagner, Nusbaum, & Goldin-Meadow, 2004) and thus function as a compensatory and facilitating device in the acquisition of a second language (Gullberg, 1998; McCafferty, 2002), other studies have found that when learning higher aspects of a L2 such as semantics, syntax, or phonetics, observing (Kelly & Lee, 2012) and producing gestures (Kelly et al., 2014; Post, Van Gog, Paas, & Zwaan, 2013) only helps when cognitive demands are not too high, otherwise becoming counterproductive and/or distracting.

In our study, participants in the NGP group might have experienced such a cognitive overload. It seems reasonable to think that for participants with no previous knowledge of Chinese, having to learn new words while having to repeat them and at the same time not mimic the target pitch gestures might be a demanding task. This may be borne out by the fact that the mean accuracy for the GO group (Table 2, Experiment 1) was much higher than the NGP group (Table 3, Experiment 2). In other words, repeating the words while seeing the words produced with pitch gesture was altogether the less effective strategy to learn both tones and words. These results may be interpreted as the consequence of a disconnection between the perceptive modality (seeing the gesture) and the productive modality (repeating speech). Because gesture and speech are highly integrated and interdependent, it is possible that this disconnection produced cognitive overload.

In general, the evidence reported in this article adds to the growing body of evidence in favor of using gestures in vocabulary and pronunciation learning, thus reinforcing the embodied cognition paradigm. This paradigm theorizes that the human perceptual and motor system play an important role in cognition and underlines the importance of body movements and multimodal supporting channels in cognition and in favoring memory traces (see Barsalou, 2008; Barsalou et al., 2003; Paivio, 1990). According to the dual coding theory (Paivio, 1990), learning is reinforced when the visual modality is added to the verbal modality. Dual coding theory supports the idea that multimodal memory traces are richer and stronger than unimodal traces that result from either the visual or verbal modality alone. Empirical evidence that mere observation of an action, like in our GO group, leads to the formation of motor memories in the primary motor cortex supports the predictions made by these theories (Stefan et al., 2005), in the sense that the addition of visual information to verbal information should create stronger memory traces.

Limitations and future directions of research

Several limitations of this study can be identified. First, in Experiment 1, the slight increase in duration found in the auditory signal of the training items corresponding to the GO condition (mean of 1 63 ms) may play a role to some extent in the positive results favoring the tones' acquisition in the gesture observation

condition. Therefore, further research could try to assess the mechanisms behind the effects of gesture, that is, whether the gesture alone could obtain an effect, or whether it is both the auditory and the visual properties (e.g., the auditory signal that is naturally modified by the production of the gesture) that are responsible for the effect.

Though our results confirm that pitch gestures can be useful for learning Chinese tones at a basic level (our participants were completely new to Mandarin Chinese), our study cannot tell whether pitch gestures will have such strong effects with more proficient learners. It would be very interesting to test the effectiveness of pitch gestures using more complex phrasal contexts such as two-syllable words and with participants that have some prior knowledge of Mandarin Chinese.

Another limitation of the study lies in the lack of a productive task in the posttests. Indeed, it would have been helpful to verify the specificity of practice effect suggested by Li and DeKeyser (2017) by exploring whether participants in the GP condition showed any advantage in productive tasks. Finally, it would be interesting to assess more precisely the respective roles of perceiving versus producing pitch gestures and determine how to use this information best to achieve particular pedagogical goals.

These limitations notwithstanding, our study shows that, at least for initial levels of L2 tone learning, observing or producing pitch

gestures can be equally effective to help students perceive the tones of the target language and learn new tonal words. From a pedagogical perspective, our findings support the use of teaching and learning methods that implement more active audio-visual and embodied cognition strategies in the second language classroom. On this basis, for example, teachers of CSL could use pitch gestures while teaching the tones for the first time or, when teaching a new word, asking learners to pay attention to the gesture while listening to the word, therefore enhancing discrimination abilities and memorization. Once learners have gained some knowledge of Chinese tones and tonal words and have observed the teacher performing pitch gestures, the teacher could ask them to repeat the words accompanied with the pitch gesture to practice oral skills. Though more applied research is clearly needed, these results constitute an incentive to start implementing more effective multimodal approaches in the CSL classroom.

3

CHAPTER 3: EMBODIED PROSODIC TRAINING HELPS IMPROVE L2 PRONUNCIATION IN AN ORAL READING TASK

Baills, F., Alazard-Guiu, C., & Prieto, P. (under review). Embodied prosodic training helps improve L2 pronunciation in an oral reading task. *Applied Linguistics.*

3.1.Introduction

There is increasing evidence of the integration of the perceptual and motor systems in the cognitive system (e.g. Barsalou 2008; Wilson & Foglia 2017; Keily 2019) and of the benefits of embodied learning in education, notably through the use of hand gestures (e.g. Macedonia 2019; Shapiro & Stolz 2019). In the field of foreign language acquisition, however, since Atkinson's call for an embodied approach to SLA (2010), relatively little work has been carried out to put this claim into perspective. Regarding phonological learning, despite numerous studies confirming the positive role of pronunciation instruction - in particular, prosodic training (e.g. Saito 2012; Gordon & Darcy 2016; Zhang & Yuan 2020) - and the important role of prosody in pronunciation evaluations (e.g. Kang 2010; Trofimovich & Baker 2006), there is a clear need for concrete, research-based, pronunciation teaching techniques that focus on highlighting L2 prosody. Based on previous evidence that prosodic features can be successfully depicted by hand movements (e.g. Connell et al. 2013; Dolscheid et al. 2014, Biau 2015), the present study explored the gains of training prosody using embodied techniques on L2 pronunciation.

3.1.1 Embodied Cognition Theory, Embodied Learning, Language and Prosody

According to Embodied Cognition Theory, sensory-motor processes and the physical body are an integral part of human cognition and modulate cognitive processing (e.g. Barsalou 2008; Wilson & Foglia 2017; Keily 2019). This theory is based on the mutual effects of perception and actions on one another and their joint effect on mental representation and is claimed to have special relevance for education (e.g. Kiefer & Trumpp 2012; Ionescu & Vasc 2014; Macedonia 2019; Shapiro & Stolz 2019).

The discovery of mirror neurons, a group of motor neurons that are activated upon watching another person perform a behavior, has led scholars to propose that these neurons may play a crucial role in understanding other peoples' actions and may be necessary for imitative learning (Rizzolatti & Craighero, 2004). They stand as a potential explanation for the positive effects of active engagement and communicative gestures. Sullivan (2018) argued that instructors' movements and their use of representational gesture stimulate mental imitation by activating the mirror neurons, which may lead to an improvement in students' academic outcomes. Meanwhile, more empirical research about embodied cognition and learning has primarily focused on how increasing students' own motor involvement during instruction increases learning outcomes (e.g. Bahnmueller et al. 2014; Smith et al. 2014).

There is evidence that language, in particular, is embodied. Research has shown that the language areas in the brain activate during sensorimotor action (e.g. Desai et al. 2010) and conversely, motor areas activate during speech (e.g. Hauk et al. 2004), including when processing non-literal action language (e.g. Yang & Shu 2016). Gestures, which are closely tied to speech (e.g. McNeill 1992) and develop together in infancy (e.g. Iverson & Goldin-Meadow 2005), may stem from spatial representations and mental images and may arise from an embodied cognitive system, as proposed by Hostetter and Alibali's Gestures as Simulated Action framework (2008). Several studies lend evidence to the theory. For example, Rieser et al. (1994) found that linguistic tasks related to spatial orientation are facilitated by the mental representation of movement both in children and adults. Descriptions of spatial associations are comprehended faster than those of spatial dissociations (Glenberg et al. 1987) and words with high 'body-object interaction' ratings (Saikaluk et al. 2008) or related to manipulable objects (Rueschemeyer et al. 2010) are recognized faster, providing further evidence of the role of motor actions on lexical-semantic processing. Moreover, there is ample evidence of the effect of actions, gestures, and exercise on memory (see Madan & Singhal 2012 for a review). For example, lessons with gestures are shown to promote deeper reasoning, synthesis, and information retention than lessons that do not feature gestures (Goldin-Meadow & Alibali, 2013). Interestingly, some work has

also focused on the important role of prosody in syntax and syntax learning through the lens of embodied interaction (e.g. Bergmann et al. 2012; Kreiner & Eviatar 2014; Matsumoto & Dobs 2016).

3.1.2 Embodied learning in SLA

Research in the field of Conversation Analysis has documented how cognitive states are expressed during interaction, not only through speech but also via gaze, facial gesture, hand gesture, posture shift, and the manipulation of documents and objects and how these embodied cognitive states participate in the management of peer interaction (e.g. Goodwin & Goodwin 1986; Drew 2006; Cekaite 2015; Eskildsen & Wagner 2013, 2015; Jakonen 2020; Majlesi 2015; Mori & Hasegawa 2009; Kääntä. 2015). To give a few examples, Mori and Hasegawa (2009) showed how two students organized themselves in a word search activity by simultaneously using different semiotic resources, such as language, body, and the structures of their textbooks and notebooks for language learning. Jakonen (2020) suggested that teachers use their body as a pedagogical device and analyzed teachers' movement trajectories and body positioning in content and language integrated learning (CLIL) classrooms. The analysis showed that walking through the classroom allowed the teacher to monitor student individual and group progress during a task, to display availability, and to invite students' interaction. Eskildsen and Wagner (2015) analyzed how gesture-speech combinations are

created by L2 learners to create a common understanding of new words and how they are reused on later occasions. Interestingly, Eskildsen and Wagner (2013) observed that the imitation of a speaker's gesture acts as a communicative resource for achieving and maintaining understanding in spontaneous conversations between pairs and with the teacher. Studies in the field of gesture are further exploring, describing, and classifying teachers' and learners' gestures as part of their linguistic conceptualization and expression (e.g. Gullberg & McCafferty 2008; Smotrova 2014; Wang & Loewen 2016). However, very few studies have been conducted to test empirically the effects of embodied learning strategies on second language acquisition. Rather, most of these studies have looked at the effect of spontaneous and nonspontaneous gestures on word recall (see Macedonia 2014; Morett 2018 for reviews; for the effect of gestures on grammar learning, see Nakatsukasa 2016).

Regarding phonological learning, strong evidence for a tight relationship between prosody and gesture (e.g. Loehr 2012; Biau 2015; Ferré 2018) suggests a positive role of embodied strategies on the learning of an L2 phonological system, especially on pronunciation. Chan (2018) advocates the integration of body movements and gestures to enhance the perception, pronunciation, and retention of L2 phonological features. There are several reasons to support her claim. First, research on embodied approaches to music education has shown that body movement can

enhance the acquisition of musical rhythmic and melodic patterns (e.g. Juntunen 2016). In view of the close resemblance between musical and prosodic structure (e.g. Heffner & Slevc 2015), we surmise that, in a similar way, hand and arm movements may help the acquisition of speech rhythm and melody. In addition, from the field of sign language, there is evidence of the existence of a visuospatial 'phonological loop' in working memory, similar to the phonological loop for speech, which is structured uniquely by language (e.g. Wilson & Emmorey 1997). In that sense, the form of a gesture may be processed in a similar way to speech sounds and associated with the corresponding phonological feature. Finally, there is evidence that the mental representation of pitch is visuospatial in nature (e.g. Connell et al. 2013; Dolscheid et al. 2014), indicating that making pitch directions and movements visible to the learners may help them process foreign language prosody. In the following section, we review the literature on the benefits of prosodic pronunciation instruction, with a focus on embodied techniques.

3.1.3 Benefits of prosodic pronunciation instruction

Prosody plays an important role in pronunciation. There is evidence that transfer from a first to a second language takes place in the prosodic domain (e.g. Ueyama 2000; Trofimovich & Baker 2006; Lomotey 2013), as well as evidence that suprasegmental patterns play a crucial role in the perception of non-native pronunciation

patterns (e.g. Kang et al. 2010; see Wang 2020, for a review) and seem to weigh more in the perception of foreign accentedness (e.g. Anderson-Hsieh et al. 1992; de Mareüil & Vieru-Dimulescu 2006; Trofimovich & Baker 2006).

A growing body of evidence has shown that pronunciation instruction focusing on speaking rate, intonation, rhythm, and word and sentence stress may improve overall measures of pronunciation more than segmental training or no training at all in sentence repetition, read-speech, and spontaneous speech tasks (e.g. Gordon et al. 2013; Saito & Saito 2017; Zhang & Yuan 2020). In a meta-analytic review, Thomson and Derwing (2015) found that 52 percent of the studies on pronunciation instruction included in their analysis investigated segmental training 18 percent focused on suprasegmental training, and 30 percent dealt with both, usually in combined lessons but occasionally as separate comparison groups. Unfortunately, these studies including long suprasegmental instruction paradigms used a varied set of techniques that ranged from explicit instruction involving theoretical presentation-practice-production sequences (e.g. Gordon et al. 2013) to more implicit techniques involving musical and rhythmic activities (e.g. Derwing et al. 1998), making a full synthesis of the results difficult. Moreover, as pointed out by the authors and also by Lee et al. (2015), most of the studies failed to provide a sufficiently thorough description of the training activities involved.

To our knowledge, only a small set of implicit prosodic training techniques involving music- and prosodic-based activities (some with visual feedback) have been empirically tested to assess their value for second language pronunciation improvement. Some studies have found that musical activities highlighting the rhythmic and melodic properties of language are helpful in improving L2 pronunciation. In a 12-week instruction study with a pre- and posttest design, Derwing et al. (1998) used song materials to train learners to count the number of syllables and stresses, tap out the beats, and use nonsense syllables to focus on rhythm. The authors found this type of training more beneficial than segmental training on the comprehensibility and fluency of spontaneous speech in a narrative task. More recently, Good et al. (2015) found a positive effect of teaching a short passage in a sung modality compared to spoken modality on the pronunciation of L2 vowels sounds. Ludke (2018) compared L2 instruction with singing and song listening activities to L2 instruction with visual arts (drawing and creating cartoons) and drama activities and found higher performance on intonation and flow of speech in the singing and song group. In a different approach, computer-assisted learning based on the development of speech analysis technology can also be used to teach L2 suprasegmental features by allowing learners to compare the visual representation of target pitch contours produced by native speakers to their own output and try to adjust it accordingly (e.g. de Bot 1983; Hardison 2004; Ramirez Verdugo 2006; Hincks

& Edlund 2009; Tanner & Landon 2009; Liu & Tseng 2019) or by juxtaposing a computerized set of percussive sounds over target sentences to provide additional rhythmic cueing (Wang et al. 2016).

3.1.4 Embodied prosodic instruction in the classroom

In practice, it is not uncommon to see teachers spontaneously use co-speech gestures when explaining difficult pronunciation features. For example, based on the observation of audiovisual corpus, Tellier (2008) gave a description of the pedagogical gestures employed to teach pronunciation, in particular gestures that enable the students to visualize and feel the prosodic characteristics of speech. She mentioned that language teachers use flat, rising, and falling hand movements to imitate sentence intonation (see also Smotrova 2014). Hudson (2011) described horizontal movements of the hands or lateral movements of the body to represent vowel duration. Finally, beat gestures, tapping, or clapping rhythms function as a way to distinguish syllables or to indicate stress position (Chan 2018; Baker 2014; Hudson 2011).

The essential haptic-integrated English pronunciation (EHIEP) framework developed by Acton and colleagues (Acton et al., 2013) proposes coupling speech with systematic hand movement, kinesthetic and tactile techniques to teach pronunciation. Acton's 'essential haptic-integrated English pronunciation' blog proposes a variety of embodied techniques for teaching segmental (vowels and

consonants) and suprasegmental (stress, rhythm, intonation) features of the English sound system. An example of such haptic techniques for prosody would be tapping on one's own shoulder and arm with different intensity when uttering stressed and unstressed syllables (The Butterfly technique, Burri & Baker 2016). Burri et al. (2016) also proposed an activity called the rhythmic fight club to teach vocabulary alongside syllable and word-stress awareness. This technique consists of performing boxing-like movements to physically experience rhythm and syllable stress. Burri et al. (2019) argued in favor of practicing intonation and rhythm haptic techniques on commonly-used chunks of language to enhance learners' spontaneous speech, although the authors did not provide any empirical evidence in this respect. Nevertheless, an evaluation of EHIEP techniques by language teachers after a 16-week practice revealed overall positive perceptions of haptic pronunciation teaching (Burri & Baker 2019). In a recent five-day intervention study, Mister et al. (2021) taught new vocabulary to 16 learners of English by focusing on word stress during both controlled and more spontaneous productive activities and by using kinaesthetic/tactile teaching techniques. Results indicate that learners increasingly improved the recall and the correct stress placement of the target words over the course of the intervention, but without contrasting these benefits to any control group.

Another approach to pronunciation teaching is known as the verbotonal method (henceforth VT, e.g. Guberina 2008, Renard

2002), which is based on the notion that prosody acts as a frame for pronunciation development and should be taught from the first stages of language learning. This is achieved notably through the repetition of logatomes combined with visuospatial hand gestures that mimic the intonation and rhythm of the sentence (Billières 2002). A logatome is a series of same consonant-vowel nonsense sequences (e.g. /dadada/) that remove any target segmental information but keep the prosodic structure of the sentence intact. Repeating meaningless CV syllables in this fashion allows learners to focus on the suprasegmental features of target utterances while keeping the segmental content controlled (see Billières 2002 for a full explanation of the use of logatomes in the VT method). In addition, the role of the body as a supporting tool is fundamental to this approach, as stated by Guberina (1965, p.151):

> "L'ensemble acoustique de toutes les langues contient certains facteurs structuraux qui sont immanents à notre être biologique. La tension, l'intensité, le rythme les tonalités sont des formes biologiques de l'homme." [The acoustic ensemble of all languages contains certain structural factors that stem from our biological nature. Tension, intensity, rhythm and tonality are all products of human biology]

Billières (2002) further describes the benefits of accompanying logatomes with hand gestures—what we will henceforth refer to as

embodied logatomes—to mimic the intonation, rhythm and stress patterns of the target sentence for learning purposes. The repetition of embodied logatomes is generally performed before the repetition of full target sentences, as the repetition of embodied logatomes is believed to have a priming effect and thereby augment the saliency of the target sentences' prosodic features.

3.1.5 Benefits of embodied prosodic training

To our knowledge, only a few studies have empirically assessed the effects of using hand gestures on pronunciation, however not directly within the framework of embodied learning. For example, the perception and production of rhythmic movements such as simple up-and-down or back-and-forth motions of the hands - also called beat gestures - have been found to aid Catalan learners' accentedness and fluency in English (Author 2017; Author 2018). Recent studies have investigated the role of handclapping in second language pronunciation and found it beneficial for the perception of Japanese long vowels in English speakers (Iizuka et al. 2020) and the accentedness of young, Catalan and Chinese naïve learners of French (Author 2021; Author 2020). Hand gestures depicting specific suprasegmental properties such as vowel duration in Japanese (durational gesture, Author 2021) and intonation contours in Spanish (pitch gesture, Author 2018) have also shown positive effects on the pronunciation of these features.

Regarding more classroom based approaches, only a few empirical studies have assessed the potential beneficial effects of embodied prosodic training for L2 pronunciation through the EHIEP and VT techniques. Mister et al. (2021) tested the rhythmic fight club technique with adult learners of English and observed that drawing attention to word stress patterns enhanced the accuracy of learners' pronunciation of words in subsequent oral production in terms of stress placement. However, this study did not include a control group and it remains unclear whether the technique employed in the training would outperform other kinds of non-embodied techniques. Author (2010) found that eight weeks of global VT phonetic training sessions improved learners' fluency in L2 French more than training sessions based on reading aloud, text comprehension and creative writing. However, it remains unclear whether the gains were due to the listen-and-repeat tasks, the use of the logatomes or the use of hand gestures. Later, Author (2013) compared the VT method to the articulatory method, which involves the explicit teaching of segments' articulatory properties, and found that after four weeks, participants following the VT method showed significantly higher gains in their fluency, in particular when their French pronunciation was worse at the outset. However, this advantage disappeared after eight weeks of training. According to the author, the introduction of written activities during the second half of the course, and more specifically the intellectualization that goes with this type of activities, instead of

improving reading fluency, may have led to a decline in pronunciation performance. These results indicate that it may also be necessary to practice oral-reading pronunciation.

All in all, while previous research has been mostly centered on testing specific prosodic aspects in laboratory settings, classroom-based studies remain scarce and reveal inconsistent results. Therefore, more empirical research is needed to assess the effects of embodied pronunciation teaching, in particular research with more classroom-based, learner-oriented experimental designs. Importantly, training pronunciation with visuospatial gestures mimicking the prosodic features of the target language does not only have pedagogical implications but also allows for the testing of the predictions of Embodied Cognition Theory for phonological learning.

3.2.6 The present study

The present study aimed to assess the efficacy of embodied pronunciation training on oral reading through visuospatial hand gesture movements mimicking the melodic and rhythmic patterns of target sentences. We hypothesised that embodied prosodic training involving the imitation of logatomes and gestures would yield greater improvements in oral-reading pronunciation than prosodic training that involved repeating logatomes, compared to a baseline condition where participants repeated speech only.

The participants for the present study were bilingual Catalan-Spanish speakers learning French as an additional language. Despite the close relationship between the Romance languages, Catalan learners face clear challenges in the acquisition of French prosody. Unlike French, Catalan does not have a phrasal-marked Accentual Phrase (AP) constituent (Prieto et al., 2015). In French, the AP may group together more than one lexical word plus the accompanying clitics, and it is characterised by the presence of an obligatory final pitch accent and an optional initial rise, which have a demarcative function (Delais-Roussarie et al. 2015). This means that while stress functions on a phrasal level in French, marking right—and optionally left—phrase boundaries (see, among others, Di Cristo & Hirst 1993; Jun & Fougeron 1995, 2000; Delais-Roussarie et al. 2015), in Catalan, as in Spanish, it works on a lexical level (Mascaró 1976; MacPherson 1975). As a consequence of the lack of lexical stress in French, there is a strong syncretism between accentuation, phrasing, and intonation, while, by contrast, Catalan and Spanish generally group two or three prosodic words, with no initial or final demarcative tonal features (Nibert 2000; Author 2015). A second basic difference between the two languages lies in the phonetic properties of stress realisation, which mainly affects the duration of the stressed syllable. French stress is realised by a more extreme lengthening of the stressed phrase-final syllable, and more particularly of the full vowel, than what is seen in Catalan (Fletcher 1991; Vaissière 1991; Di Cristo &

Hirst 1993; Astésano 2001). This was demonstrated by Author (2021), who carried out an exploratory acoustic analysis comparing the duration of sentence-final stressed and unstressed syllables in 20 pairs of cognate words in Catalan and French (e.g. balcó – balcon 'balcony') and detected significantly longer phrase-final syllable durations in the French words.

Reading aloud is a common language classroom practice in any number of activities, despite the fact that it may hinder comprehension (e.g. Gabrielatos 2002; but see Gibson 2008 for exceptions). For the purpose of teaching pronunciation, it may strengthen the grapho-phonemic correspondences of the L2 (e.g. Gibson 2008) and improve learners' fluency (e.g. Klomjit 2013). Riquelme Gil et al. (2017) tested the Repeated Reading method by asking young Spanish learners of English to read short passages from a story book both silently and aloud over the course of six weeks. They found that, after the intervention, participants produced less pronunciation errors in three different tasks carried out at pre- and posttest: re-reading of the original text, reading of an unknown text and spontaneous speech. In this study, we adopted the Repeated Reading method as a way to introduce our stimuli and training materials.

Finally, oral proficiency in a language is most often rated in terms of comprehensibility, fluency and accentedness (e.g. Munro & Derwing 2015). However, more concrete features, from a wide

range of suprasegmental features (e.g. Munro 1995; Trofimovich & Isaacs 2012; Saito et al. 2016) to segments with high functional load (e.g. Suzukida & Saito 2019), can also be rated to evaluate pronunciation accuracy. Thus, five dimensions were selected here to assess participants' pronunciation before and after training: comprehensibility, fluency, accentedness, segmental accuracy and suprasegmental accuracy. This is in keeping with the view expressed by Saito and Plonsky (2019) that a truly comprehensive assessment of the effects of pronunciation instruction on L2 speech must take into account both holistic and specific levels of measurement.

3.2 Methods

3.2.1 Participants

Seventy-five first- or second-year students doing undergraduate degrees in Translation and Interpreting or Applied Language at the Universitat Pompeu Fabra in Barcelona participated in this study. They were all enrolled in an intermediate-level French course, which consisted of 90 minutes of language theory and 60 minutes of language practice (including a variety of oral and written activities) per week over a four-month term. This pronunciation training study was incorporated into the French course and took place over five weeks. Participation was therefore mandatory for all students. The actual pronunciation training was carried out by the first author.

All of the students reported themselves to be Catalan-Spanish bilinguals. Results of a preliminary questionnaire showed that as a group they used Catalan 61% of the time on average in their daily lives (SD = 28.4). Participants self-reported their French proficiency to be between CEFR levels A2 and B1. They also reported that they had studied English as a foreign language to one extent or another. Prior to participation in this study, they all signed a form consenting to the use of audio recordings of their speech for the purposes of this research.

The 75 participants were randomly assigned to one of the three conditions such that the speech-only group contained 27 participants (Mage = 20.04, SD = 2.87 2 males), the non-embodied logatome group contained 22 (Mage = 19.79, SD = 1.32, 4 males), and the embodied logatome group contained 26 (Mage = 19.80, SD = 1.37 2 males).

An a priori power analysis was conducted using G*power3 to test the interaction between groups and tests (ANOVA: repeated measures, within-between factors; medium target effect size η2 = 0.04, alpha = .05). Results showed that a total sample of 66 participants was required to achieve a power of .95.

3.2.2 Materials

a) Audiovisual stimuli for the pronunciation training sessions

All the materials used in this experiment can be seen in Appendix A and they are openly available at https://osf.io/93pdw/?view_only=d2c77e66c557404da94d0428ebfaeaf0.

The materials used in the training sessions consisted of dialogues taken from a French language textbook that focuses on teaching oral skills through meaningful, enjoyable texts (Martins & Mabilat 2003). Nine dialogues were used in the training sessions, with a different set of three employed in each of the three sessions. While the intention was to select target dialogues that did not include

novel vocabulary, a short glossary in Catalan was provided adjacent to each text to be read clarifying any words that might be unfamiliar to lower-level participants. In addition, dialogues were chosen such that the oral performance of the dialogues would include a variety of intonation contours arising from different situational contexts.

A total of five sentences in each dialogue were selected (around 42% of the total number of sentences) to be target stimuli for repetition during the training sessions. Video recordings were then made of three instructors performing these five stimuli in the three experimental conditions. The instructors (2 female, 1 male) were two specialists in the VT method and the first author of this study. Recording took place over four hours at the second author's university broadcasting studio with professional equipment and help from a technical assistant (See Appendix B for a detailed description of the recording procedure).

In all recordings, the frame of the image was set to show the upper half of each instructor's body to allow a clear view of the face and all hand movements. For the speech condition, the instructors simply pronounced the target sentences clearly while standing still. For the non-embodied logatome condition, the logatome consisted of pronouncing the syllable "da" instead of the phrase's syllables, but without changing the intonation of the phrase. As for the embodied logatome condition, as the logatome was uttered, the

right hand, palm open facing downward, made a sweeping left-to-right movement across the body at chest level that mimicked through upward and downward movements the rises and falls of the pitch contours of their oral utterance as they spoke. Importantly, these movements served to depict not only intonational pitch movements but also the rhythmic features of their speech by increased or decreased velocities and short pauses in the hand's movement. Figure 1 shows sequences of video stills from a sample stimulus trial in the non-embodied (top panel) and embodied logatome condition (middle panel), as well as the pitch contour and corresponding logatome "da" syllables (bottom panel).

The video clips were edited in Adobe Premiere Pro 13 to create three sets of stimulus materials corresponding to one of the experimental conditions (speech-only, non-embodied logatome plus speech, embodied logatome plus speech). Figure 2 shows the training sequence for each sentence. The instructor pairs varied throughout the stimuli for the nine dialogues; however, the combination for each dialogue was consistent across the three conditions. Therefore, all participants were able to listen to the three instructors throughout the course of the training.

Finally, the nine dialogues were acted out by amateur actors in appropriate locations, either in France or in Catalonia (but using French native speakers) and video-recorded. After each dialogue had been trained, the participants would be shown the

corresponding enactment as a kind of wrap-up activity (these video files are available at https://osf.io/93pdw/?view_only=d2c77e66c557404da94d0428ebfaeaf0).

Thus, the final material for each session consisted of three training videos (five sentences each) in one of the experimental conditions, each one followed by the enactment of the full source dialogue as a wrap-up. This material was embedded in an online presentation format using Alchemer software, accompanied by written instructions. Since training involved three separate sessions, three such presentations were prepared for each condition.

Figure 1

Stills from stimulus videos showing an instructor performing in the non-embodied (top panel) and embodied logatome conditions (middle panel). In this case the target sentence is Je suis désolée, votre lettre n'est pas là 'I am sorry, your letter is not here'. Acoustic data and the intonation pattern of the logatome sequence is shown in the bottom panel

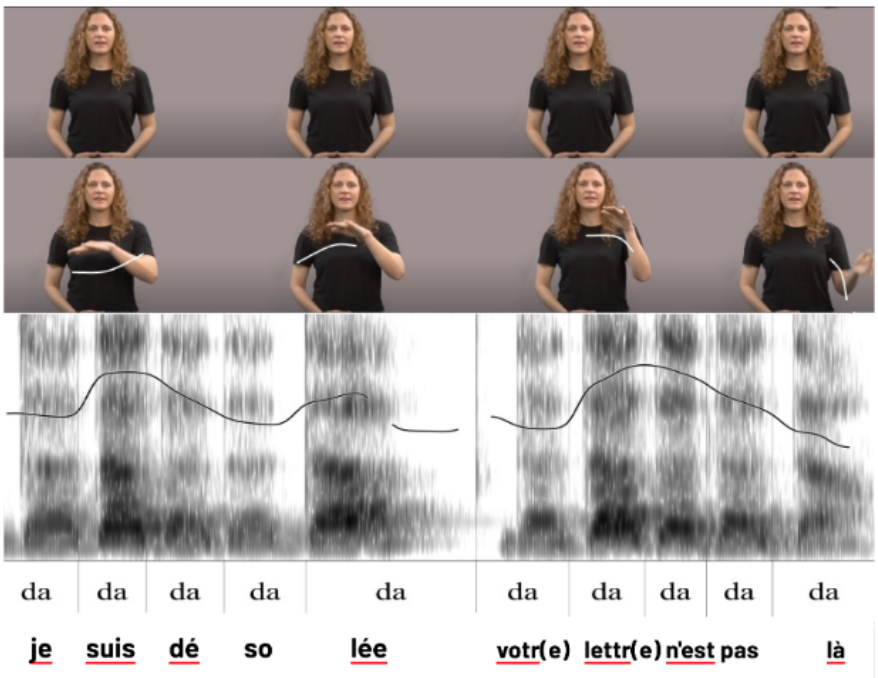
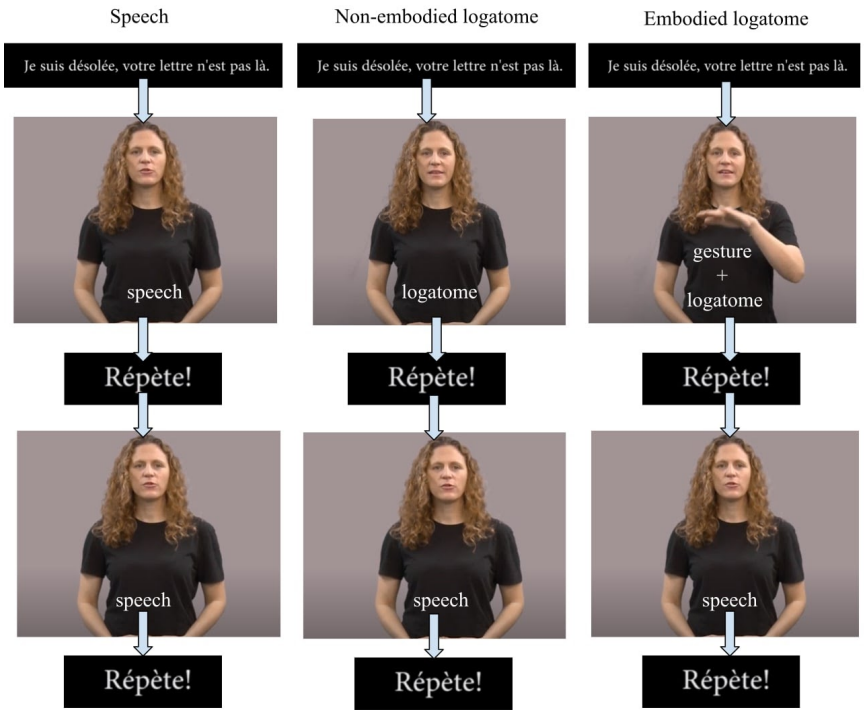


Figure 2

Audiovisual training sequence for each sentence



b) Pretest and posttest materials and control measures

Participants' pronunciation was tested before and after training by means of a dialogue-reading task. The pretest and posttest were identical and consisted of four dialogues to be read aloud, three of them also appearing in a training session (one dialogue from each set of three used in the training sessions) and the fourth being untrained. These materials as well as corresponding instructions were uploaded to the same online presentation platform as the

training materials and could be accessed by a link provided by the teacher.

Two sets of data were gathered to control for potential differences between groups. The first set covered participants' self-reported proficiency in French and prior experience learning that language. This questionnaire yielded four separate scores per participant: the number of years spent learning French; the number of months spent learning French as an extracurricular activity (outside school/university); the number of months spent abroad in a French-speaking country; and a nominal value from 1 to 6 indicating self-reported proficiency in French (A1 = 1, A2 = 2, B1 = 3, B2 = 4, B2= 4, C1 = 5, C2 = 6).

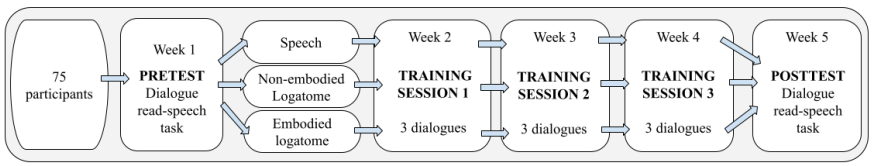
The second set of data, which was gathered at the end of the posttest, was a rating of the participants' satisfaction with regard to the pronunciation training they had received. This online questionnaire asked the participants to rate their satisfaction with the pronunciation training they had received by reacting to the following statements on a scale from 1 ('I strongly disagree') to 9 ('I strongly agree'): a) I liked these pronunciation training sessions; b) I improved my pronunciation; and c) I would like to repeat this kind of activity with other texts.

3.2.3 Procedure

Figure 3 provides an overview of the experimental design of the three-session training programme with pre- and posttest. A week prior to the first training session, participants received from their respective French language instructors a link to the website containing the materials for the pretest task, which consisted of video-recording themselves as they read aloud four dialogues. The full task took on average ten minutes. They were required to complete the task and upload the resulting video files to a shared folder within three days of having received the link from their instructor. Participants were asked to carry out the pretest using their own computer and headset in a quiet environment. The purpose of video recording was to ensure that the tasks were done properly, and uploaded student recordings were regularly checked by the first author for this purpose. The audio tracks from the recordings were then extracted and saved for further analysis.

Figure 3

Diagram of the experimental design of the training programme

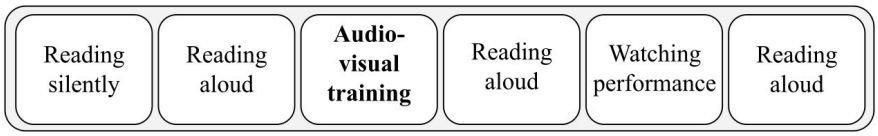


The experimental training took place in three separate sessions over three weeks during the regular class period of participants' course in French language. All sessions took place on the university premises in individual soundproofed booths equipped with computers and microphones. Before starting the first training session, the students answered the language questionnaire. The teacher then emailed a link to one of three separate sets of training materials, depending on the experimental group to which the participant had been randomly assigned previously. After reading some initial instructions, participants then completed the training procedure individually at their own pace, recording their speech output throughout using Audacity software. The training procedure consisted of completing a set of subtasks associated with three dialogues. Figure 4 shows the sequence of subtasks related to one dialogue. Participants moved from one step in the process to the next by means of clicking on their keyboard. The order of presentation of the dialogues was randomized automatically by the

software. Each full training session lasted roughly 30 minutes, about 10 minutes per dialogue unit.

Figure 4

Procedure of a trial involving one dialogue. Each full training session consisted of three such sequences



Once they had completed all three training tasks, participants stopped the recording process and uploaded the resulting audio file to a shared folder. While the training session was in progress, the instructor monitored participant behavior from outside the individual booths, particularly to ensure that participants in the embodied logatome training were duly performing the required hand movements. Because the class period was longer than the time required for the training session, once they had completed the training task, participants then proceeded to complete other language-learning activities assigned by their instructor. One week after the third and last training session took place, participants took the posttest, which, like the pretest, consisted of recording themselves reading four dialogues aloud and then uploading the recordings to a shared folder.

Pronunciation assessment

Assessment was carried out by three raters (2 female 1 male), all native speakers of French with extensive L2 teaching experience with Catalan learners. They took part in a one-hour training session to receive detailed explanations about the five dimensions they were to evaluate and instructions on how to apply the nine-point assessment scales (1 = worst score, 9 = best score) with which they would rate participant output on each of the five dimensions. They then individually practiced applying these scales using five sample dialogues read by the participants and the first author provided feedback to ensure a clear understanding of the five dimensions.

Each rater evaluated the totality of speech samples taken from participant-recorded pretest and posttest audio files ((4 pretest dialogues + 4 posttest dialogues) × 75 participants = 600 audio files) for each of the five pronunciation dimensions, giving a total of 3,000 scores. The speech samples consisted of the full dialogues (durations in s: M = 27.86, SD = 6.07 for dialogue 1, M = 41.28, SD = 7.78 for dialogue 2, M = 27.61, SD = 4.83 for dialogue 3, and M = 27.39, SD = 4.84 for dialogue 4)) and were randomized and grouped into sixteen different batches using Alchemer online software. Each batch took about one hour to rate. While we advised the raters to take only short breaks during each batch rating, so as not to lose the data, we recommended they rested as much as needed between each batch in order to avoid listener fatigue. The

raters rated the dataset at home over the course of seven days by completing one to three batches per day. They received monetary compensation for their work.

Items' internal consistency was checked by means of Cronbach's alpha and satisfactory coefficients were obtained (0.93 for comprehensibility, 0.92 for fluency, 0.79 for accentedness, 0.92 for segmental accuracy and 0.88 for suprasegmental accuracy). Interrater reliability was assessed by calculating the intraclass correlation coefficient (two-way random, absolute agreement, see Landers 2015), showing moderate to good agreement among the raters: ICC = 0.56, $F(599\ 1198) = 2.29$, $p < .001$, 95% CI [0.50, 0.62] for comprehensibility, ICC = 0.64, $F(599\ 1198) = 2.79$, $p < .001$, 95% CI [0.59, 0.69] for fluency, ICC = 0.73, $F(599\ 1198) = 3.76$, $p < .001$, 95% CI [0.69, 0.77] for accentedness, ICC = 0.61, $F(599\ 1198) = 2.59$, $p < .001$, 95% CI [0.56, 0.66] for segmental accuracy and ICC = 0.72, $F(599\ 1198) = 3.64$, $p < .001$, 95% CI [0.68, 0.76] for suprasegmental accuracy.

Statistical analysis

Statistical analyses were carried out with IBM SPSS 23. Two databases were set up, one sorted by participant and the other sorted by item (i.e., stimulus sentence). In order to test for homogeneity across the three groups, the participant-sorted database was used to show individual scores for the self-reported French language proficiency measures and satisfaction with the

training experience. As the measures showed a skewed distribution, differences between groups and mean satisfaction scores across groups were explored by means of a non-parametric Kruskal-Wallis H test.

The item-sorted database was used to analyze the effect of type of training (speech vs. logatome vs. embodied logatome) on participant pronunciation measures. Five general linear mixed models (GLMMs) were run, each with the one of the following dependent variables: comprehensibility, fluency, accentedness, segmental accuracy and suprasegmental accuracy. For all these variables, Shapiro-Wilk tests showed that the scores were positively skewed. Therefore, an inverse Gaussian distribution with a log function was specified in each model. Group (3 levels: speech only, logatome, embodied logatome) and Session (2 levels: pretest and posttest), Group × Session, and Familiarity (2 levels: trained and untrained items) were set as fixed factors; random intercepts were set for participants and for items. Sequential Bonferroni pairwise comparisons were used.

3.3 Results

3.3.1 Homogeneity across groups

Results of the Kruskal-Wallis H test showed that there was no significant difference between the three groups in terms of age, $\chi^2(2) = 0.62$, $p = 0.73$, years of learning French, $\chi^2(2) = 0.36$, $p = 0.84$, months of extra-curricular French lessons, $\chi^2(2) = 0.80$, $p = 0.77$, months of stay abroad, $\chi^2(2) = 0.45$, $p = 0.80$ and self-assessed French proficiency, $\chi^2(2) = 3.02$, $p = 0.22$ (see Table 1).

The result of the GLMM with comprehensibility as the dependent variable showed a significant effect of session, $F(1, 1793) = 18.63$, $p < .001$, $\eta^2 = .01$, 90% CI [.004, .02]. No significant effect of group, Session × Group or familiarity were found. Post hoc analyses revealed a significant effect of session for the three groups, $F(1, 1793) = 4.28$, $p = .04$, $\eta2 = .002$, 90% CI [.0076, .0079] for the speech only group, $F(1, 1793) = 3.57$, $p = .059$, $\eta^2 = .002$, 90% CI [0, .007] for the non-embodied logatome group, and $F(1, 1793) = 12.56$, $p < .001$, $\eta^2 = .007$, 90% CI [.002, .015] for the embodied logatome group.

Table 1

Descriptive statistics and rank mean values for age and French proficiency measures in each group

	Group	*M*	*SD*	*SE*	95% CI	Rank mean
Age	Speech only	20.04	2.91	0.55	[18.91 21.16]	35.64
	Non-embodied logatome	19.82	1.37	0.29	[19.21 20.42]	39.93
	Embodied logatome	19.76	1.30	0.26	[19.22 20.30]	38.94
Formal instruction in French	Speech only	4.25	2.58	0.49	[3.25, 5.25]	36.20
	Non-embodied logatome	4.41	2.01	0.43	[3.51, 5.30]	39.73
	Embodied logatome	5.08	3.84	0.76	[3.49, 6.66]	38.50
Extra-curricular instruction in French	Speech only	0.71	1.01	0.19	[0.32 1.11]	38.20
	Non-embodied logatome	0.81	1.11	0.24	[0.31 1.30]	39.95
	Embodied logatome	0.73	1.36	0.27	[0.17 1.29]	36.06
Stay abroad	Speech only	0.69	1.53	0.29	[0.09 1.28]	36.91
	Non-embodied logatome	0.84	1.72	0.36	[0.08 1.60]	37.14
	Embodied logatome	0.66	1.26	0.25	[0.14 1.18]	39.98
Self-reported proficiency	Speech only	3.71	0.71	0.13	[3.44, 3.99]	35.57
	Non-embodied logatome	3.73	0.93	0.20	[3.31, 4.14]	34.52
	Embodied logatome	4.04	0.89	0.18	[3.67, 4.41]	43.78

3.3.2 Training effects

A general improvement between pre- and posttest was observed in all the measures. A general view of the results is presented in Figure 5. The descriptive results are gathered in Table 2. Below, we report only the significant results. All the inferential statistical results are available in Appendix C.

The result of the GLMM with fluency as the dependent variable showed a significant effect of session, $F(1, 1793) = 96.50$, $p < .001$, $\eta^2 = 0.05$, 90% CI [.973, .976]. No significant effect of group, Group × Session or familiarity were found. Post hoc analyses revealed a significant effect of session for the three groups, $F(1, 1793) = 28.03$, $p < .001$, $\eta^2 = .01$, 90% CI [.007, .026] for the speech only group, $F(1, 1793) = 25.84$, $p < .001$, $\eta^2 = .01$, 90% CI [.006, .025] for the non-embodied logatome group, and $F(1, 1793) = 44.57$, $p < .001$, $\eta^2 = .02$, 90% CI [.01, .04] for the embodied logatome group.

The result of the GLMM with comprehensibility as the dependent variable showed a significant effect of session, $F(1, 1793) = 18.63$, $p < .001$, $\eta^2 = .01$, 90% CI [.004, .02]. No significant effect of group, Session × Group or familiarity were found. Post hoc analyses revealed a significant effect of session for the three groups, $F(1, 1793) = 4.28$, $p = .04$, $\eta2 = .002$, 90% CI [.0076, .0079] for the speech only group, $F(1, 1793) = 3.57$, $p = .059$, $\eta^2 =$

.002, 90% CI [0, .007] for the non-embodied logatome group, and $F(1, 1793) = 12.56$, $p < .001$, $\eta^2 = .007$, 90% CI [.002, .015] for the embodied logatome group.

The result of GLMM with accentedness as the dependent variable showed a significant effect of session, $F(1, 1793) = 68.14$, $p < .001$, $\eta^2 = .03$, 90% CI [.02, .05], and Group × Session, $F(2, 1793) = 7.38$, $p = .001$, $\eta^2 = .008$, 90% CI [.002, .016]. No significant effect of group or familiarity were found. Post hoc analyses revealed a significant effect of session for the three groups, $F(1, 1793) = 9.28$, $p = .002$, $\eta^2 = .005$, 90% CI [.001, .012] for the speech only group, $F(1, 1793) = 11.53$, $p = .001$, $\eta^2 = .006$, 90% CI [.002, .014] for the non-embodied logatome group, and $F(1, 1793) = 61.91$, $p < .001$, $\eta^2 = .03$, 90% CI [.02, .05] for the embodied logatome group, as well as a significant difference between groups at posttest only, $F(2, 1793) = 3.50$, $p = .03$, $\eta^2 = .004$, 90% CI [.0001, .009], with significantly higher improvement in the embodied logatome group than in the speech only group, $p = .04$.

Table 2

Mean, standard deviation, standard error and 95% confidence intervals at pre- and posttest for the speech only group, the non-embodied logatome group and the embodied logatome group, and for trained and untrained items (Familiarity) in the five pronunciation assessment measures

		Comprehensibility				Fluency			
		Mean	*SD*	*SE*	95% CI	*Mean*	*SD*	*SE*	95% CI
Speech only	pretest	7.29	1.41	.08	[7.13, 7.44]	7.29	1.41	.08	[7.13, 7.44]
	posttest	7.49	1.30	.07	[7.35, 7.63]	7.49	1.30	.07	[7.35, 7.63]
Non-embodied logatome	pretest	7.32	1.43	.09	[7.15, 7.50]	7.32	1.43	.09	[7.15, 7.50]
	posttest	7.53	1.36	.08	[7.36, 7.70]	7.53	1.36	.08	[7.36, 7.70]
Embodied logatome	pretest	7.30	1.39	.08	[7.30, 7.46]	7.30	1.39	.08	[7.30, 7.46]
	posttest	7.67	1.10	.06	[7.54, 7.79]	7.67	1.10	.06	[7.54, 7.79]
Familiarity	trained	7.44	1.35	.04	[7.37, 7.51]	7.44	1.35	.04	[7.37, 7.51]
	untrained	7.40	1.33	.06	[7.28, 7,52]	7.40	1.33	.06	[7.28, 7,52]

		Accentedness				Segmental accuracy			
		Mean	*SD*	*SE*	95% CI	*Mean*	*SD*	*SE*	95% CI
Speech only	pretest	5.93	1.27	.07	[5.80, 6.07]	6.29	1.31	.07	[6.14, 6.43]
	posttest	6.18	1.23	.07	[6.05, 6.31]	6.49	1.26	.07	[6.35, 6.62]
Non-embodied logatome	pretest	5.94	1.22	.07	[5.79, 6.09]	6.27	1.34	.08	[6.11, 6.43]
	posttest	6.25	1.15	.07	[6.11, 6.39]	6.60	1.30	.08	[6.35, 6.62]
Embodied logatome	pretest	5.92	1.16	.07	[5.79, 6.06]	6.28	1.34	.08	[6.12, 6.43]
	posttest	6.61	1.05	.06	[6.49, 6.73]	6.63	1.25	.07	[6.49, 6.78]
Familiarity	trained	6.14	1.20	.03	[6.08, 6.20]	6.42	1.31	.04	[6.35, 6,49]
	untrained	6.13	1.23	.06	[6.02, 6.24]	6.43	1.31	.06	[6.31, 6.55]

		Suprasegmental accuracy			
		Mean	*SD*	*SE*	95% CI
Speech only	pretest	6.69	1.16	.06	[6.57, 6.81]
	posttest	7.18	0.99	.05	[7.07, 7.28]
Non-embodied logatome	pretest	6.61	1.20	.07	[6.46, 6.76]
	posttest	7.24	1.04	.06	[7.12, 7.37]
Embodied logatome	pretest	6.67	1.22	.07	[6.53, 6.80]
	posttest	7.55	0.95	.05	[7.44, 7.66]
Familiarity	trained	7.01	1.17	.03	[6.95, 7.07]
	untrained	6.93	1.08	.05	[6.83, 7.03]

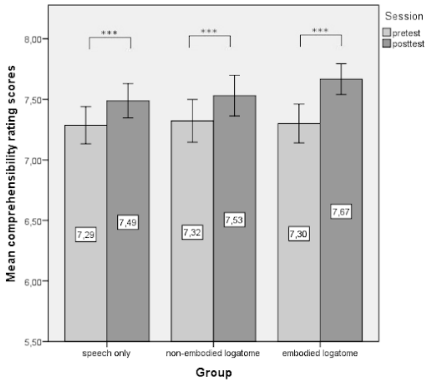
The result of the GLMM with segmental accuracy as the dependent variable showed a significant effect of session, $F(1, 1793) = 27.44$, $p < .001$, $\eta^2 = 0.01$, 90% CI [.007, .02]. No significant effect of group, Session × Group or familiarity were found. Post hoc analyses revealed a significant effect of session for the three groups, $F(1, 1793) = 4.79$, $p = 0.03$, $\eta^2 = .003$, 90% CI [.0001, .008] for the speech only group, $F(1, 1793) = 10.21$, $p = .001$, $\eta^2 = .006$, 90% CI [.001, .01] for the non-embodied logatome group, and $F(1, 1793) = 12.28$, $p < .001$, $\eta^2 = .007$, 90% CI [.002, .01] for the embodied logatome group.

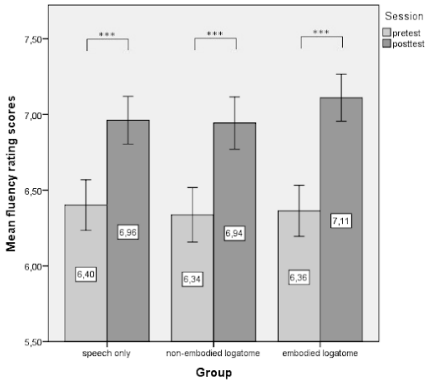
The result of the GLMM with suprasegmental accuracy as the dependent variable showed a significant effect of session, $F(1, 1793) = 197.89$, $p < .001$, $\eta^2 = .10$, 90% CI [.08, .12], and Session × Group, $F(2\ 1793) = 6.25$, $p = .002$, $\eta^2 = .007$, 90% CI [.002, ..01]. No significant effect of group or familiarity were found. Post hoc analyses revealed a significant effect of session for the three groups, $F(1, 1793) = 39.59$, $p < .001$, $\eta^2 = .02$, 90% CI [.01, .03] for the speech only group, $F(1, 1793) = 52.89$, $p < .001$, $\eta^2 = .03$, 90% CI [.02, ..04] for the non-embodied logatome group, and $F(1, 1793) = 116.32$, $p < .001$, $\eta^2 = .06$, 90% CI [.04, .08] for the embodied logatome group, as well as a significant difference between the embodied logatome group and the speech only group at posttest, $F(2, 1793) = 3.40$, $p = .03$, $\eta^2 = .004$, 90% CI [.0001, .009].

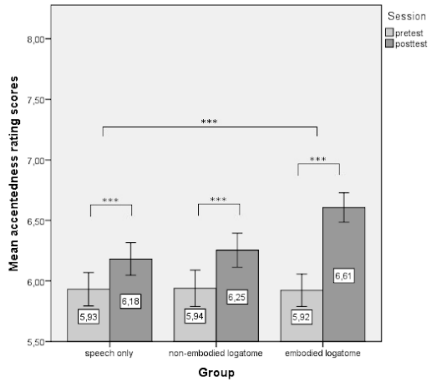
In sum, results showed a significant improvement in read-speech comprehensibility, fluency, accentedness, segmental accuracy, and suprasegmental accuracy after training in all the three groups, with higher effect sizes for the embodied logatome group in all the measures. In addition, the embodied logatome group improved significantly more than the speech-only group in terms of accentedness and suprasegmental accuracy, while the non-embodied logatome group did not (see Appendix C for fixed effects and contrast estimates). No significant differences were found between trained and untrained items, showing that participants improved equally in their pronunciation of French when reading a text aloud regardless of whether they had received prior training with that particular text or not.

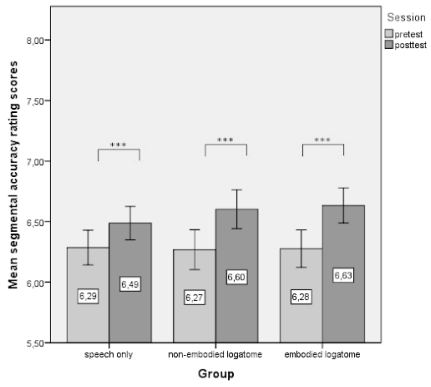
Figure 5

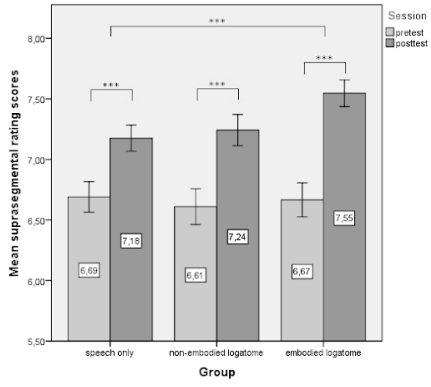
*Mean rating scores at pre- and posttest for the five pronunciation measures. Significant contrasts are labeled with asterisks (***: $p < .001$)*


Session
pretest
posttest
Mean comprehensibility rating scores
8,00
7,50
7,00
6,50
6,00
5,50
***
7,29
7,49
7,32
7,53
7,30
7,67
speech only
non-embodied logatome
embodied logatome
Group
Error Bars: +/- 2 SE


Session
pretest
posttest
Mean fluency rating scores
7,50
7,00
6,50
6,00
5,50
***
6,40
6,96
6,34
6,94
6,36
7,11
speech only
non-embodied logatome
embodied logatome
Group
Error Bars: +/- 2 SE


Session
pretest
posttest
Mean accentedness rating scores
8,00
7,50
7,00
6,50
6,00
5,50
***
5,93
6,18
5,94
6,25
5,92
6,61
speech only
non-embodied logatome
embodied logatome
Group
Error Bars: +/- 2 SE


Session
pretest
posttest
Mean segmental accuracy rating scores
8,00
7,50
7,00
6,50
6,00
5,50
***
6,29
6,49
6,27
6,60
6,28
6,63
speech only
non-embodied logatome
embodied logatome
Group
Error Bars: +/- 2 SE


Session
pretest
posttest
Mean suprasegmental rating scores
8,00
7,50
7,00
6,50
6,00
5,50
***
6,69
7,18
6,61
7,24
6,67
7,55
speech only
non-embodied logatome
embodied logatome
Group
Error Bars: +/- 2 SE

3.3.3 Satisfaction with training

Results of the general satisfaction questionnaire showed high degrees of satisfaction for the three measures, as shown in Table 4. The Kruskal-Wallis H test showed that there was no statistically significant difference among the groups in terms of likeability of the activity ($\chi^2(2) = 4.18$, $p = .12$, with a mean rank score of 79.72 for the speech only group, 64.31 for the non-embodied logatome group, and 68.41 for the embodied logatome group), self-perception of improvement ($\chi^2(2) = 3.58$, $p = .17$, with a mean rank score of 79.54 for the speech group, 68.17 for the non-embodied logatome group, and 65.11 for the embodied logatome group), and interest in repeating the activity ($\chi^2(2) = 76.06$, $p = .53$, with a mean rank score of 76.06 for the speech group, 66.83 for the non-embodied logatome group and 70.41 for the embodied logatome group.

Table 3

Mean results for the satisfaction questionnaire across the three groups based on a 1-9 scale, from 1 ('I strongly disagree') to 9 ('I strongly agree')

	Group	*Mean*	*SD*	*SE*	95% CI for Mean
I liked the pronunciation training sessions.	Speech only	8.30	0.82	.11	[8.07, 8.52]
	Non-embodied logatome	7.81	1.19	.18	[7.44, 8.18]
	Embodied logatome	7.74	1.58	.23	[7.27, 8.21]
I think I improved my pronunciation.	Speech only	6.93	1.40	.19	[6.54, 7.31]
	Non-embodied logatome	6.48	1.67	.26	[5.96, 7.00]
	Embodied logatome	6.30	1.64	.24	[5.82, 6.79]
I would like to repeat this kind of activity with other texts.	Speech only	7.30	1.47	.20	[6.89, 7.70]
	Non-embodied logatome	6.71	2.10	.32	[6.06, 7.37]
	Embodied logatome	6.87	2.05	.30	[6.26, 7.48]

3.4 Discussion

The present study explored the effects of embodied prosodic training via visuospatial gestures depicting rhythm and intonation on overall (comprehensibility, fluency, accentedness) and specific (segmental and suprasegmental accuracy) measures of pronunciation with Catalan intermediate learners of French. This embodied training was embedded in repeated reading and oral imitation activities, while the effects on pronunciation were assessed through an oral-reading task. One week after the last session of our intervention, the speech-only group, the non-embodied prosodic group (logatome only), and the embodied logatome group (logatome and gesture) significantly improved in all the measures compared to pretest. Our results revealed that participants in the embodied logatome group obtained significantly higher gains compared to the speech-only group in terms of accentedness and suprasegmental accuracy, while the non-embodied logatome group did not. Nonetheless, the difference between the non-embodied and embodied logatome groups was not significant in any of the measures, despite systematic larger effect sizes in the improvement between pretest and posttest for the embodied logatome group. These results demonstrate that only when accompanied by a gesture did the logatome a superior effect on learning outcomes.

Our results provide evidence that the embodiment of a phonological feature in a foreign language helped learners process this specific feature more efficiently: embodying prosody directly improved the scores on suprasegmental accuracy. The motoric action provided by the perception and the production of the gesture may have reached the visuospatial phonological loop (Wilson & Emmorey 1997) and may have been associated to the adequate mental representation of rhythmic and melodic patterns, facilitating the processing and the acquisition of such features. In the absence of the visuospatial gesture, the ability of prosodic training with only logatomes to convey the saliency of suprasegmental features may not have been sufficient to make a difference. Our study thus supports the claims in favor of embodied techniques for teaching pronunciation (e.g. Billières 2002; Acton et al. 2013; Chan 2018) and sheds a new light on the mechanism behind the positive effects of rhythmic embodied training involving rhythmic gestures (Author 2017; Author 2019) or hand clapping (Author 2020; Lee et al. 2020; Author 2021), and embodied gesture training focusing on specific segmental or suprasegmental features (Author 2018; Author 2020).

At the practical level, our findings offer additional evidence of the efficacy of one of the main features of the VT method, namely the use of embodied logatomes. Though our findings confirm previous results that the use of embodied logatomes may not provide an advantage for fluency measures (Author 2013) or segmental

accuracy (Author 2018), they indicate that this method is able to boost pronunciation learning in terms of accentedness and suprasegmental accuracy. It is of interest to note in the present study the high level of participant satisfaction in all groups, indicating that they felt at ease with a repetition paradigm involving short dialogues, whether this included logatomes and embodiment or not. This suggests that the introduction of novel, maybe unusual methodologies, including using one's body was not a hindrance to learning - on the contrary. Hence, the use of embodied techniques may be of particular interest for language teachers who detect the need to improve their learners' pronunciation at any time during their class, without requiring any materials or heavy preparation.

The lack of any difference between the three training conditions for comprehensibility and fluency measures could be explained by the fact that suprasegmental features may weigh less in these measures than in the accentedness measure (e.g. Trofimovich & Isaacs 2012; Saito et al. 2016). However, the larger effect sizes obtained for embodied prosodic training in both measures may also point to a certain advantage for this type of training, which might be amplified if the duration of the training period were extended. As Author (2013) pointed out, a longer training period may be necessary to widen the differences among the groups with respect to fluency scores. In addition, regarding comprehensibility scores, it may be the case that these scores were already too high to be able

to detect sufficiently large differences between groups and that effects may have been observed with learners of lower proficiency.

In line with previous research that demonstrated the value of pronunciation instruction, our results showed that 30 minutes of pronunciation training once a week over three weeks helped improve significantly comprehensibility, fluency and accentedness in L2 read speech regardless of the training method, as our three experimental groups obtained significantly higher scores in those measures at posttest. Moreover, following the recommendation by Saito and Plonsky (2019), this study encompasses the three traditional overall measures of pronunciation, as well as specific segmental and suprasegmental measures, whereas previous literature tends to focus on only one of these aspects. Furthermore, this improvement in pronunciation was evident even in the one read dialogue for which they had not been trained, showing that participants may have been able to generalize what they learned during training to an untrained item and adding some evidence on the generalization of pronunciation gains after pronunciation training, an issue that is seldom raised (e.g. Levis & Pickering 2004).

There are several limitations to the present study. First, we only obtained moderate inter-rater reliability between the three raters. We think that perceptively evaluating long samples of read-speech (between 20 and 40 seconds) on a scale from one to nine might

have allowed for more variability than evaluating single sounds or words. Second, our results are restricted to pronunciation in read speech. Although our findings on oral reading can be considered useful for improving learners' pronunciation – notably, because oral reading is a common task in the second language classroom - it is not clear whether the benefits of the embodied logatome technique would extend to spontaneous speech. As suggested by Saito and Plonsky (2019), more evidence is needed on the effect of perceptive and productive phonologic training on learners' pronunciation skills in spontaneous speech. In order to broaden the scope of the present findings, future studies should take into account spontaneous speech at both the training and testing stages through, for example, picture description tasks. Third, in the present case the posttest took place one week after training. In light of research showing that gestures aid vocabulary and grammar retention (Macedonia & Klimesh 2014; Nakatsukasa 2016) and phonological learning over time (Li et al. 2021), it is likely that a delay longer than one week between training and posttest would provide important information about the durability of the benefits of embodied prosodic training on the development of learners' pronunciation.

Our study does not disentangle the respective benefits of producing and observing the gesture in the embodied logatome group, that is, the effects of training with gesture as opposed to just observing the models an equal number of times. Despite Eskildsen and Wagner's

(2013) observation that imitating a speaker's gesture may induce and sustain understanding of the item being learned, the positive effects of gestures may stem from seeing the gesture performed by the instructor rather than from making the gesture. In that respect, there are few empirical studies directly comparing the effects of gesture perception and production. While Author (2019) did not find any difference between the perception and the production of pitch gestures for learning Mandarin tones and words, Author (2021) showed that gesture production can be more beneficial only when the learner performs the gesture correctly. Hence, further research should look at learners' gestural performance as an important factor when comparing gesture perception and production. For these reasons, it would have been interesting to add a gesture-observation group to the study and to control for individual differences in terms of gesture production accuracy.

Finally, the design of this study did not allow for any interaction with or feedback from the instructors, it was essential to strictly control potential differences between the groups. However, it is highly likely that individual feedback would have enhanced the pronunciation learning outcomes, as previous evidence suggests (Saito & Lyster 2012; Gordon et al. 2013; Lee et al. 2015). Most importantly, the role of gesture in corrective feedback may be highly relevant (Nakatsukasa 2016; Wang & Loewen 2016; Thompson & Renandya 2020).

Conclusion

he Embodied Cognition paradigm has already opened up many possibilities in the field of education, thanks in particular to the proven effects of embodiment on memory for language learning (Madan & Singhal 2012; Kiefer & Trumpp 2012; Macedonia 2019). In the field of second language learning, gestures and movements embodying actions or objects in a foreign language help learners retain new vocabulary (e.g. Quinn-Allen 1995; Tellier 2008; Macedonia & Klimesch 2014). All in all, the results of the present study confirm the predictions of the Embodied Cognition hypothesis for phonological learning and thus favor the embodying of phonological prosodic features in the teaching of pronunciation. In particular, we demonstrate the value of embodied oral reading in the development of L2 reading skills. Adding visuospatial gestures depicting prosody and probably other phonological features should be added to the toolkit of the second language teacher.

4

CHAPTER 4: EMBODYING RHYTHMIC PROPERTIES OF A FOREIGN LANGUAGE THROUGH HAND-CLAPPING HELPS CHILDREN TO BETTER PRONOUNCE WORDS

Baills, F., & Prieto, P. (2021). Embodying rhythmic properties of a foreign language through hand-clapping helps children to better pronounce words. *Language Teaching Research, First Online.* http://www.doi.org/10.1177/1362168820986716

4.1 Introduction

Many studies have shown the importance of rhythm perception in language development (Johnson & Jusczyk, 2001; Morgan & Saffran, 1995; for a review, see Gordon et al., 2015a) and language processing (Magne et al., 2007; Pitt & Samuel, 1990; Roncaglia-Denissen, Schmidt-Kassow & Kotz, 2013). The importance of rhythmic abilities for language learning has been assessed regarding various competencies, especially in children. For example, rhythmic abilities have been shown to influence children's syntactic competency (Gordon et al., 2015b). The accurate perception of language rhythmic structure has also been claimed to be crucial for phonological development and the processing of word metric structure (Goswami et al., 2002), as well as for speech intelligibility (Zion Golumbic, Poeppel & Schroeder, 2012). Further evidence shows that reading struggles in children are related to underlying difficulty in neural rhythmic entrainment, which can be detected by impaired auditory rhythm perception (Corriveau & Goswami, 2009; Goswami, 2011) and impaired musical beat perception (Goswami et al., 2013).

Below, we review the literature showing how rhythmic priming and rhythmic training can facilitate speech processing and help children improve phonological awareness (e.g. Cason et al., 2015b), as well as overcome reading difficulties (e.g. Bhide, Power & Goswami, 2013; Nelson, 2016). The present article assesses the potentially

beneficial effects of rhythmic training through hand-clapping on another area of language learning which has been less investigated, namely the learning of foreign language pronunciation by children

4.1.1 Effects of rhythmic priming on speech processing

There is growing evidence that rhythmic priming is beneficial for different aspects of language processing in adults. Falk, Lanzilotti and Schön (2017a) presented participants with sentences in French which were preceded by matching or non-matching musical rhythmic priming and observed that phase coupling, as measured by EEG (electroencephalography), was enhanced by the rhythmic auditory input when the latter was coupled with accented syllables. Their findings support the hypothesis that rhythmic cues mapping onto speech metrical structure enhance temporal expectancy and facilitate the processing of upcoming events in speech at predicted times (Falk & Dalla-Bella, 2016; Falk, Volpi-Moncorger & Dalla Bella, 2017b; Kotz & Gunter, 2015).

Other priming studies by Cason and collaborators have shown that the phonological processing of speech by adult participants is enhanced by the temporal expectancy generated by a musical rhythmic prime (Cason & Schön, 2012; Cason et al., 2015a). First, Cason and Schön (2012) presented French participants with matching and mismatching percussive rhythmic primes followed by nonwords respecting French phonotactics, and asked them to state whether a target phoneme had been pronounced in the

nonword. Behavioral measures in the form of reaction times (RTs) showed that target phonemes were detected faster when positions matched the prime beat. Additionally, when a beat expectancy violation occurred, ERP measurements (event-related potentials, also obtained by EEG) showed a larger-amplitude and longer latency response at P300. These findings were successfully reproduced in a follow-up study (Cason et al., 2015a) with spoken sentences in French preceded by a prime musical meter to induce metrical expec- tancy about both stress patterns and the number of syllables. Additionally, in this study, a group of participants underwent a short audio-motor training session several times during the experiment (just before and halfway through each block) which consisted of repeating vocally the prime rhythm using different sounds to distinguish between strong and weak musical beats. The results revealed that the priming effect was enhanced by the audio-motor training.

4.1.2 Benefits of rhythmic training

The benefits of rhythmic training on children's developing phonological and reading skills have been investigated thoroughly. For example, Bhide et al. (2013) compared the effect of a two-month rhythmic nonverbal training program to the effect of rhyme-based training software on the reading and phonological skills of 19 children aged 6 and 7 who were considered poor readers. The rhythmic training consisted of activities such as

tapping in time to a metronome, differentiating between tempos and rhythm, mimicking a rhythmic sequence, clapping or marching to a song or playing hand-clap games. The results showed that after intervention, the reading and phonological skills of participants in both training conditions improved with comparable effect sizes. Additionally, the authors found a strong correlation between children's improvement in rhythmic entrain- ment as an effect of the intervention and improvement in the overall reading score between pre- and posttest. These results suggest that interventions using purely musical rhythms may have a positive impact on reading skills. Similarly, Nelson (2016) integrated rhythmic activities into an 8-week literacy intervention program and found better results in rhyme awareness for the preschoolers that followed that program compared to those that followed regular classroom activities.

4.1.3 Rhythmic training for L2 phonological development

Second language teachers regard pronunciation as an important aspect of language to be mastered by learners in order to achieve successful communication (e.g. Nagle et al., 2018). Numerous studies on pronunciation instruction show the positive effect of overtly teaching pronunciation to foreign and second language learners (for reviews, see J. Lee, Jang & Plonsky, 2015; Saito, 2012). Most classroom pronunciation training has tended to center around segmental instruction (that is, it focuses solely on specific

speech sounds) and second language prosody is often overlooked (for a review, see Gordon & Darcy, 2016; Thomson & Derwing, 2015). However, recent work has pointed to the need for L2 prosodic instruction, as having non-target prosody in the L2 affects negatively accentedness, comprehensibility and intelligibility (Anderson- Hsieh, Johnson & Koehler, 1992; Kang, Rubin & Pickering, 2010). In this context, several studies have highlighted the importance of suprasegmental instruction for improving learners' overall fluency and comprehensibility and reducing their foreign accent (see, for example, Derwing et al., 1998; Derwing & Rossiter, 2003; Gordon et al., 2013; Behrman, 2014).

Little is known about whether rhythmic training activities can enhance phonological awareness and pronunciation in a second language teaching context. Several complementary lines of evidence lead us to think that a short rhythmic intervention can enhance second language production patterns, including pronunciation. First, various studies have demonstrated that musical aptitude, more particularly rhythmic receptive and productive abilities, are correlated with phonological abilities and pronunciation in a foreign language (e.g. Arellano & Draper, 1972; Cohrdes, Grolig & Schroeder, 2016; Gilleece, 2006; Milovanov et al., 2008; Morgan, 2004; Nardo & Reiterer, 2009; Slevc & Miyake, 2006). Second, there is evidence that rhythmic priming has immediate positive effects on the phonological production skills of hearing-impaired children speaking their first language. For

example, Cason et al. (2015b) looked at the effect of rhythmic priming on the oral accuracy of 14 hearing-impaired children with cochlear implants. In this study, children had to repeat the prime vocally and then immediately pronounce the sentence. As in the previous experiments, the primes either matched or mismatched the metrical structure of the target sentences. A comparison of the children's oral production before and after the priming session showed significantly improved pronunciation accuracy for both vowels and consonants as well as syllable and word accuracy in the matching condition only, suggesting that rhythmic priming enhances phonological production.

More research is needed on the potential positive effects of rhythmic training on L2 pronunciation. To our knowledge, only a few studies from different domains of research have been conducted, exploring the potential benefits of rap music (Fischler, 2009), a computer-based rhythm generator (Wang, Mok & Meng, 2016), rhythmic beat gestures (Gluhareva & Prieto, 2017; Kushch, 2018) and hand-clapping (Iizuka, Nakatsukasa & Braver, 2020; Zhang, Baills & Prieto, 2018) on L2 pronunciation, with mixed results. During a four-week intensive course, Fischler (2009) taught sentence and word stress in English to six advanced adolescent learners with different L1 backgrounds through activities related to rhythm and rap music. A qualitative analysis of the number of errors in stress placement and of intelligibility during reading and narrative-picture tasks before and after training showed a general

improvement for the reading task only. However, in the absence of a control group, the author could not claim that the participants benefited specifically from the training method. Following a different approach, Wang et al. (2016) tested the effect of a computer application that automatically generated a percussive rhythm on the pronunciation of sentences by 20 Chinese learners of English. Participants were asked to pronounce 15 English sentences before and after the rhythmic cue. Only those who obtained the lowest scores in terms of native-likeness before the rhythmic priming significantly improved their pronunciation. Adopting another approach, Gluhareva and Prieto (2017) tested whether the observation of rhythmic beat gestures, simple up-and-down or back-and-forth hand movements naturally coordinated with the prominent parts of speech, was beneficial for the pronunciation of English sentences by Catalan intermediate learners during a short training session. The results pointed to a positive effect of rhythmic beat training on elicited semi-spontaneous speech in terms of accentedness reduction.

The facilitating effect on the pronunciation of words by marking syllables by hand-clapping, an activity that lends itself very easily to the classroom context, has been investigated only recently in two studies, with mixed results. First, a study by Zhang et al. (2018) in which, during a short audiovisual training session, two groups of 25 Chinese adolescents repeated unknown French words while either clapping out their rhythmic structure or not. Accentedness

ratings of participants' oral production before and after training showed only a near-significant difference in improvement between the two groups. However, acoustic analysis of final rhyme duration indicated that participants in the clapping group lengthened the final syllable more appropriately than did participants who were not trained to clap, indicating that hand-clapping helped participants acquire the rhythmic structure of the words. However, in this study, participants had to learn the meaning and pronunciation of words at the same time, rendering it not possible to determine whether the effects of clapping on pronunciation might not have been negatively impacted by cognitive overloading. Second, Iizuka et al. (2020) assessed the effect of watching and performing hand-clapping of Japanese moras on the percep- tion and pronunciation of long vowels, geminates and moraic nasals presented in loan- words by adult English native speakers and found a significant benefit of hand-clapping for the perception of these segmental features in a delayed posttest. However in this study, despite reducing the cognitive load of meaning retrieval by using loanwords, the results of the production task failed to show a superior effect of repeating words with hand-clapping compared to repeating speech only. Overall, given the mixed results obtained in the literature, further research is needed to empirically test the effects of hand-clapping on L2 pronunciation. In addition, to our knowledge no previous study has assessed the role of a short

hand-clapping training session on L2 pronunciation patterns in children.

4.2 The present study

The aim of the present study was to assess whether a short training session using hand- clapping to highlight the rhythmic structure of words can improve the pronunciation of newly learned cognate words in French. The 28 Catalan-speaking children who participated in the training session had no prior knowledge of French. Crucially, the 20 items chosen for the training session were French-Catalan 'cognates', that is, words with identical meanings and similar forms, like avion / avió 'airplane'. This was done deliberately on the grounds that it would facilitate word recall (de Groot & Keijzer, 2000) and allow participants to focus exclusively on pronunciation rather than word meaning, thus avoiding the potential cognitive overload present in the study by Zhang et al. (2018) noted above. Importantly, while the transparency of meaning offered by cognates can facilitate comprehension and memorization, the similarity in phonological forms may enhance phonological transfer from their L1, thus penalizing pronunciation (Flege, 1987).

Catalan is considered a stress-accented language in which lexically stressed syllables generally serve as the main landing site for phrasal pitch accents. Word stress is realized on one of the last

three syllables of the morphological word, and at the phrasal level, the last content word in the intonational phrase receives the main phrasal stress (Prieto et al., 2015). Unlike Catalan, French has no lexical stress and is considered to be an edge- prominence language. Stress is assigned at the phrasal level, as follows: (1) an obligatory phrase-final primary stress is generally assigned to the last metrical syllable of a content word and has a demarcative function which marks the right edge of a prosodic phrase; and (2) an optional secondary stress can be assigned phrase-initially (see, among others, Di Cristo & Hirst, 1993; Jun & Fougeron, 1995, 2000; Delais-Roussarie et al., 2015).

In Catalan and in French, as in many other Romance Languages, the nuclear accent falls at the end of the sentence (Nuclear Stress Rule: Halle & Vergnaud, 1987; see Frota & Prieto, 2015). However, regarding the phonetic properties of stress realization, there seems to be a basic difference between the two languages which affects the duration of the stressed syllable. In comparison with Catalan, French stress is realized by a more extreme lengthening of the stressed phrase-final syllable and more particularly of the full vowel (Astésano, 2001; Delattre, 1966; Di Cristo & Hirst, 1993; Fletcher, 1991; Vaissière, 1991). To our knowledge, although thus far no study has systematically analysed cross- linguistic differences in final lengthening between Catalan and French, two types of acoustic evidence point to a difference in final lengthening patterns between these two languages. First,

acoustic comparisons between French and Spanish (a language with a rhythmic structure and durational patterns that are similar to Catalan; see Prieto et al., 2012) tend to point to more exaggerated final lengthening patterns in French. While Rao (2010) found that final lengthening patterns before a pause in three different varieties of Spanish may reach an average of 30%, studies scrutinizing final lengthening in French (Bartkova et al., 2012; Zellner, 1996) have found as much as a 50% increase in syllable-final durations. However, in these studies, syllable structure was not controlled for, the speakers' samples for the analysis were small, and, importantly, final lengthening was calculated for paroxytone words only. In a recent study with a large dataset (15 hrs of speech), Gendrot, Adda-Decker & Santiago (2019) compared the duration of final vowels produced at the right edges of Intonation Phrases in both French and Spanish. The results showed that, in oxytonic positions, French vowels tend to be longer than Spanish vowels. Second, for the purpose of this study, an exploratory acoustic analysis was carried out which compared the duration of stressed and unstressed syllables in 20 pairs of cognate words in Catalan and French (i.e. *balcó – balcon* 'balcony'). The Catalan words (N = 20) were pronounced by two 8-year-old speakers of Catalan and the French words (N = 20) were pronounced by two 8-year-old speakers of French (see Table 1). The mean ratio between stressed and unstressed syllables was calculated for each language and for the three stress positions (e.g. oxytonic, paroxytonic and

proparoxytonic positions) in Catalan. We found that in French, accented syllables were 1.91 times longer than the preceding unstressed syllable, whereas in Catalan, accented syllables were 1.75 times longer than the preceding unstressed syllable for paroxytones, 1.28 times longer for paroxytones and 1.65 times longer than the following syllable for proparoxytones (for the description of the procedure, see Figure 1; see also Appendix A in supplemental material).

Table 1

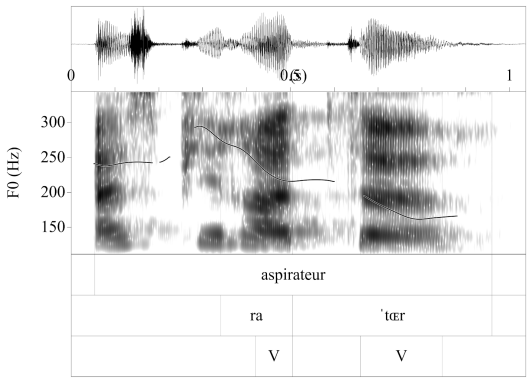
Target French words for the training session and their Catalan cognates

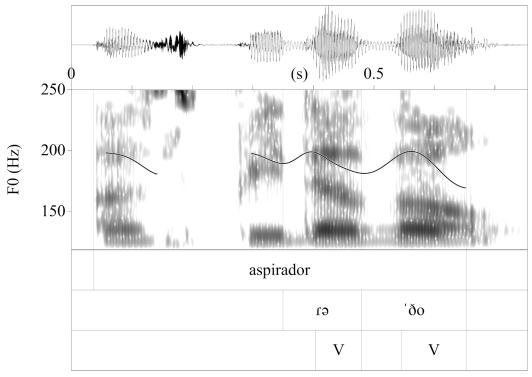
French	Catalan: oxytone	English gloss	French	Catalan: paroxytone + proparoxytone	English gloss
balcon [balˈkɔ̃]	*balcó* [bəɬˈko]	'balcony'	*oreille* [ɔˈʁɛj]	*orella* [uˈɾɛʎə]	'ear'
tambour [tɑ̃ˈbuʁ]	*tambor* [təmˈbo]	'drum'	*famille* [faˈmij]	*família* [fəˈmiljə]	'family'
purée [pyˈʁe]	*puré* [puˈɾe]	'puree'	*musique* [myˈzik]	*música* [ˈmuzikə]	'music'
avion [aˈvjɔ̃]	*avió* [əˈβio]	'airplane'	*pizza* [piˈdza]	*pizza* [ˈpid͡zə]	'pizza'
crocodile [kʁɔkɔˈdil]	*cocodril* [kukuˈðɾil]	'crocodile'	*confiture* [kɔ̃fiˈtyʁ]	*confitura* [kuɱfiˈtuɾə]	'jam'
biberon [bibˈʁɔ̃]	*biberó* [biβəˈɾo]	'baby bottle'	*spaghettis* [spageˈti]	*espaguetis* [əspəˈɣetis]	'spaghetti'
céréales [seʁeˈal]	*cereals* [sərəˈals]	'cereals'	*éléphant* [eleˈfɑ̃]	*elefant* [ələˈfan]	'elephant'
aspirateur [aspiʁaˈtœʁ]	*aspirador* [əspirəˈðo]	'vacuum cleaner'	*mandarine* [mɑ̃daˈʁin]	*mandarina* [məndəˈɾinə]	'tangerine'
télévision [televiˈzjɔ̃]	*televisió* [tələβiˈzjo]	'television'	*ambulance* [ɑ̃byˈlɑ̃s]	*ambulància* [əmbuˈɬansjə]	'ambulance'
ordinateur [ɔʁdina ˈtœʁ]	*ordinador* [uɾdinəˈðo]	'computer'	*hélicoptère* [elikɔpˈtɛʁ]	*helicòpter* [əliˈkɔptəɾ]	'helicopter'

For the purposes of the present study, the realization of the more extreme lengthening patterns in the final stressed syllable may be of crucial importance for the production and perception of French prosodic phrasing for L2 learners. Schwab (2012) analysed the production of adult intermediate Spanish learners of French and found some evidence that they transferred Spanish stress realization when speaking French. In a follow-up study, Schwab (2013) showed that adult intermediate Spanish speakers were able to produce the intended stressed syllable at the right edge of the accentual phrase, marking the stressed syllable by means of variations in duration and F0. These results were confirmed by Santiago and Mariano's (2019) study analysing a corpus of adult intermediate Spanish learners of French. However, additional empirical studies are needed to examine the exact phonetic realization of French final lengthening by L2 learners. Learners with less knowledge of the target language or exposed to less input may benefit from specific training to speed up the acquisition of a more extreme durational production of word- final syllables in French. In addition, a series of recent studies have shown that L2 learners' general pronunciation may be improved by training them in the production of rhythmic prosodic features (e.g. Gluhareva & Prieto, 2017; Li, Baills & Prieto, 2020; Yuan et al., 2019), corroborating the idea that suprasegmental features count as a major factor in measures of accentedness and perception of oral proficiency (see, for example, Kang et al., 2010).

Figure 1

Acoustic representation of the French–Catalan cognate pair of words aspirateur / aspirador 'vacuum cleaner'. The comparison of vowel duration measures show that while the French stressed syllable [tŒʁ] in aspirateur is 2.22 times longer than the preceding syllable [ʁa] (upper panel), the Catalan stressed syllable [ðo] is 1.4 times longer than the preceding syllable [ɾə] (lower panel).





The rhythmic training proposed in this study consisted of audio-visually highlighting the rhythmic structure of words through hand-clapping. We hypothesized that hand-clapping can serve to acoustically and visually highlight the prosodic patterns of speech. Acoustically, the clapping sound will auditorily highlight the syllabic structure of the target words. Visually, the fact that the hands stay longer together on the stressed syllable calls attention to the longer duration of this syllable. Embodying these prosodic patterns might reinforce the phonological learning process by increasing phonological awareness, which can ultimately lead to better pronunciation as measured by accentedness ratings and acoustic analysis. We surmise that the effect of the training might be detectable through an acoustic analysis that can assess a more target-like duration of the stressed syllable by L2 learners.

4.2.1 Hand-clapping as prosodic embodiment

Hand-clapping is intrinsically related to the concept of rhythm and, as such, falls simultaneously within the two domains of music and language. It is present throughout infancy in the form of hand-clapping games and songs, and can be observed in numerous cultures (Cameron & Grahn, 2014; see also Romero Naranjo, 2013). Clapping one's hands, like tapping one's foot or dancing to musical rhythms, is a natural way to express the temporal structure of music with body movements (Repp & Su, 2013). There are reasons to believe that the reinforcement by means of a motor

action (e.g. clapping) can be helpful for L2 learners to better process and produce a prosodic feature of speech (in this case the rhythmic structure of words) that may be difficult for them to acquire.

According to the theory of grounded cognition, the activation of appropriate perceptual and motor interactions during learning should enhance the development of cognitive functions (Borghi & Caruana, 2015). Indeed, neuroscientific studies have shown that not only perception but also motor brain networks are activated when participants engage in different tasks involving abilities such as memory, knowledge, language and thought (for a review, see Barsalou, 2008). Embodied theories of language processing suggest that motor action and semantic processing are closely interrelated (see, e.g. Glenberg & Kaschak, 2003; Zwaan & Taylor, 2006) and that the execution of motor actions has a selective effect on the linguistic processing of words (Rueschemeyer et al., 2010). In the field of gesture studies, research has shown that learners achieve better results in different memory and cognitive tasks when producing hand gestures than when merely observing them (Goldin-Meadow, 2014; Goldin-Meadow, Cook & Mitchell, 2009; for a review of the effects of enactment and gestures on memory recall, see Madan & Singhal, 2012). There is also evidence from neurophysiological research that self-performing a gesture when learning verbal information favors the formation of sensorimotor networks that contribute to the representation and storage of words

in a native language (Masumoto et al., 2006) as well as foreign language (Macedonia, Müller & Friederici, 2011).

The implications of the benefits of embodiment are crucial for education (for reviews, see Kiefer & Trumpp, 2012; Wellsby & Pexman, 2014). Embodied approaches to music pedagogy are a good illustration of how body movements facilitate the understanding and enhance the retention of complex musical concepts (Juntunen, 2016). In the field of the acquisition of L2 phonological patterns, a recent series of studies on pitch gestures also show the benefits for word recall and the perception of pitch information of watching and producing up-and-down hand movements that represent rising and falling pitch (Baills et al., 2019; Kelly, Bailey & Hirata, 2017; Morett & Chang, 2015). Such prosodic hand gestures also seem to facilitate the production of difficult pitch contours in a foreign language by tonal language speakers (Yuan et al., 2019).

4.2.2 Individual differences in pronunciation learning

Since the main goal of the present study was to assess the role of hand-clapping in second language pronunciation learning, three types of individual measures related to working memory, speech imitation skills and musical abilities were taken into account, as they have been shown to play an important role in L2 learners' pronunciation. First, phonological working memory has attracted attention as a contributing factor to pronunciation talent

(Aliaga-Garcia, Mora & Cerviño-Povedano, 2011; Darcy, Park & Yang, 2015; Rota & Reiterer, 2009). Second, speech imitation talent and pronunciation skills in a foreign language have been shown to be highly interdependent in research by Nardo and Reiterer (2009). Reiterer et al. (2013) also found that speech-motor flexibility may be among the best predictors of speech imitation capacities, leading to better pronunciation in an unknown language. Speech imitation abilities would then be of major importance when assessing learners' pronunciation. Finally, some studies have shown that musical abilities are related not only to receptive but also to productive phonological learning skills in an L2 (Delogu & Zheng, 2020; Milovanov et al., 2008, 2010; Slevc & Miyake, 2006).

4.3 Methods

In a between-participants training study with a pre- and posttest design, a group of 28 Catalan children were asked to learn 20 new cognate French words under one of two audiovisual conditions: (1) training which involved observing and replicating the behavior of a native speaker simultaneously saying a word and clapping to highlight the pro- sodic structure of the word; or (2) training which involved observing and replicating a native speaker who merely spoke the word without clapping. We hypothesized that observing and subsequently performing hand-clapping would lead to a greater improvement in pronunciation of the French words both in terms of perceived accentedness ratings and in terms of acoustic patterns (e.g. a more native-like lengthened production of the words' final rhyme and final vowel).

4.3.1 Participants

Twenty-eight 7- to 8-year-old children from the city of Girona, Catalonia, took part in the experiment at their school premises after their parents signed a written consent form. They were all Catalan-Spanish bilinguals with Catalan as their dominant language (percentage of time Catalan used in their daily life: M = 87%, as reported by participants' caregivers). None of them had any prior knowledge of French. They were informed that they would learn words in French and were randomly divided into two groups, namely the clapping group (n = 14, M_{age} = 7.43; SD = 0.5, 7

females) and the non-clapping group (n = 14, M_{age} = 7.29; SD = 0.46, range 7 females).

4.3.2 Materials

a) Training session

Materials for the training session consisted of two 10-minute videos prepared at the professional broadcasting studio of the Universitat Pompeu Fabra in Barcelona. For both conditions, the videos were designed to teach the 20 target French words with two instructors (see Table 1).

The rationale for the selection of words was that (1) their meaning should be transparent to Catalan speakers (e.g. Catalan *avió* / French *avion* 'plane', Catalan *ordinador* / French *ordinateur* 'computer'); and (2) they should name objects that would be easy to represent by means of a simple black and white line drawing in order to avoid any written input.1 A variety of consonantal environments were proposed, mainly constrained by the obligation to work with cognates in the two languages and to maintain the number of syllables constant.

Crucially, though the target French words were all cognates, they included a variety of sounds in the target language that are not part of the Catalan sound inventory, such as the labiodental [v] and the uvular rhotic [ʁ] for consonant sounds, as well as the rounded front vowels [y] and [œ] and nasal vowels [ɑ̃] and [ɔ̃]. Importantly, apart

from these segmental differences, a salient phonological feature of all the French words as compared to their corresponding Catalan cognates was the presence of a phonetically strong lengthened phrasal-final stress in French, which would compete with lexical stress in Catalan. For ten of the French items, stress was located in the same position as in their Catalan oxytone cognates (balcon – balcó). For the ten remaining items, the Catalan counterparts were nine paroxytones (oreille – orella) and one proparoxytone (musique – música). In these cases, the stress was either on the same syllable (6 items) or on a different syllable (4 items) (see Table 1). The words included two-syllable words (8 items, 4 oxytones, 3 paroxytones, 1 proparoxytone), three-syllable words (8 items, 3 oxytones, 5 paroxytones) and four-syllable words (3 oxytones, 1 paroxytone).

The two video stimuli (the non-clapping video and the clapping video) were prepared as follows. Two female native French speakers were video-recorded when producing all 20 target words in Table 1 as if speaking to a class of learners in a very clear manner. A total of 80 videos were recorded in this fashion (20 words × 2 instructors × 2 conditions).

For the clapping stimuli, instructors first spoke one word without moving their hands, and immediately repeated the same word while simultaneously clapping once on each of the target syllables of the word and then returning their hands to their rest position. For each

target word, each syllable was marked by a regular hand-clapping sound, highlighting its syllabic structure. Visually, the duration of the hand-claps highlighted the prosodic prominence patterns, with the syllables preceding the last stressed syllable not bearing any prosodic emphasis. A frame-by-frame analysis of the 20 clapping videos showed that the hands remained in contact longer in the last syllable before returning to the rest position (*M* = .499 sec, *SD* = .249, 95% CI [.419, .578] for the final syllable, *M* = .069 sec, *SD* = 0.026, 95% CI [0.062, 0.074] for the other syllables), thus visually highlighting the longer duration of the final syllable. We asked the instructors to avoid using a higher clapping intensity (volume) on the stressed syllable for two reasons: first, because it might interfere with perception of the speech signal; and second, because stressed syllables in French are not characterized by higher intensity, and louder clapping might have cued participants to use intensity instead of duration to mark stress. Intensity was further measured at each hand-clap in the 40 target stimuli to ensure equivalent levels (*M* = 78.053 dB, *SD* = 3.225, 95% CI [76.802, 79.304] for the final syllables, *M* = 75.364 dB, *SD* = 6.521, 95% CI [73.470, 77.257] for the other syllables). We also checked that the sound produced by the clapping at no time masked the voice of the instructor.

For the non-clapping stimuli, the two instructors spoke the words twice without moving their hands. No pause was produced between syllables. Moreover, whether they were accompanying their speech

with claps or not, the two instructors were asked to use natu- ral head movements and facial expressions while they spoke but refrain from emphasizing stressed syllables by head, eyebrow or chin movements.

In order to check that speech rate and word duration did not differ between the spoken and hand-clapped words, the first author of this study extracted the soundtracks of the 80 items in the training videos produced by the two instructors for the two conditions and labeled those items in Praat (Boersma & Weenink, 2017). The values for speech rate and word duration were automatically extracted by using Praat scripts (de Jong & Wempe, 2009; Elvira-Garcia, 20142). Potentially significant differences between the speech rate and syllable duration patterns across the two conditions (clapping vs. non-clapping) were tested by means of two independent sample t-tests. The speech rate of the instructors' production of the target items in the clapping condition (M = 1.96 sec, SD = .44) compared to the speech rate of the instructors for the non-clapping condition, M = 1.90 sec, SD = .42) did not differ significantly ($t(65) = -.602$, p = .549, 95% CI [– .271, – .145]. Similarly, no significant differences were found for word duration, $t(65) = .888$, p = .378, 95% CI [– .075, .197], between the clapping condition (M = 1.42, SD = .28, 95% CI [1.32,1.52]) and the non-clapping condition (M = 1.48 sec, SD = .27, 95% CI [1.38, 1.58]).

Each of these 80 video recordings was embedded between two still sequences. The first, which lasted for 3 seconds, showed a black and white drawing illustrating the target French word about to be spoken. The still sequence following the stimuli lasted for 5 seconds and showed a black screen with no image (see Figure 2). This 5-second blank screen was intended to give the viewer time to replicate what they had seen and heard, depending on the group to which they had been assigned. If they had been assigned to the clapping group, they would repeat the word and clap as they had seen it done (Figure 2, top panel). If they had been assigned to the non-clapping condition, they would merely repeat the word as they had heard it spoken (Figure 2, bottom panel).

In order to balance the presence of the two speakers, for each condition, two blocks were created with a total of 20 sequences/items with the two speakers appearing an equal number of times, making sure that each consecutive item was produced by a different instructor. All in all, the 20 target words were trained twice, each time with a different instructor. To ensure variability in order of presentation of the stimuli across participants, six videos with different orders of presentation of the target items were created for each condition.

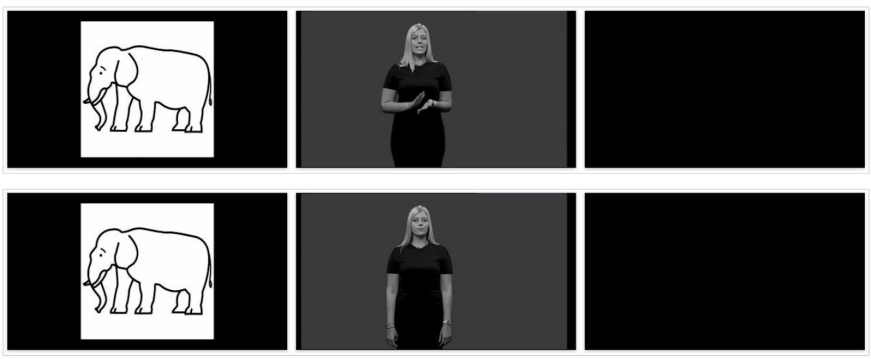
b) Pre- and posttest materials

To test participants' pronunciation and to avoid the influ- ence of written input, a word repetition task was created. The materials

required for the pre- and posttest word repetition task consisted of 15 French words which the participants would hear and then repeat. Ten of these words were words related to the training session (*balcon, tambour, musique, purée, ambulance, crocodile, biberon, mandarine, confiture, aspirateur*; see Table 1) and merely consisted of the audio tracks from the video recordings by the two instructors. The other five words were completely new (*calendrier* 'calendar', *garderie* 'kindergarten', *sportive* 'sporty', *imprimante* 'printer', *cheveu* 'hair', *râteau* 'rake') and were not cognate words. In this case new audio recordings were made by one of the two native speakers that featured in the videos.

Figure 2

Stills from the training video for the word éléphant 'elephant' in the clapping condition (top panel) and non-clapping condition (bottom panel).



Sixteen different orders of presentation of the 15 words were created using a presenta- tion software and participants were

assigned to different combinations of these orders at pre- and posttest. The audio files were directly embedded in the presentation software.

c) Individual measures

As noted above, in order to control for the effect of individual differences related to cognitive, linguistic and phonological abilities, the following five tasks were administered:

1. Short-term memory task: Short-term memory was assessed through a memory span task where participants had to repeat different lists of Catalan words (see Appendix B in supplemental material) ranging from three to six words (Bunting, Cowan & Saults, 2006).

2. Imitation talent task: Participants' imitation abilities were tested through a word repetition task involving 12 words in six different languages. The items were 2- or 3-syllable words containing segmental information that can be considered difficult for Catalan speakers and which are not part of the consonantal and vocalic inventory of the Catalan language (see Appendix C in supplemental material).

3. Phonological perceptual ability task: Participants undertook a standard phonological discrimination task for children whereby they had to listen to pairs of French nonwords and decide if they were the same or different (Macchi et al., 2013).

4. Rhythmic perceptual ability tasks: Participants undertook two standard discrimi- nation tests for musical rhythm (8 items) and musical accent (10 items) extracted from a free musical perception test called PROMS that can be tailored for children in terms of the number and difficulty of items (Law & Zentner, 2012). The procedure followed for both tests was the same, namely the children listened to one sequence twice and then to a last sequence which they had to qualify as same, different, or unsure. While the rhythm subtest consists of discriminating among simple patterns of quarter notes, eight notes, and sixteenth notes, the musical accent subtest assesses the ability to distinguish the relative emphasis given to certain notes in a rhythmic pattern. As such, it is related to the concepts of meter in music and stress in speech (for a detailed description of the subtests, see Law and Zentner, 2012).

5. Rhythmic production ability task: In this hand-clapping replication test, participants heard a rhythm sample and immediately had to replicate the rhythm by clapping (six samples, all 4/4 time, 2 measures).

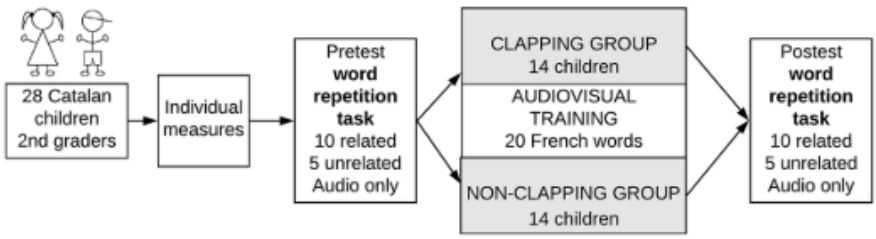
4.3.3 Procedure

The experimental procedure, which consisted of pretest – training session – posttest, lasted 20 minutes (see Figure 3). The experiment was carried out with each child individually in a quiet room at their school. The child was seated in front of a tablet computer and wore a comfortable high quality headset equipped with a high quality

microphone. The first author of the study remained in the room with the child to ensure that the training session was completed successfully. As noted above, prior to the training as such, each of the 28 participating children took a battery of five tests to measure their cognitive, linguistic and phonological abilities in a separate session that lasted around 40 minutes. A research assistant helped the first author to collect the individual measures.

Figure 3

Experimental procedure



The pretest consisted of a short word repetition task. Children were asked to touch a key to play audio recordings of 15 French words, one at a time, repeating each word before moving on to the next. They were thus able to set their own pace. The duration of the pretest was roughly 5 minutes. The participant's speech was recorded throughout the pretest. After the pretest, participants watched one of the two training videos (clapping or non-clapping) depending on which group the child had been randomly assigned to. Children assigned to the clapping group were randomly exposed

to one of the six variants of the clapping stimulus video. For each trained word in the clapping condition, children first saw the drawing depicting the word, then watched the video clip of the instructor producing the word while clapping her hands, and finally (as they viewed the empty black screen) repeated the word while also clapping their hands. By contrast, children in the non-clapping group saw the drawing, watched the instructor producing the word without clapping, and simply repeated the word. As the stimulus video con- sisted of two blocks, each child was exposed to each item twice. The total duration of the training session was roughly 10 minutes. When the training session finished, children performed the posttest word repetition task, which was identical to the pretest task. Their verbal output was likewise recorded during the posttest. The full procedure, including the prior individual measures testing, pretest, training session and posttest, lasted a total of approximately 60 minutes.

4.3.4 Data coding

The collected data underwent two types of analyses, namely (1) perceived accentedness as judged by three native French speakers listening to the recordings, and (2) acoustic measures of the final rhyme duration and final vowel duration. The results of the five tasks designed to collect individual measures (short-term memory, imitation ability, phonological perception, rhythmic perception and production) were coded and added to the database (see below).

a) Perceived accentedness ratings

The 840 audio files of the children's oral productions during pre- and posttest (15 words × 2 tests × 28 children) were rated by three non-linguist French native speakers, who were unaware of the purpose of the experiment. To avoid fatigue effects, the audios were split into four different blocks and uploaded in four different rating surveys. The raters took between 45 and 60 minutes to complete each block. They were asked to complete the task within four consecutive days. For each word, the raters listened to the original audio prompt that featured in the test and then heard the children's productions in random pretest/posttest pairs. In other words, they did not know whether a particular item came from the pretest or the posttest. The raters were asked to compare the children's productions with the original auditory stimuli and evaluate the general accentedness of the target words on a scale from 1 'not accented' to 7 'extremely accented', that is, the degree to which their pronunciation approximated the native model (Munro, Derwing & Morton, 2006). We preferred to use an accentedness measure over a comprehensibility measure because it has been shown that accentedness scores by native listeners are more closely associated with target pronunciation features (e.g. vowels, consonants, stress errors) than comprehensibility scores, which have been found to be associated with non-phonological variables like lexis or grammar (for a review, see Saito, Trofimovich & Isaacs, 2017).

Inter-rater reliability was assessed using IBM SPSS Statistics 23 by calculating the Intraclass Correlation Coefficient based on the scores given by the three raters for each item. The results pointed to a high degree of reliability (ICC = .97, $F(839, 1678) = 99.87$, $p < .001$, 95% CI [0.95, 0.97]).

b) Acoustic analysis

Since training only had an effect on the pronunciation of trained items, as measured by accentedness ratings (see Section IV), the acoustic analysis was carried out with the trained items only, for a total of 560 audio files (10 words × 2 tests × 28 children). In order to analyse the duration patterns of the target words, word boundaries, word-final rhyme boundaries and word-final syllable boundaries of the children's oral productions at pre- and posttest were manually annotated in Praat by the first author, following Machač and Skarnitzl's guidelines (2009). Absolute duration measures were then extracted with an automatic script (Dan McCloy, original version by Mietta Lennes). This process yielded, in seconds (sec), word duration, word-final rhyme duration and word-final vowel duration. Using this data, the ratio obtained by dividing the word-final rhyme duration and word duration, and the ratio obtained by dividing the word-final vowel duration and the word duration (as a %) were calculated in order to control for speakers' speech rate differences and for differences related to word duration.

c) Individual differences

The memory span score corresponded to the number of words participants remembered in at least three lists with the same number of words, regardless of the order in which the words were recalled.

• Imitation talent: The first author, a phonetician, rated participants' oral production by comparing them to the native pronunciation on a scale between 1 ('very close to target pronunciation') and 7 ('very different from target pronunciation'). For each item, the rating consisted of listening to the word pronounced by the native speaker first and then immediately to the same word pronounced by the partici- pant. The rater compared how close to the target the sounds were produced.

• Phonological perceptual ability: The score for this task corresponded to the num- ber of correct answers, with a maximum of 36 points.

• Rhythmic perceptual ability: The final score was automatically calculated from the online software for the two subtests.

• Rhythmic production ability: Each sequence that was accurately replicated in terms of number of beats and rhythmic pattern was coded by the first author as 1. Inaccurate replications scored 0.

4.3.5 Statistical analysis

All statistical analyses were run using IBM SPSS Statistics 23. For each model, the tests of significance were two-tailed with an alpha level of .05, and post hoc comparisons were adjusted with the Bonferroni correction. A participant-sorted database was created displaying individual measure scores per participant, their mean accentedness rating at pre- and posttest, and their mean duration ratios for final rhyme and final vowel at pre- and posttest. The individual measures for each participant were used to test for (1) potential differences between the between-participant groups, (2) potential effects of individual differences on perceived accentedness scores and (3) potential effects of individual differences on acoustic measures.

First, in order to test for homogeneity between the two groups, we ran an independent sample t-test with individual measures (age, short-term memory, imitation, phonological discrimination, rhythm perception, rhythm production) as tested variables and group (two levels: clapping vs. non-clapping) as the grouping factor. Then, to explore the potential effects of individual differences on our results, three stepwise multiple regression analyses were run with mean accentedness score, rhyme duration ratio and vowel duration ratio as the dependent variables and the individual measures short-term memory, imitation, phonological discrimination, rhythm perception and rhythm production as fixed factors. Consequently, the

individual differences that were found to have an effect on the accentedness results were added as covariates to the models testing the effect of training on accentedness and final lengthening (see sections below).

An item-sorted database was created, displaying the three raters' accentedness scores, the duration ratio for the final rhyme and the duration ratio for the final vowel for each of the 15 items at pre- and posttest for each participant.4 For each item it was also indicated if the word was one of the 10 words which appeared in the training sequence (which we labeled 'trained') or was one of the five that were not ('new'), and the number of syllables and the stress position were specified.

To analyse the effect of the type of training (clapping vs. non-clapping) on the accent- edness ratings of participants' pronunciation, a generalized linear mixed model (GLMM) was run with accentedness as the dependent variable. Training group (two levels: clapping vs. non-clapping), session (two levels: pretest and posttest), Training group × Session, familiarity (two levels: trained vs. new), number of syllables (three levels: two, three, or four), stress position (three levels: oxytone, paroxytone or proparoxytone) and the interactions Training group × Familiarity, Session × Familiarity, Training group × Session × Familiarity were set as fixed factors. One random effects block was specified with the variables participant and item, and the covariates imitation and

phonological discrimination (the ones found to have a significant effect) were added to the model.

To analyse the effect of the type of training (clapping vs. non-clapping) on the acoustic measures assessing final lengthening, two GLMMs were run with the dependent vari- ables word-final rhyme duration ratio and word-final vowel duration ratio. Fixed and random factors were the same as for the GLMM for accentedness ratings. In addition, the interactions Training group × N of syllables, Session × N of syllables, Training group × Session × N of syllables were set as fixed factors. Rhythm perception and phonological discrimination were added as covariates.

4.4 Results

4.4.1 Differences between groups

An independent sample t-test indicated that there was no significant difference between groups for any of the five individual measures short-term memory, imitation, phonologi- cal discrimination, rhythm perception and rhythm production. These results confirmed that children in the two groups were equally distributed in terms of individual aptitudes (see Table 2).

Table 2

Participants' scores on individual measures (means, SDs and confidence intervals per condition)

	Non-clapping group				Clapping group			
	M	*SD*	95% Confidence Interval		*M*	*SD*	95% Confidence Interval	
Age	7.43	0.50	7.23	7.62	7.29	0.46	7.11	7.46
Short-term memory	4.36	0.83	4.04	4.68	4.29	0.60	4.05	4.52
Imitation	4.36	1.42	3.81	4.91	4.36	1,37	3.83	4.89
Phonological discrimination	23.14	4.35	21.46	24.83	22.50	2.27	21.62	23.38
Rhythm perception	17.43	3.41	17.95	20.34	19.14	3.08	17.95	20.34
Rhythm production	4.46	2.21	3.61	5.32	4.43	1.41	3.88	4.98

4.4.2 Effects of individual differences on accentedness and acoustic measures

Results of the multiple regression analysis revealed that imitation and phonological discrimination abilities were significant predictors of participants' accentedness scores. Imitation and phonological discrimination scores explained 38.5% of the variance (R^2 = .38; F(2, 7,830) = 1.71, p = .002). Mean accentedness decreased .15 points for each point of improvement in the imitation test (β = − .46, p = .007, 95% CI [–0.25, –0.04]) and fell .05 points for each additional point in the phonological discrimination test (β = − .39, p = .019, 95% CI [– .83, – .01]). Consequently, as explained above, these two variables were added as covariates to the model testing the effect of training on perceived accentedness.

The results of the multiple regression analyses revealed that rhythm perception and phonological discrimination abilities were significant predictors of participants' lengthening measures. For the final-rhyme duration ratio, rhythm perception and phonological discrimination explained 1.4% of the variance (R^2 = .01; F(2,550) = 3.86, p = .022). The final-rhyme duration ratio increased 0.1 points for each point of improvement in the rhythm perception test (β = .26, p = .031, 95% CI [.02, .50]) and increased 0.1 points for each additional point in the phonological discrimination test (β = .40, p = .02, 95% CI [.06, .73]). For the word-final vowel duration ratio,

rhythm perception explained 0.7% of the variance (R^2 = .007; F(1,551) = 4.92, p = .027). The final-vowel duration ratio increased 0.09 points for each point of improvement in the rhythm perception test (β = .20, p = .027, 95% CI [.02, .38]). Consequently, as explained above, these variables were added as covariates to the models testing the effect of training on the two acoustic measures.

4.4.3 Effects of type of training and item familiarity on perceived accentedness

Table 3 and Figure 4 show the mean accentedness scores at pre- and posttest for the non-clapping and clapping groups and for trained and new items. Results of the GLMM showed significant main effects of session, F(1, 2,505) = 54.69, p < .001), and Training group × Session, F(1, 2,505) = 10.51, p = .001, on accentedness scores. Post-hoc analyses revealed a significant effect of session for both groups, meaning that there was a significant decrease in perceived accentedness between pretest and posttest scores in both the non-clapping group, F(1, 2,505) = 8.61, p = .003) and the clapping group, F(1, 2,505) = 56.64, p < .001). The contrast estimates indicated a larger effect size in the clapping group (β = 0.54, p < .001) than in the non-clapping group (β = 0.21, p = .003). A significant difference between the clapping group and the non-clapping group was found at pretest, with significantly lower accentedness scores for the non-clapping group (F(1, 2,505) = 4.56, p = .033). However, no difference between groups was

found at posttest. In other words, the non-clapping group were significantly better rated for accentedness at pretest than the clapping group; however, their improvement, although significant, did not reach the size of the improvement experienced in the clapping group.

Table 3

Mean accentedness ratings for the word imitation task across groups (clapping, non-clapping), sessions (pretest, posttest) and familiarity (trained, new)

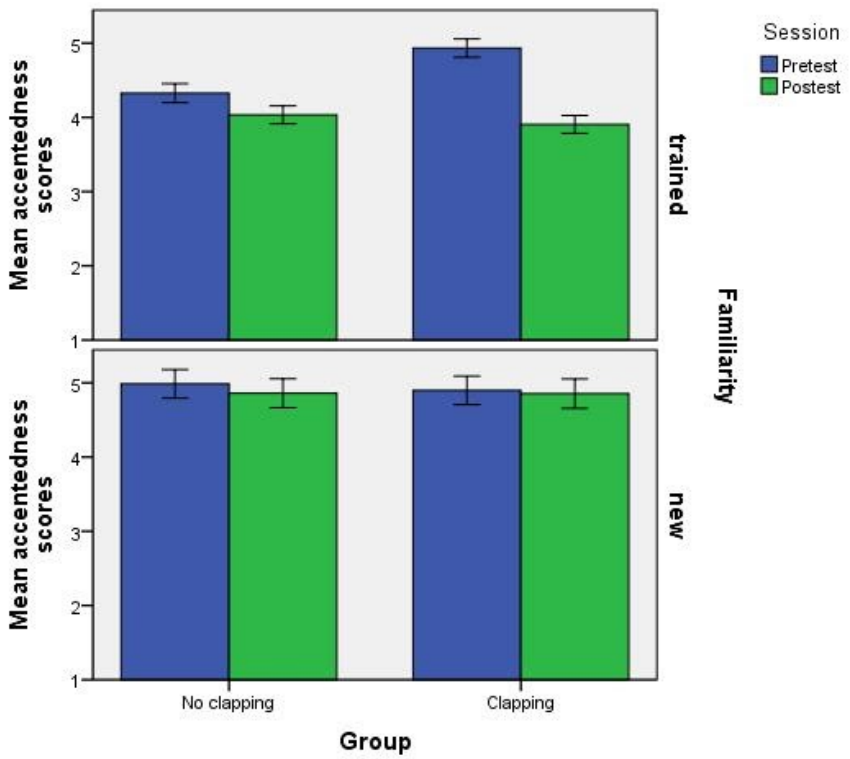
Group	Session	Familiarity	*M*	*SD*	*SE*	95% Confidence interval	
						Inf.	Sup.
Non-clapping	Pretest	trained	4.28	1.38	0.15	3.98	4.58
		new	4.94	1.31	0.17	4.61	5.26
	Posttest	trained	3.99	1.40	0.15	3.68	4.29
		new	4.81	1.24	0.17	4.48	5.14
Clapping	Pretest	trained	4.95	1.39	0.15	4.66	5.25
		new	4.91	1.27	0.16	4.59	5.23
	Posttest	trained	3.92	1.42	0.15	3.63	4.22
		new	4.87	1.22	0.16	4.54	5.19

Familiarity, $F(1, 2{,}505) = 8.57$, $p < .003$, Training group × Familiarity, $F(1, 2{,}505) = 7.91$, $p = .005$, Session × Familiarity, $F(1, 2{,}505) = 32.68$ $p < .001$, and Training group × Session × Familiarity, $F(1, 2{,}505) = 16.46$, $p < .001$, also had a significant effect on accentedness ratings. Post-hoc analyses revealed a significant improvement after training only for trained words in both the non-clapping group, $F(1, 2{,}505) = 12.11$, $p = .001$, and the clapping group, $F(1, 2{,}505) = 4.93$, $p < .001$, meaning that training

had no effect in either group on participants' pronunciation of items they had not been trained for. Moreover, there was a significant difference between the clapping group and the non-clapping group at pretest for the trained items only ($F(1, 2{,}505) = 4.72$, $p = .03$).

Figure 4

Mean accentedness ratings for the word imitation task across groups (clapping, non-clapping), sessions (pretest, posttest) and familiarity (trained, new). Notes. Error Bars: +/− 2 SE



No effect of number of syllables or stress position was found, revealing that differences between accentedness scores for the two groups were not related to word length or stress position.

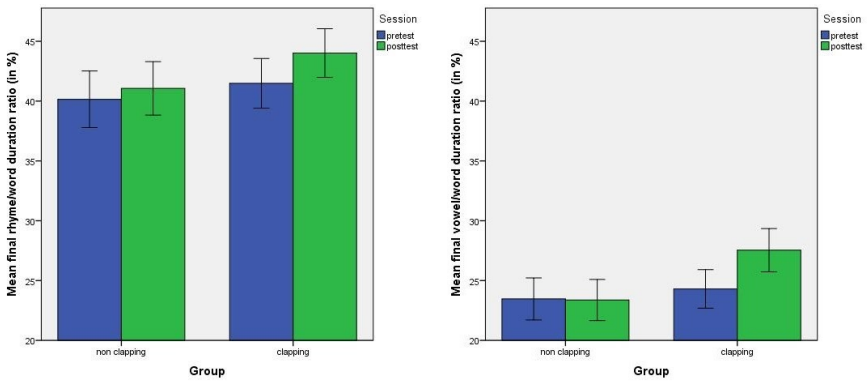
4.4.4 Effects of type of training and word length on acoustic measures

The two graphs in Figure 5 show the mean final rhyme duration ratio (in %, left graph) and the mean final vowel duration ratio (in %, right graph) at pre- and posttest for the non-clapping and clapping groups (for descriptive results, see Table 4). Results of the GLMM with final rhyme duration as a dependent variable showed a significant effect of training group ($F(1, 539) = 7.62$, $p = .006$) and session ($F(1, 539) = 12.65$, $p < .001$). However, no significant Training group × Session interaction was found.

As expected, a significant effect of number of syllables ($F(1, 539) = 11.88$, $p = .001$) revealed a significant difference in rhyme duration ratio depending on the length of the word. However, importantly, there was no significant interaction between Number of syllables × Training group or Number of syllables × Session.

Figure 5

Mean ratio (in %) of (a) the relative duration of the final rhyme and (b) the relative duration of the final vowel, broken down by group and session. Notes. Error Bars: +/− 2 SE.



By contrast, results of the GLMM with final vowel duration as a dependent variable revealed a significant effect of Training group ($F(1, 539) = 4.91$, $p = .027$), Session ($F(1, 539) = 7.69$, $p = .006$) and Training group × Session ($F(1, 539) = 4.62$, $p = .032$). Post-hoc analyses showed a significant difference between clapping and non-clapping at posttest only ($F(1,539) = 10.15$, $p = .002$) and a significant difference between pre- and posttest in the clapping group only ($F(1,539) = 19.19$, $p < .001$).

4.5 Discussion and conclusion

The results of the present investigation show that a short 20-minute training session involving hand-clapping during word learning in a second language helped children more than only audiovisual training in (1) reducing their accentedness as perceived by native speakers and (2) increasing the lengthening of the final vowel of the items, thus more closely approximating the way stress is phonetically realized in French. Two complementary sets of results back up our interpretation. First, regarding perceived accentedness, although both groups improved their pronunciation after training, our results show that the children in the clapping group reduced their accentedness scores significantly more than the children in the non-clapping group. Second, the results of the acoustic analyses show that the children belonging to the clapping group produced significantly longer (hence more target-like) final vowels5 after training than the children in the non-clapping group. Since French phrasal stress/rhythm is essentially characterized by a significant lengthening of the final full vowel (Delais-Roussarie et al., 2015; Delattre, 1966; Di Cristo & Hirst, 1993; Fletcher, 1991; Vaissière, 1991) and it is considered a stable indicator of prominence in this language, an appropriate realization of final lengthening is indeed of crucial importance for the production of French rhythm.

All in all, the results of the study corroborate and expand previous results on the beneficial role of rhythmic training for phonological

learning. From previous studies, we know that short training with rhythmic primes has positive effects on the phonological perception of speech in a first language (Cason & Schön, 2012; Cason et al., 2015). There is also extensive evidence that musical rhythmic activities increase children's phonological awareness and help develop their pre-reading skills (e.g. Herrera et al., 2011; Nelson, 2016; for a review, see Tierney & Kraus, 2013), and they can also be used as part of a method to help children with reading disorders (Flaugnacco et al., 2015; Habib et al., 2016; Overy, 2003). Rhythmic training has also been shown to have an immediate positive effect on the first-language phonological production of hearing-impaired children (Cason et al., 2015b).

In the context of the acquisition of second language pronunciation, our study backs up the general claim made by various researchers that suprasegmental (e.g. Darcy, Ewert & Lidster, 2012) or rhythmic training including a variety of activities can enhance the learning of prosody in a second language (e.g. on the use of rap music, see Fischler, 2009; on the use of beat gestures, see Gluhareva & Prieto, 2017; Kushch, 2018; on the use of hand-clapping, see Iizuka et al., 2020; Zhang et al., 2018). The present study thus extends the findings from the abovementioned studies by showing that a short audio-visual training session based on repeating words while hand-clapping their rhythmic structure can be of benefit in the second language classroom. First, while beat gestures may highlight higher levels of prosodic structure by

marking nuclear pitch accents (Gluhareva & Prieto, 2017; Kushch, 2018), and beat gestures that accompany L2 speech help speakers externalize the prosodic features of a foreign language (McCafferty, 2006), hand-clapping lends itself to indicating durational structure at the syllabic level.

Importantly, our study complements previous mixed findings by Zhang et al. (2018) and Iizuka et al. (2020) on the beneficial effects of hand-clapping. First, in Zhang et al.'s study participants obtained a benefit from hand-clapping training, yet effects were smaller than those reported in the present study. While hand-clapping training was significantly helpful for the Catalan children who participated in the present study in terms of general pronunciation assessments, this was not the case with Chinese adolescents as reported by Zhang et al. (although a positive effect of acoustic lengthening was found). The reason for this difference might be that the participants in Zhang et al.'s study had to learn word pronunciation and meaning at the same time, while the present study avoided the potential cognitive overload that this more complex learning task may have entailed by using cognate words (French words that were similar in phonological form and mean- ing to their Catalan counterparts). The stronger effect found here might thus be due to the fact that while the Catalan children could fully direct their attention and efforts towards how to pronounce the words, Chinese adolescents had to also learn the meanings of the French words at the same time. That said, the two studies complement each other by

both showing the beneficial effect of hand-clapping, regardless of the cognitive difficulty of the associated word learning task. Second, while participants in the hand-clapping group in Iizuka et al.'s (2020) study increased their perception abilities of segmental L2 features, they did not improve on their pronunciation of long vowels, geminates and moraic nasals. The fact that Iizuka et al.'s materials involved not only learning vowel duration patterns but also consonantal duration patterns may have made the task more difficult, and it may be that additional training was required to achieve appropriate pronunciation of these consonantal features. Moreover, since English loanwords in Japanese do not share the same phonological form as their English counterparts, a picture-naming task may have been too difficult to elicit pronunciation at an early stage of acquisition. Further research will be needed to tackle the role of hand-clapping on various aspects of L2 pronunciation at different stages of learning.

We believe the success of training with rhythmic hand-clapping may be explained by the properties of the hand-clapping sounds and movements. First, clapping on each syllable that makes up a word highlights the representation of the suprasegmental characteristics of the words through two perceptual channels, namely, the auditory and the visual channels. While both emphasize the temporal regularity of the syllables, the longer time spent with hands in contact emphasizes the stronger prominence present in the last syllable. Clapping to the rhythm of the words

may thus have served to reinforce participants' ability to perceive the prosodic pattern of foreign speech and led to better overall pronunciation scores. In addition, although hand-clapping cannot visually represent duration in space the same way as pitch gestures do, training effects were also found on the realization of final lengthening. In addition, the subsequent embodied reproduction of hand-clapping by the participants themselves may have reinforced the overall perceptive effect even further. This would be consistent with the theory of embodied cognition (Barsalou, 2008), according to which the motor modality is closely linked to the perceptual modality (Borghi & Caruana, 2015). This suggests that appropriate sensory and motor interactions may have triggered a more efficient development of human cognition, an idea that has crucial implications for learning and education (for reviews, see Kiefer & Trumpp, 2012; Wellsby & Pexman, 2014), including second language acquisition (Macedonia, 2019).

Regarding the effect of individual differences, an initial analysis revealed significant effects of language imitation aptitude and productive rhythmic skills, whereby better scores in these tasks predicted a reduction in the perceived accentedness of second language pronunciation. Good productive rhythmic skills, realized through hand-clapping, may have helped the children who ranked high in this aptitude to better follow the training session. By contrast, no significant effects were found for phonological and rhythmic perception scores, suggesting that rhythmic skills based

on perception alone play a more limited role in production. Finally, perhaps surprisingly, we found that working memory did not play any role in our results. However, since the trained words were French/Catalan cognates, the pretest and posttest tasks did not place high demands on working memory, unlike in Zhang et al. (2018), where the two languages involved were entirely unrelated and the meanings of new words were thus totally opaque to the learners.

Limitations and future research

There are several limitations to this study. First, training with hand-clapping was found to improve the pronunciation of trained items only, while generalization to new, non-cognate items was not found. Perhaps a longer training period would be needed to detect generalization effects. Crucially, the untrained words were not cognate words; results might have been different if the new words had been cognates, as in the training session. In addition, results regarding the new words might also have been different if the training session had been carried out with non-cognate words. A follow-up study with non-cognate words could further assess the potential role of word familiarization when assessing the effectiveness of hand-clapping.

The small number of participants in this study is another limitation and suggests further testing with larger groups. Moreover, in the case of the analysis of accentedness, a significant difference

between groups should have been observed at posttest only. With better scores in the non-clapping group at pretest, participants in this group may have had a narrower margin for improvement. We think that with a larger sample size, we might have been able to avoid the difference between groups at pretest. At this juncture, the acoustic analysis offered a more convincing type of evidence for the benefits of rhythmic training in this dataset. Importantly, further studies need to assess the effects of hand-clapping on other populations. For example, the effect of hand-clapping may show different outcomes depending on whether learners are still in a sensitive period for neural plasticity, as in the case of children, or not, as in the case of adults. Research has shown that adults tend to rely more on explicit knowledge to achieve learning than on a bottom-up process based only on direct experience (see, e.g. White et al., 2013). It would there- fore be of interest to replicate this experiment using older second language learners (but see Iizuka et al., 2020; Zhang et al., 2018). Further, specific types of learners (e.g. kinaesthetic learners) may benefit more from this type of training, and this aspect should be taken into account in future studies. Regarding language proficiency, future assessments might want to recruit participants who are actually in the process of learning the foreign language rather than merely exposed to a set of words for experimental purposes. Training could then be of a longer duration and findings extended to the learning of prosody across full sentences rather than single words.

In addition, the benefits of hand-clapping on L2 pronunciation should be further investigated with languages of different typologies. By highlighting the specific properties of a language in terms of syllabic structure, stress placement or prosodic patterns, hand-clapping could at the very least raise learner awareness of these elements and help improve pronunciation. Finally, in order to widen our understanding of the embodiment theory, more evidence is needed to assess the benefits of producing hand-clapping after watching the teacher as compared to merely watching the teacher and not replicating his/ her behavior.

Conclusion

The results of the present study offer two complementary pieces of evidence – one perceptual, the other acoustic – of the usefulness of word-based rhythmic training with hand-clapping for the acquisition of L2 pronunciation. First, the study shows that rhythmic training with hand-clapping leads to greater improvement in accentedness scores. Second, acoustic analysis yields more specific evidence of the benefits of rhythmic training for the acquisition of final vowel lengthening. All in all, these results expand and complement the previous mixed evidence reported in Zhang et al. (2018) and Iizuka et al. (2020) on the role of rhythmic training through hand-clapping on pronunciation learning.

Learners' development of L2 pronunciation in an instructional context requires effective and practical tools. Given that mastering

L2 suprasegmental features helps learners reduce their accentedness and improve their oral proficiency (e.g. Kang et al., 2010), training students in these prosodic features should be considered an important part of pronunciation instruction. In this context, from a practical perspective, hand-clapping is clearly a technique that would be easy to implement in the foreign language classroom as a tool to teach language rhythm patterns. Additionally, it would make the implementation of repetition-based drills or singing activities more engaging and pleasant and would potentially enhance motivation, especially with children. We surmise that the combination of acoustic and visual information channels, together with the motor experience involved in hand-clapping, can lead to a more optimized learning of the rhythmic structure of a novel language and consequently to better production of these rhythmic patterns.

Finally, from an educational point of view this approach would nicely mesh first with recent proposals advocating multisensory and embodied trainings in the classroom (e.g. Kiefer & Trumpp, 2012) and, second, a fuller integration of music and foreign language learning (for a review, see Viladot & Casals, 2018), which is consistent with an interdisciplinary approach to education (e.g. Jones, 2010). In our view, considerable work still remains to be done to develop and empirically test second language programs and educational tools based on rhythmic and melodic training which at

the same time favor communicative situations and goal-based meaningful activities in the second language classroom.

5

CHAPTER 5: GENERAL DISCUSSION AND CONCLUSION

5.1 Summary of findings

The general goal of this dissertation was to experimentally assess the potential benefits of different types of embodied pronunciation training techniques that highlight prosodic features of a target language. Crucially, the thesis adopts a multisensory approach to phonological learning in a foreign language. While previous experimental studies on embodied prosodic training have mainly focused on the role of gestures that highlight articulatory features (e.g. P. Li et al., 2021; Xi et al., 2020) or rhythmic features (e.g., Gluhareva & Prieto, 2017; Kushch, 2018; Llanes-Coromina et al., 2018), the main goal of the present dissertation is to broaden the scope of investigation and focus on the role of visuospatial hand gestures and hand-clapping (e.g., two types of embodied enactments of speech prosody) on the phonological learning of non-native prosodic features.

We carried out three between-subject training studies with a pretest and posttest design to assess the role of hand gestures and percussive hand movements in the learning of three types of prosodic features, namely (a) lexical tones (Study 1), (b) phrasal intonation (Study 2) and (c) word rhythm (Study 3), with each study addressing one aspect. The assessment of the potential pronunciation gains from these embodied techniques was tailored to the specific difficulties of the feature trained in each study and the proficiency of the participants in the target language.

The first study (Chapter 2) investigated whether seeing and producing visuospatial hand gestures mimicking pitch contrasts at the syllabic level helped Catalan speakers without any knowledge of Mandarin Chinese to identify Mandarin Chinese lexical tones and words. The study looked exclusively at the perception of Mandarin Chinese lexical tones in monosyllabic words by 106 naïve learners and how pitch gestures helped them retrieve the meaning of the word when competing in a minimal pair. The results of Study 1 showed that participants who watched pitch gestures during training on Mandarin Chinese lexical tones and words improved significantly more from pretest to posttest compared to participants who were not exposed to gestures in a tone identification task. In addition, the experimental group obtained significantly higher scores in a meaning-association task involving minimal pairs of Mandarin Chinese words contrasting in tone. Similar results were found when the experimental group was asked to imitate the pitch gestures during training. In a second step, the results of the gesture observation and gesture production groups were compared and no significant difference between them was found. All in all, these results suggest that both watching and producing visuospatial hand gestures encoding pitch contours may help novice learners acquire novel prosodic contrasts in an L2.

The second study (Chapter 3) explored the effects of visuospatial hand gestures encoding pitch and rhythmic properties at the

phrase-level with 75 Catalan learners of French. The study assessed the pronunciation of French sentences by intermediate Catalan learners in a comprehensive manner by measuring perceived comprehensibility, fluency, and accentedness as well as the perceived pronunciation of segmental and suprasegmental features in an oral-reading task both at pre- and posttest. The results of Study 2 showed that embodied prosodic training with visuospatial hand gestures encoding phrasal melodic and rhythmic features of French yielded an greater improvement in the pronunciation of suprasegmental features and reduced accentedness from pretest to posttest in an oral-reading task compared to training with mere oral sentence repetition. As for comprehensibility, fluency, and segmental accuracy scores, both training groups (embodied and non-embodied) showed similar levels of improvement.

Finally, the third study (Chapter 4) assessed the potential benefits of hand-clapping cueing the rhythm of French words both acoustically (clapping rate on each syllable) and visually (joined hands longer on final lengthening) on the pronunciation of these words by 28 naïve Catalan children. The study assessed the pronunciation of French cognate words, whose meaning was transparent to our young naïve learners of French, both before and after training. The results of Study 3 revealed that children who performed hand-clapping on the syllabic and rhythmic structure of French words while repeating the words (a) reduced their

accentedness scores on the pronunciation of these words in an imitation task at posttest; and (b) showed a more target-like final lengthening patterns, with longer durations of the target vowel in the last syllable, compared to children who merely repeated the words during training. These results suggest that producing percussive hand movements encoding rhythmic features may help novice learners to reduce accentedness and more accurately produce non-native final lengthening patterns.

All in all, the three studies jointly show that embodied interventions involving visuospatial hand gestures and percussive hand movements encoding a variety of prosodic features (i.e. pitch, rhythm, and melodic features) are beneficial for phonological learning. Importantly, different aspects of phonological learning were touched upon, namely perception and production skills. Production skills were assessed by using overall measures (fluency, comprehensibility, and accentedness), as well as specific perceptual constructs (segmental and suprasegmental features, acoustic analysis) of pronunciation evaluation (Saito and Plonsky, 2019). In the next section we discuss the specific effects of embodied prosodic training on pronunciation gains and discuss why they work.

5.2 Effects of prosodic embodiment techniques on L2 phonological learning

5.2.1. Effects of observing vs producing prosodic embodiment

One of the goals of this doctoral dissertation was to comprehensively assess the role of embodied techniques with different types of hand movements in foreign language phonological learning, from the perspective of training novel phonological features with gesture observation and gesture production.

A specific goal of Study 1 was to compare the effects of observing versus producing pitch gestures on the perception of Chinese lexical tones. A variety of studies have suggested that the production of gestures by the learners is more effective than observing them alone in various learning contexts (e.g., Goldin-Meadow, 2014; Goldin-Meadow et al., 2009; Macedonia et al., 2011; Masumoto et al., 2006; see also Saltz & Donnenwerth-Nolan, 1981, for motoric enactment). Study 1 was the first to test this hypothesis specifically in regards to pitch gestures in L2 phonological learning. We found that learning Mandarin lexical tones by (a) observing an instructor utter the words and produce the gesture and (b) observing and then imitating

the instructor's speech and gesture were equally beneficial in both tone- and word-learning tasks. Our results thus add new evidence on the positive role of perceiving and producing gestures encoding phonological features in strengthening the link between semantic meaning and phonological forms, in line with previous studies showing the beneficial role of different types of gestures in this domain (e.g., Kushch et al., 2018; P. Li et al., 2021; Morett & Chang, 2015; So et al., 2012).

The positive results on the use of pitch gestures contrast with previous findings on observing durational gestures showing that hand gestures may have limited effects on identifying non-native phonological contrasts such as duration (e.g., Hirata et al., 2014; Hirata & Kelly, 2010, Kelly et al., 2017; P. Li et al., 2020) and aspiration (Xi et al., 2020). However, our results support previous findings by Kelly et al. (2017), who showed that congruent pitch gestures favored the identification of intonational contrasts by English-speaking learners of Japanese. The discrepancy in terms of benefits for the perception of L2 phonological features between the visuospatial gestures representing pitch and the one representing duration at the syllable level cannot yet be easily explained and more experimental evidence would be needed to continue exploring this issue.

Interestingly, M. Li and De Keyser (2017) provided strong evidence that tone-word perception and production skills each

depend on the type of practice, that is perception training tends to enhance perception scores while production training tends to enhance production scores. In Study 1, even though both gesture perception and production were trained, only participants' perception was evaluated, showing no difference between the groups. It may well be the case that if participants were asked to produce the words themselves, superior outcomes could be found for participants who performed the gestures. Along this line, Studies 2 and 3 show that producing gestures and percussive movements had a positive effect on production (pronunciation) skills, in line with previous studies testing gesture production (e.g., Morett & Chang, 2015; Kushch, 2018, Llanes-Corominas et al., 2018; F. Zhang, 2006) or kinesthetic training (e.g., Hamada, 2018, Iizuka et al., 2020; B. Lee et al., 2020; Yang, 2016). Further research would be needed to test the effects of embodied productive training on perceptive outcomes, in order to further assess the value of gesture production and perception patterns for phonological training.

5.2.2. Specific effects of embodied prosodic training on pronunciation

A specific goal of this dissertation was to evaluate the direct effect of embodied prosodic training on the pronunciation of suprasegmental features. As expected, our three studies confirm this hypothesis. In Study 1, participants who perceived or produced

the pitch gestures obtained significantly higher scores in lexical tone identification. In Study 2, our results provide evidence that the embodiment of phrasal prosodic features in a foreign language helped learners produce suprasegmental features more efficiently at posttest: as expected, embodying prosody directly improved the scores on suprasegmental accuracy on the oral tasks at posttest. In Study 3, children who performed hand-clapping on the metrical structure of French words during training managed to produce significantly longer and more target-like durations of the prominent word-final vowels. The motoric actions triggered by the perception and the production of the gestures and percussive hand movements may have fostered the adequate phonological representation of rhythmic and melodic patterns, boosting the processing and the acquisition of such features. The abovementioned findings are in line with results of previous studies showing that embodied pronunciation interventions with hand gestures encoding specific prosodic features directly improve participants' learning outcomes of these prosodic features, such as lexical tones (Morett & Chang, 2015), intonation (Kelly et al., 2017; Yuan et al., 2019), vowel duration (P. Li et al., 2020), as well as word stress (Ghaemi & Rafi, 2018).

Interestingly, the results of Study 2 and Study 3 not only showed the positive results of embodied prosodic training on the target suprasegmental features, but also its overall beneficial impact on

pronunciation in terms of accentedness. This also confirms the results of previous studies focusing on effects of beat gestures and hand-clapping on pronunciation scores (e.g., Gluhareva & Prieto, 2017; Kushch, 2018; B. Lee et al., 2020; P. Li et al., 2020; Llanes-Coromina et al., 2018; Y. Zhang et al., 2020a). As for comprehensibility and fluency scores, the results of Study 2 showed that both training groups (embodied and non-embodied) attained similar levels of improvement. This may be explained by the fact that suprasegmental features may weigh less in the assessment of comprehensibility and fluency measures than in the accentedness measure (e.g., Trofimovich & Isaacs 2012; Saito et al. 2016). In addition, the larger effect sizes obtained for embodied prosodic training in both measures may also point to a certain advantage for this type of training. For this reason, complementary research extending the duration of the training period or performing a delayed posttest would be necessary, as suggested by Alazard (2013). Another possible explanation for our results is that learners with higher proficiency levels may have already overcome most comprehensibility and fluency difficulties in a reading task, while this remains a challenging task for beginner to intermediate learners. A different pattern of results may be expected from a less controlled oral production task, with higher cognitive demands for the mobilization of syntactic, lexical, and phonological resources.

Regarding embodied kinesthetic training, the results of Study 3 demonstrate the effectiveness of hand-clapping for pronunciation learning, a technique frequently used in musical education (e.g. Romero Naranjo, 2013) and when teaching literacy skills to children (e.g., Batchelor & Bintz, 2012; Kern, 2018). Study 3 replicated the findings from a parallel study (Y. Zhang et al., 2020a) by showing that hand-clapping helps naive young learners of French to produce more accurate French final lengthening patterns. Moreover, our findings complement and extend previous evidence on the positive effects of hand-clapping on listening comprehension (B. Lee et al., 2020), on the identification of long phonemes in Japanese (Iizuka et al., 2020), and on pronunciation accuracy (B. Lee et al., 2020). Hand-clapping is a natural hand percussive movement that is frequently used by children in song play and when following musical rhythms (e.g., Brodsky & Sulkin, 2011). By imitating the instructor, our young participants were able to follow and reproduce the target rhythm more closely. Hence, the improvement in pronunciation may have come from the integration of the language and motor systems through the synchronization of verbal (word syllables) and movement sequences (hand claps) (Sulkin & Brodsky, 2007). All in all, Study 3 extends our general knowledge of embodied prosodic techniques using hand gestures to other types of hand movements, in this case, hand-clapping, and opens the door to explore other techniques related to kinesthetic movement and musical rhythm.

Finally, regarding the potential role of prosodic training on segmental accuracy, Study 2 did not find a direct beneficial effect of our training with phrasal-level prosodic gesture on segmental accuracy in an oral-reading task, contrasting with previous findings (Missaglia, 1999, 2008; Saito & Saito, 2017; Hardison, 2004). However, in a recent study (Li et al., under review), embodied prosodic training using phrasal-level prosodic gestures was tested in a relatively similar design as Study 3. In this study, the frequency of the target difficult segments (the French front rounded vowels /y, ø, œ/), was increased in the dialogue stimuli. Results showed that the participants in the embodied prosodic training group not only improved their accentedness and their pronunciation of suprasegmental features but also produced more target-like French rounded vowels compared to participants who followed training based on speech repetition only. Therefore, it seems that embodied prosodic training may improve segmental accuracy when the frequency of the target items is increased (e.g., Gullberg et al., 2010, 2012; Lyster, 2017)

5.2.3 Controlling for individual differences

A trending topic in second language acquisition and in phonological acquisition in particular is the role of cognitive individual differences such as sound discrimination ability and working memory in learning processes (e.g., Baker-Smemoe & Haslam, 2013; Kachlicka et al., 2019; Saito, 2017; Saito &

Hanzawa, 2016; Safronova, 2016; Saito, Suzukida, et al., 2019; Saito et al., 2020; Zheng et al., 2020). Musical perceptive abilities have been revealed to be predictive factors in L2 pronunciation abilities (e.g., Kempe et al., 2015; M. Li & DeKeyser, 2017; Milovanov et al., 2010; Moyer, 2014; Piske et al., 2001; Richter, 2018; Slevc and Miyake, 2006). Furthermore, individual differences in motivation (Dörney, 2009) may well be crucial to obtain the desired result of a training experiment. In the three studies presented in this dissertation, individual differences have been assessed to different extents. In Study 1, we controlled for participants' phonological working memory, which was important for the word learning task. In Study 2, the motivation of the students and their perception of their own achievements was assessed after training, showing very positive evaluations of the training program in all the groups. In Study 3, a more comprehensive battery of tests was taken to assess individual differences in terms of phonological discrimination abilities, language imitation skills, phonological working memory, as well as musical rhythmic perceptive and productive abilities. It was ensured that the between-subject groups did not differ in these measures, and when some of the measures (phonological disrimination and language imitation skills) were found to have an effect on the results of a task (accentedness ratings), they were added in the random effect structure of the general statistical model. It is desirable that the design of future training studies take

into account the relevant individual differences not only to control for balanced experimental groups, but also to assess potential interference with training effects.

5.3 Why is prosodic embodiment so effective for pronunciation learning? Implications for the Embodied Cognition paradigm

As stated in Chapter 1, the three studies in this thesis stem from the theoretical framework of Embodied Cognition (e.g., Barsalou, 2008; Foglia & Wilson, 2013) and its implications for education (Ionescu & Vasc, 2014; Kiefer & Trumpp, 2012; Shapiro & Stolz, 2019; Wilson, 2002). The three studies were designed to test the predictions of this theory on phonological learning.

According to the Embodied Cognition hypothesis, cognition takes place both in the brain and in the motor and perceptual systems (e.g., Barsalou, 2008, 2010; Foglia & Wilson, 2013; Gallagher, 2005; Lakoff & Johnson, 1999), notably through the mechanism of simulation of action and the simulated reenactment of perceptual, motor, and introspective states taking place during any interaction with the world (e.g., Barsalou, 2008; Decety & Grezes, 2006; Goldman, 2006). Mirror neurons, which are activated during action observation, are assumed to play an important role for action processing and learning through imitation (Fu & Franz, 2014; Rizzolatti & Craighero, 2004). The cognitive role of body movement itself is also addressed under the Embodied Cognition paradigm with the concept of cognitive offloading: humans compensate their limited information-processing abilities, by distributing cognitive demands onto the world or the body, i.e. by

body movement or by gesturing (Glenberg & Robertson, 1999; Risko & Gilbert, 2016; Wilson, 2002). By extension, gestures can be considered useful to reduce cognitive load (Post et al., 2013).

The present dissertation had the goal of broadening the scope of embodied cognition and embodied learning and extending it to the domain of phonological learning of a foreign language. While such embodied theories as Mahon and Caramazza's grounding by interaction hypothesis (2008) focused on the facilitating role of action perception and action realization on the processing of concepts, the present dissertation concentrated on phonological processing, adding evidence to the fact that the implications of embodied cognition may be extended to more domains of understanding and learning. Similarly, McNeill's Growth Point theory (2005) stated that gesture and speech coexpressively embody a single underlying meaning during communication and together participate to the semantic and pragmatic processing of utterances; however, the beneficial effects of gesture-speech integration reported in this thesis also show that the benefits of this integration may also apply to lower levels of processing such as phonological processing.

Previous research on embodied cognition related to second language learning focused on successful classroom interactions (e.g., Eskildsen & Wagner, 2013, 2015; Hasegawa, 2009; Jakonen, 2020) and lexical learning (Asher, 1972; Pesce et al., 2009;

Mavilidi et al., 2015), our studies set out to test the beneficial role of motor actions related to prosody in phonological learning. The studies in this dissertation proposed to implement both imagery and motoric enactment respectively through the perception and imitation of visuospatial hand gestures and percussive hand movement representing prosodic features. Following the *grounding by interaction hypothesis* (Mahon & Caramazza, 2008), the visuospatial hand gestures and hand-clapping have provided a richer conceptualization of the prosodic features under scrutiny. Through the visual modality, pitch gestures have *grounded* the concept of Chinese lexical tones in a spatial dimension thanks to the internalized metaphor of height for pitch contours, adding this knowledge to the abstract conceptualization of lexical tones. In the same way, phrase-level prosodic gestures have *grounded* the concept of French intonation and phrasal rhythmic properties also in a spatial dimension thanks to the same metaphor of height for pitch contours, and also thanks to the linear representation of time in space. Finally, hand-clapping has mobilized both visual and auditory modalities to encode rhythm in addition to speech. Seeing and feeling the hands move in rhythm while listening to the sound of the claps might have helped grounding the concept of rhythmic properties of French words, one of them being the realization of the prominent syllable thanks to seeing the maintenance of the hands together implying duration. Therefore, even if the participants in the three studies may not have any abstract conceptualization of the

prosodic features, embodied prosodic training might have triggered grounded conceptualization of prosodic features.

Finally, the significant positive effect of visuospatial gestures and kinesthetic movements found in the three studies in comparison to non-gestural control groups must be nuanced. The sizes of the effects (based on moderate improvements measured on Likert scales or acoustic data) in our short interventions remained modest, which is in line with most of the studies with similar experimental designs looking at the effect of gestures on L2 learning (see section 1.3). In addition, some studies have also shown that gestures may sometimes trigger null or detrimental effects on L2 comprehension and recall (e.g. Rohrer et al. 2020), on the perception of L2 phonological features (e.g., Hirata & Kelly, 2010; Hirata et al. 2014; Kelly et al., 2014) and on L2 pronunciation (e.g. Hoetjes & Van Maastricht, 2020, Iizuka et al., 2020). Hence, as explained in sections 5.4 and 5.5, more research is needed to continue testing the general efficacy of gesture in L2 learning depending on structural factors (e.g. the adequacy of the gesture) and on learners' individual differences (e.g. gesture performance).

5.4 Practical implications for pronunciation teaching: a multisensory approach

The results of the three studies in this dissertation provide further empirical evidence on the value of multisensory training involving prosody in language teaching practice. While a number of classroom observations (e.g., Hudson, 2011; Rosborough, 2010) and teaching proposals (e.g., Chan, 2018; Odisho, 2007; Roberge et al., 1996; Smotrova, 2017) have promoted the use of gestures, tactile/kinesthetic actions, and body movements conveying phonological features during L2 pronunciation instruction, it is not until recently that researchers have started to empirically assess the effectiveness of these techniques in learning L2 pronunciation. Our results have added new evidence in favor of using embodied approaches to L2 pronunciation teaching that include multiple sensory experiences. Importantly, they empirically support previously proposed methodologies that encourage the use of such multisensory activities in pronunciation learning for segments (e.g., Esteve-Gibert et al., 2021; Hardison 2003, 2005; Haught & McCafferty, 2008; Hazan et al., 2005, 2006; Hirata & Kelly, 2010; Inceoglu, 2016; Y. Li & Somlak, 2017 ; Ozakin et al., 2021). Importantly, our studies provide further evidence of the benefits of such training on prosodic learning, in line with teaching method proposals favoring multisensory prosodic activities and body

movement (e.g., Acton et al., 2013; Llorca, 2001; Odisho, 2007; Soulaine, 2013).

In particular, using the hands to highlight prosodic features has been proven to be an effective technique that can come handy in the classroom, as it does not require any specific technology and can be put into practice whenever needed. However it is important that the gestures used in the classroom are adequate and rightly depict the target features. In our experiments, the movements were not difficult to imitate by the learners. In Study 1, the gestures depicting pitch movements taking place on lexical tones were based on the well known metaphor between space and pitch height. In Study 2, hand-clapping represented a very straightforward and familiar way to mark rhythmic patterns. Finally, the movements in Study 3 were more challenging, but the repetition of the same schematized prosodic structures over a variety of sentences helped learners to process and understand these structures, and they were able to reproduce these movements themselves. In the eventuality that instructors need to create their own gesture based on their teaching needs, they should be cautious and employ gestures that are easily understood by the learner, and if possible these gestures should be easy to imitate for the learners. For example, when using hand gestures, the shape and the movement of the hand should be appropriately designed and performed. As shown by P. Li et al., 2020; Xi et al., 2020, and P. Li et al., 2021, if gestures misrepresent the target feature to be learned, or if learners cannot manage to

imitate the gestures appropriately, even if they are well designed, embodied training with hand gestures may not helpful. According to P. Li et al. (2021), the quality of gesture performance might indicate learners' cognitive load, learning motivation, effort, and so on. In Study 3, it was checked that the learners did not find it awkward to have to mimic the phrase-level prosodic gestures and their motivation was controlled for across groups.

All in all, the results obtained in the three studies allow the author of this dissertation to encourage language teachers to adopt embodied and multisensory approaches to pronunciation learning. In particular, we recommend focusing on embodied *prosodic* techniques, which trigger benefits not only on the production of suprasegmental features but also on general pronunciation. As advocated by different teaching approaches within the communicative framework (e.g., ACCESS, Gatbonton & Segalowitz, 1988, 2005; TBPT, Ellis, 2009), focus-on-form and by extension prosodic training can be integrated in the language classroom without compromising the main objectives of interaction and meaningful activities (e.g., Gordon, 2021). Likewise, we believe that integrating visuospatial gestures and kinesthetic techniques into the focus-on-form activities of communicative lesson plans would raise learners' awareness on prosodic features and boost phonological learning. In addition, such techniques respect four of the five principles of teaching pronunciation proposed by Colantoni et al. (2021): the importance of

perception-based activities, the teaching of prosodic features, the incorporation into a communicative context, and the focus on features with high functional load. In more practical terms, in view of the importance of repetition for phonological learning (e.g., Bradlow et al., 1997; Jung et al., 2017; Lord, 2005; Saito, 2015; Trofimovich & Gatbonton, 2006), we also suggest that teachers check that learners get used to the specific techniques, by using them often during class and by encouraging learners to produce them.

5.5. Limitations and future research

One of the limitations of the present results is the lack of assessment of long-term learning effects. As many previous studies with an embodied pronunciation paradigm (e.g., Gluhareva & Prieto, 2017; Hirata & Kelly, 2010; Hoetjes & van Maastricht, 2020; Xi et al., 2020; Yuan et al., 2019), the three studies in the present dissertation tested training effects through an immediate posttest. However, there is recent evidence that embodied techniques may be even more effective than non-embodied techniques at delayed posttests either by helping students maintain training effects (e.g., P. Li et al., 2021b, R. Zhang & Yuan, 2020) or even by further improving pronunciation scores (e.g., B. Lee et al., 2020, P. Li et al., 2021a). Therefore, future studies would need to further explore the impact of embodied pronunciation training in the long run.

A second limitation relates to the lack of a fine-grained systematic assessment of the pronunciation gains in our embodied prosodic training that relates to the issue to what extent suprasegmental and segmental components are affected. First, while Study 3 complemented the perceptual assessments of pronunciation with acoustic measures of the realization of French final lengthening, in Study 2 it may be interesting to contrast the results of the perceptual evaluations with acoustic measures of fluency and

segmental and suprasegmental features to understand further what are the most influential components of fluency, comprehensibility, and accentedness. Second, the effects of embodied prosodic training on segmental accuracy should be further investigated to further strengthen the potential bootstrapping role of prosody in phonological learning (see Li et al., under review, for recent positive evidence of prosodic trainings on the acquisition of segmental features).

Regarding perception vs. production outcomes in embodied prosodic training, while Kushch (2018) found more benefits of beat gesture production compared to beat gesture perception, Study 1 in the present dissertation did not find any differences between the effects of pitch gesture observation and production. However, note that these studies are hardly comparable, as they observe very different effects (accentedness ratings in Kushch, 2018 and identification scores of tonal contrasts in Study 1). With respect to the rest of embodied prosodic techniques that have been tested so far, very little information is available to assess if production is better than perception. However, a comparison could be established between our results in Study 2 and the results obtained by P. Li et al. (under review), who proposed a similar training paradigm with the same type of participants, but with gesture observation. The authors found beneficial effects of phrasal-level prosodic gestures not only on accentedness and suprasegmental features as in Study 2, but also on segmental accuracy. Regarding kinesthetic training

techniques as in Study 3 (hand-clapping), we believe that the comparison would not be applicable, as these techniques specifically require to activate the haptic sensory modality or whole body movements. Crucially, in their meta-analysis comparing comprehension-based instruction to production-based instruction for grammar learning, Shintani et al. (2013) observed that whereas both types of instruction produced large effects for both receptive and productive knowledge at immediate posttests, production-based instruction showed a significant advantage in the long run (75 days after treatments). If learning by doing is advantageous for grammar learning, it may well be the case that a similar scenario can be observed for pronunciation learning. Therefore, more research is needed to assess perception vs. production outcomes in the long term.

Following previous research suggesting the importance of (a) the adequacy of gesture choice in the training of segmental features (Hoetjes & van Maastricht, 2020; Xi et al., 2020), and (b) the participants' gesture performance during training (P. Li et al., 2021a), it is important to highlight the fact that the accuracy of the target gestures used in embodied training (in terms of both the adequacy of the target gesture choice and the adequacy of the learner's gesture performance during training) should be assessed further. Crucially, when a common representational mapping between motor, sensory and abilities can be established, learning is likely to be enhanced (e.g., Zhen et al., 2019). If pitch gestures and

hand-clapping seem easy enough to be performed spontaneously by a teacher, producing an accurate phrase-level prosodic gesture over a sentence may be more challenging. In view of the evidence above, further research is needed to evaluate the adequacy of certain visuospatial gestures as transparent metaphorical representations of phonological features.

Related specifically to embodied and multisensory training involving the production of gestures or movements, individual differences in terms of motoric imitation are of utmost importance, as has been recently discovered. P. Li et al. (2021a) showed that participants who did not appropriately imitate the instructor's gestures during training did not benefit from the use of these gestures. For the production of gestures, these differences may not be explained by motor timing skills or fine motor skills (e.g. Lorås et al., 2013), however, there is some evidence that hand-clapping skills are related to foreign language imitation skills (Y. Zhang et al., 2020b). In any case, the findings by P. Li et al. (2021a) suggest that the movements of the learners during embodied training need to be adequate, and so, some action in terms of motivation, or familiarization with the new gestures or movements may be necessary. The studies of the present dissertation did not control for participant's gesture performance, which may have influenced our results.

As a final remark, the notion of learner motivation has become central in second language acquisition (e.g. Carrio-Pastor & Mestre Mestre, 2014; Dörney, 2009, Gardner, 2010), and has important implications for phonological learning (Purcell & Suter, 1980; Elliot, 1995a; 1995b; Moyer, 1999; Muñoz & Singleton, 2007; Shively, 2008). Some recent qualitative evidence indicates that embodied activities in the language classroom have a positive impact on learners' motivation (Zirak & Chicho, 2021). Further studies should therefore assess the potential activating role of embodied and multisensory phonological training on learners' extrinsic motivation.

5.6 General conclusion

The present dissertation represents one step forward towards the implementation of a set of prosodic-based embodied techniques in the field of pronunciation instruction. One of the main contributions of this work is to show that visuospatial hand gestures and movements mimicking prosodic features can be used to improve foreign language learners' pronunciation, thereby representing an important tool for integrating pronunciation teaching and learning into the foreign language classroom and improving oral skills. The three studies reported in this thesis contribute in different ways to mount experimental evidence on the benefits of a multisensory approach to language teaching with the activation of aural, visual, and motor channels. Results give clear evidence of the benefits of using a variety of visuospatial hand gestures and movements depicting prosodic features during pronunciation training across different phonological tasks - perception, imitation and reading aloud - and with students at various levels of proficiency. Studying the gestures produced by the language teachers and the learners, either spontaneously or with a planned methodology, and examining their contribution to the improvement of L2 speaking skills, is a necessary step in building a strong theory for embodied pronunciation learning as well as for the constitution of evidence-based embodied programs for teacher training.

Bibliography

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Abrahamson, D. & Sánchez-García, R. (2016). Learning is moving in new ways: The ecological dynamics of mathematics education. *Journal of the Learning Sciences, 25*(2), 203–239.

Abril, C. R. (2011). Music, movement, and learning. In R. Colwell & P. R. Webster (Eds.), *The MENC handbook of research in music learning, volume 2: Applications* (pp. 92–129). New York, NY: Oxford University Press.

Acton, W., Baker, A. Ann., Burri, M., & Teaman, B. (2013). Preliminaries to haptic-integrated pronunciation instruction. In J. M. Levis & K. LeVelle (Eds.): *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 234-244). Ames, IA: Iowa State University.

Alazard, C. (2013). *Rôle de la prosodie dans la fluence en lecture oralisée chez des apprenants de Français Langue Etrangère* [doctoral dissertation, Université Toulouse le Mirail - Toulouse II]. TEL Archives Ouvertes.

Alazard, C., Astésano, C., & Billières, M. (2010, May). The implicit prosody hypothesis applied to foreign language learning: From oral abilities to reading skills. In

Proceedings of the International Conference on Speech Prosody (p. 648), Chicago, USA.

Alazard-Guiu, C., Santiago, F., & Mairano, P. (2018, June). L'incidence de la correction phonétique sur l'acquisition des voyelles en langue étrangère : étude de cas d'anglophones apprenant le français. In *XXXII^e Journées d'Études sur la Parole* (pp. 116–124), Aix-en-Provence, France.

Aliaga-Garcia, C. (2017). *The effect of auditory and articulatory phonetic training on the perception and production of L2 vowels by Catalan-Spanish learners of English* [doctoral dissertation, Universitat de Barcelona]. Tesis en Xarxa.

Aliaga-Garcia, C., Mora, J.C., & Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Studies in Contemporary Linguistics, 47*(1), 1–14.

Alibali, M. W., & Kita, S. (2010). Gesture highlights perceptually present information for speakers. *Gesture, 10*(1), 3–28.

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15*(6), 593–613.

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science*, *22*(9), 1138–1144.

Allen, L. Q. (1995). The effects of emblematic gestures on the development and access of mental representations of French expressions. *The Modern Language Journal*, *79*(4), 521–529.

Allen, L. (2000), Nonverbal accommodations in foreign language teacher talk. *Applied Language Learning, 11*(1), 155–176.

Alves, U. K., Magro, V. (2011). Raising awareness of L2 phonology: Explicit instruction and the acquisition of aspirated /p/ by Brazilian Portuguese speakers. *Letras de Hoje, 46*, 71–80.

Amengual, M. (2012). Influence bilingual in interlingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition, 15*(3), 517–530.

Amand, M., & Touhami, Z. (2016). Teaching the pronunciation of sentence final and word boundary stops to French learners of English: Distracted imitation versus audio-visual explanations. *Research in Language*, *14*(4), 377–388.

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

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38*(4), 561–613.

Antonucci, S. M., and Alt, M. (2011). A lifespan perspective on semantic processing of concrete concepts: does a sensory/motor model have the potential to bridge the gap? *Cognitive, Affective, & Behavioral Neuroscience, 11*(4), 551–572.

Asher, J. J. (1969). The total Physical Response approach to second language learning. *Modern Language Journal, 53*(1), 3–17.

Asher, J. J. (1972). Children's first language as a model for second language learning. *Modern Language Journal, 56*(3), 133–139.

Asher, J. J., & Price, B. S. (1967). The learning strategy of the Total Physical Response: Some age differences. *Child Development, 38*(4), 1219–1227.

Astésano, C. (2001). *Rythme et Accentuation en Français: Invariance et Variabilité Stylistique*. París: L'Harmattan.

Atilgan, H., & Bizley, J. K. (2021). Training enhances the ability of listeners to exploit visual information for auditory scene analysis. *Cognition, 208*, 104529.

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., & Bizley, J. K. (2018). Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron, 97*(3), 640-655.

Austin, E. E., & Sweller, N. (2014). Presentation and production: The role of gesture in spatial communication. *Journal of Experimental Child Psychology*, *122*(1), 92–103.

Bahnmueller, J., Dresler, T., Ehlis, A. C., Cress, U., & Nuerk, H. C. (2014). NIRS in motion – Unraveling the neurocognitive underpinnings of embodied numerical cognition. *Frontiers in Psychology, 5*(743), 1–8.

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly*, *48*(1), 136–163.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics, 44*(3), 231–259.

Baker-Smemoe, W., & Haslam, N. (2013). The effect of language learning aptitude, strategy use and learning context on L2 pronunciation learning. *Applied Linguistics, 34*(4), 435–456.

Barenberg, J., Berse, T., & Dutke, S. (2011). Executive functions in learning processes: do they benefit from physical activity? *Educational Research Review, 6*(3), 208–222.

Barriuso, T. A., & Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *The CATESOL Journal*, *30*(1), 177–194.

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

Barsalou, L. W. (2010). Grounded cognition: past, present, and future. *Topics in Cognitive Science, 2*(4), 716–724.

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

Batchelor, K. E., & Bintz, W. P. (2012). Hand-clap songs across the curriculum. *Reading Teacher*, *65*(5), 341–345.

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *Developmental Psychobiology*, *40*(3), 293–310.

Bavelas, J. B., & Chovil, N. (2006). Nonverbal and verbal communication: Hand gestures and facial displays as part of language use in face-to-face dialogue. In V. Manusov & M. Patterson (Eds.), *The Sage handbook of nonverbal communication* (pp. 97–115). Thousand Oaks, CA: Sage Publishing.

Beattie, G., & Coughlan, J. (1999). An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. *British Journal of Psychology, 90*(1), 35–56.

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

Belhiah, H. (2009). Tutoring as an embodied activity: How speech, gaze, and body orientations are coordinated to conduct ESL tutorial business. J*ournal of Pragmatics, 41*(4), 829–841.

Belhiah, H. (2013). Using the hand to choreograph instruction: On the functional role of gesture in definition talk. *The Modern Language Journal, 97*(2), 417–434.

Bell, P. (2009). Le cadeau or la cadeau?: The role of aptitude in learner awareness of gender distinctions in French. *The Canadian Modern Language Review, 65*(4), 615–643.

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

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA, London: The MIT Press.

Best, C.T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171 – 232). Baltimore: York Press.

Best, C.T. & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. -S. Bohn & M. J. Munro (Eds.), *Language experience*

in second language speech learning: In honor of James Emil Flege (pp. 13 – 34). Amsterdam: John Benjamins Publishing.

Bhatara, A., Boll-Avetisyan, N., Höhle, B., & Nazzi, T. (2018). Early sensitivity and acquisition of prosodic patterns at the lexical level. In P. Prieto & N. Esteve-Gibert (Eds.), The development of prosody in first language acquisition (pp. 37–57). Amsterdam: John Benjamins Publishing Company.

Biau, E., Moris Fernandez, L., Holle, H., Avila, C. & Soto-Faraco, S. (2016). Hand gestures as visual prosody: BOLD responses to audio-visual alignment are modulated by the communicative nature of the stimuli. *Neuroimage, 132*, 129–137.

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

Billières, M. (2002). Le corps en phonétique corrective. In R. Renard (Ed.): *Apprentissage d'une langue étrangère seconde 2. La phonétique verbo-tonale* (pp. 35–70). Bruxelles: De Boeck Université.

Billières, M. (2017). *Corps, prosodie, travail phonétique* [PowerPoint slides]. Alliance Française di Padova. https://www.verbotonale-phonetique.com/geste-parole/

Blesedell, D. S. (1991). *A study of the effects of two types of movement instruction on the rhythm achievement and*

developmental rhythm aptitude of preschool children [doctoral dissertation, Temple University]. ProQuest Dissertations and Theses.

Blonder, L. X., Burns, A. F., Bowers, D., Moore, R. W., & Heilman, K. M. (1995). Spontaneous gestures following right hemisphere infarct. *Neuropsychologia, 33*(2), 203–213.

Bolinger, D. (1983). Intonation and Gesture. *American Speech*, *58*(2), 156–174.

Borghi, A.M., Glenberg, A., & Kaschak, M. (2004). Putting words in perspective. *Memory & Cognition, 32*, 863 – 873.

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature, 391*, 756.

Bowen, J. D. (1972). Contextualizing pronunciation practice in the ESOL classroom. *TESOL Quarterly, 6*(1), 83 –94.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics, 61*(5), 977 –985.

Brass, M., & Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences, 9*(10), 489–495.

Brice, M. (2003). *Pédagogie de Tous les Possibles. La Rythmique Jaques-Dalcroze*. Genève: Éditions Papillon.

Broaders, S., Cook, S.W.; Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture reveals implicit knowledge and leads to learning. *Journal of Experimental Psychology, 136*(4), 539–550.

Brodsky, W., & Sulkin, I. (2011). hand-clapping songs: A spontaneous platform for child development among 5-10-year-old children. *Early Child Development and Care*, *181*(8), 1111–1136.

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly, 22*(4), 593–606.

Burgess, J. (1994). Ideational frameworks in integrated language learning. *System, 22*(3), 309–318.

Burgess, J., & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System*, *28*(2), 191–215.

Burr D., & Gori M. (2012). Multisensory integration develops late in humans. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes* (Chapter 18). Boca Raton, FL: CRC Press/Taylor & Francis.

Burri, M., & Baker, A. (2016). Teaching rhythm and rhythm grouping: The butterfly technique. *English Australia Journal: the Australian journal of English language teaching, 31*(2), 72-77.

Burri, M. & Baker, A. (2019). "I never imagined" pronunciation as "such an interesting thing": Student teacher perception of

innovative practices. International *Journal of Applied Linguistics, 29*(1), 95-108.

Burri, M., Baker, A. & Acton, W. (2019). Proposing a Haptic Approach to Facilitating L2 Learners' Pragmatic Competence. *Humanising Language Teaching, 21*(3), 1-15.

Butcher, C., and Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: when hand and mouth come together. In D. McNeill (Ed.), *Language and gesture* (pp. 235–257). Cambridge: Cambridge University Press.

Cai, Z. G., & Connell, L. (2012). Space-time interdependence and sensory modalities: Time affects space in the hand but not in the eye. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 168-173). Austin, TX: Cognitive Science Society.

Cai, Z. G., Connell, L., & Holler, J. (2013). Time does not flow without language: Spatial distance affects temporal duration regardless of movement or direction. *Psychonomic Bulletin & Review, 20*(5), 973–980.

Calbris, G. (2003). *L'Expression gestuelle de la pensée d'un homme politique*. Paris: CNRS Éditions.

Calvo-Merino, B., Grezes, J., Glaser, D.E., Passingham, R.E., & Haggard, P. (2006). Seeing or doing? Influence of visual

and motor familiarity in action observation. *Current Biology, 16*(19), 1905–1910.

Cameron, H., & Xu, X. (2011). Representational gesture, pointing gesture, and memory recall of preschool children. *Journal of Nonverbal Behavior, 35*(2), 155–171.

Camponogara, I., & Volcic, R. (2021). Integration of haptics and vision in human multisensory grasping. *Cortex*, *135*, 173–185.

Capirci, O., Contaldo, A., Caselli, M. C., and Volterra, V. (2005). From action to language through gesture: a longitudinal perspective. *Gesture, 5*(1-2), 155–177.

Capirci, O., and Volterra, V. (2008). Gesture and speech: The emergence and development of a strong and changing partnership. *Gesture, 8*(1), 22–44.

Carpenter, M., & Call, J. (2009). Comparing the imitative skills of children and nonhuman apes. *Revue de primatologie, 1*, 263.

Carroll, J., Snowling, M., Hulme, C., & Stevenson, J. (2003). The development of phonological awareness in preschool children. *Developmental Psychology, 39*(5), 913–923.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*(2), 579–593.

Casasanto, D., Phillips, W., & Boroditsky, L. (2003). Do we think about music in terms of space? Metaphoric representation of musical pitch. In R. Alterman & D. Kirsch (Eds.),

Proceedings of the 25th Annual Conference of the Cognitive Science (p. 1323), Boston, USA.

Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: Rhythmic priming and audio-motor training affect speech perception. *Acta Psychologica*, *155*, 43–50.

Cason, N., Hidalgo, C., Isoard, F., Roman, S., & Schön, D. (2015). Rhythmic priming enhances speech production abilities: Evidence from prelingually deaf children. *Neuropsychology*, *29*(1), 102–107.

Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, *50*(11), 2652–2658.

Cassidy, J. W. (1993). Effects of various sight singing strategies on non music majors' pitch accuracy. *Journal of Research in Music Education*, *41*(4), 293–302.

Cebrian, J., & Carlet, A., (2014). Second language learners' identification of target language phonemes: A short-term phonetic training study. *Canadian modern language review, 70*(4), 474–499.

Cekaite, A. (2009). Soliciting teacher attention in an L2 classroom: Affect displays, classroom artefacts, and embodied action. *Applied Linguistics, 30*(1), 26–48.

Cekaite, A. (2015). The coordination of talk and touch in adults' directives to children: Touch and social control. *Research on Language and Social Interaction, 48*(2), 152–75.

Celce-Murcia, M. (2001). *Teaching English as a second or foreign Language (3rd edition)*. Boston: Heinle & Heinle Publisher.

Celce-Murcia, M., Brinton, D., Goodwin, J. ., & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide*. Cambridge: Cambridge University Press.

Chamberlin-Quinlisk, C. (2008). Nonverbal communication and second language classrooms: A review. In S.G. McCafferty & G. Stam (Eds), *Gesture: Second language acquisition and classroom research* (pp. 25–45). New York: Routledge.

Chan, M. J. (2018). Embodied pronunciation learning: Research and practice. *The Catesol Journal*, *30*(1), 47–68.

Chang, W. (2008). The role of L1 phonological feature in the L2 perception and production of vowel length contrast in English. *Speech Sciences*, *15*(1), 37–51.

Chang, Y. K., Labban, J. D., Gapin, J. I., & Etnier, J. L. (2012). The effects of acute exercise on cognitive performance: a meta-analysis. *Brain Research, 1453*, 87–101.

Chao, S- C., Ochoa, D., & Daliri, A. (2019). Production variability and categorical perception of vowels are strongly linked. *Frontiers in Human Neuroscience, 13*, 96.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception–behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893.

Chen, C. -M. (2013). Gestures as tone markers in multilingual communication. In I. Kecskes (Ed.), *Research in Chinese as*

a second language (pp. 143–168). Berlin, Boston: De Gruyter Mouton.

Cherdieu, M., Palombi, O., Gerber, S., Toccaz, J., & Rochet-Capellan, A. (2017). Make gestures to learn: Reproducing gestures improves the learning of anatomical knowledge more than just seeing gestures. *Frontiers in Psychology, 8*, 1689.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology*, *22*(1), 267–281.

Colantoni, L., Escudero, P., Marrero-Aguilar, V., & Steele, J. (2021). Evidence-based design principles for Spanish pronunciation teaching. *Frontiers in Communication*, *6*, 30.

Coles, K., & Tomporowski, P. D. (2008). Effects of acute exercise on executive processing, short-term and long- term memory. *Journal of Sports Sciences, 26*(3), 333–344.

Colletta, J-M., (2004). *Le développement de la parole chez l'enfant âgé de 6 à 11 ans. Corps, langage et cognition*. Wavre: Éditions Mardaga.

Colletta, J.-M., Guidetti, M., Capirci, O., Cristilli, C., Demire, O. E., Kunene, R. N., et al. (2015). Effects of ageake gestures to learn: Reproducing gestures improves the learning of anatomical knowledge more than just seeing gestures. and language on co-speech gesture production: an investigation

of French, American, and Italian children's narratives. *Journal of Child Language, 42*(1), 122–145.

Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., & Goldin-Meadow, S. (2017). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learning and Instruction, 50*, 65–74.

Connell, L., Cai, Z. G., & Holler, J. (2013). Do you see what I'm singing? Visuospatial movement biases pitch perception. *Brain and Cognition*, *81*(1), 124–130.

Cook, S.W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106*(2), 1047– 1058.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes, 27*(4), 594–610.

Cortés Moreno, M. (1997). Sobre la percepción y adquisición de la entonación española por parte de hablantes nativos de chino. *Estudios de Fonética Experimental, 5*, 67–134.

Couper, G. (2011). What makes pronunciation teaching work? Testing for the effect of two variables: socially constructed metalanguage and critical listening. *Language Awareness, 20*(3), 159–182.

Craig, R. J., & Amernic, J. H. (2006). PowerPoint presentation technology and the dynamics of teaching. *Innovation in Higher Education, 31*, 147–160.

Crowder, E. (1996). Gestures at work in sense-making science talk. *Journal of the Learning Sciences, 5*(3), 173–208.

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *Modern Language Journal*, *99*(1), 80–95.

Crumpler, S. E. (1982). *The effects of Dalcroze Eurhythmics on the melodic musical growth of first grade students* [doctoral dissertation, Louisiana State University]. ProQuest Dissertations and Theses.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America, 111*(6), 2862–2873.

Dabène, L. (1984). Pour une taxinomie des opérations méta-communicatives en classe de langue étrangère. *Études de Linguistiques Appliquée, 55*, 39–46.

Danner, S. G., Barbosa, A. V., & Goldstein, L. (2018). Quantitative analysis of multimodal speech data. *Journal of Phonetics, 71*, 268–283.

Darcy, I., Ewert, D., & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: an ESL teachers'

pronunciation “toolbox.” In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 93–108). Ames, IA: Iowa State University.

Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, *40*, 63–72.

Dargue, N., & Sweller, N. (2018a). Donald Duck’s garden: the effects of observing iconic reinforcing and contradictory gestures on narrative comprehension. *Journal of Experimental Child Psychology, 175*, 96–107.

Dargue, N., & Sweller, N. (2018b). Not all gestures are created equal: the effects of typical and atypical iconic gestures on narrative comprehension. *Journal of Nonverbal Behavior, 42*(3), 327–345.

Dargue, N., & Sweller, N. (2020a). Learning stories through gesture: Gesture’s effects on child and adult narrative comprehension. *Educational Psychology Review*, *32*(1), 249–276.

Dargue, N., & Sweller, N. (2020b). Two hand and a tale: When gestures benefit adult narrative comprehension. *Learning and Instruction, 68*, 101331.

Dargue, N., Sweller, N., and Jones, M. P. (2019). When our hands help us understand: a meta-analysis into the effects of

gesture on comprehension. *Psychological Bulletin, 145*(8), 765–784.

Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*(1), 51–62.

de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech, 26*(4), 331-350.

Decety, J., & Grezes, J. (2006). The power of simulation: Imagining one's own and other's behaviour. *Cognitive Brain Research, 1079*(1), 4–14.

Degerman, A., Rinne, T., Salmi, J., Salonen, O., & Alho, K. (2006). Selective attention to sound location or pitch studied with fMRI. *Brain Research, 1077*(1), 123–134

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(5), 893–916.

de la Mota, C. (2019). Improving non-native pronunciation: teaching prosody to learners of Spanish as a second/foreign language. In Rao, R. (Ed.), *Key issues in the teaching of Spanish pronunciation* (pp. 226–254). New York: Routledge.

Delvaux, V., Demolin, D., & Soquet, A. (2004, April). Interactions mimétiques entre locuteurs: une étude expérimentale. *Actes*

Des XXV^e Journées d'Étude Sur La Parole (pp. 153–15). Fès, Maroc.

de Mareüil, P. B., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica*, *63*(4), 247–267.

de Nooijer, J. A., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of imitating gestures during encoding or during retrieval of novel verbs on children's test performance. *Acta Psychologica*, *144*(1), 173–179.

de Ruiter, J. P. (2017). The asymmetric redundancy of gesture and speech. In Church, R.B., Alibali, M.W., & Kelly, S.D. (Eds.), *Why gesture? How the hands function in speaking, thinking and communicating* (pp. 59–75). Amsterdam: John Benjamins Publishing Company.

Derwing, T. (2010). Utopian goals for pronunciation teaching. In J. M. Levis, & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference* (pp. 24–37). Iowa State University Print.

Derwing, T. M., Diepenbroek, L. G., & Foote, J.A. (2012). How well do general skills ESL textbooks address pronunciation? *TESL Canada Journal, 30*(1), 22–44.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(4), 476–490.

Derwing, T. M. & Munro, M. J. (2015). *Pronunciation fundamentals*. Amsterdam: John Benjamins Publishing Company.

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

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54*(4), 655–680.

Desai, R. H., Binder, J. R., Conant, L. L., & Seidenberg, M.S. (2010). Activation of sensory-motor areas in sentence comprehension. *Cerebral Cortex*, *20*(2), 468–478.

Di Cristo, A., & Hirst, D. (1993). Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français. *Travaux de l'Institut de Phonétique d'Aix*, *15*, 9–24.

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

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2012). The sound of thickness: Prelinguistic infants' associations of space and pitch. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the*

Cognitive Science Society (pp. 306–311). Austin, TX: Cognitive Science Society.

Dolscheid, S., Willems, R. M., Hagoort, P., & Casasanto, D. (2014). The relation of space and musical pitch in the brain. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 421–426). Quebec, Canada.

Donald, M. (1993). *Origins of the modern mind: Three stages in the evolution of culture & cognition* (Reprint). Cambridge, MA: Harvard University Press.

Donnelly, J. E., Hillman, C. H., Castelli, D., Etnier, J. L., Lee, S., Tomporowski, P., et al. (2016). Physical activity, fitness, cognitive function, and academic achievement in children: a systematic review. *Medicine & Science in Sports and Exercise, 48*(6), 1197– 1222.

Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–11). Clevedon: Multilingual Matters.

Drew, Paul (2006). When documents 'speak': Documents, language and interaction. In P. Drew, G. Raymond, & D. Weinberg (Eds.), *Talk in interaction in social research methods* (pp. 63–80). Thousand Oaks, CA: Sage Publications.

Drijvers, L., and Özyürek, A. (2018). Native language status of the listener modulates the neural integration of speech and

iconic gestures in clear and adverse listening conditions. *Brain and Language, 177–178*, 7–17.

Drijvers, L., Vaitonytė, J., and Özyürek, A. (2019). Degree of language experience modulates visual attention to visible speech and iconic gestures during clear and degraded speech comprehension. *Cognitive Science, 43*(10), e12789.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., and Peperkamp, S. (2008). Persistent stress 'deafness': the case of French learners of Spanish. *Cognition, 106*(2), 682–706.

Ekman, P., & Friesen, W.V., (1969). The repertoire of nonverbal behavioral categories: Origins, usage and coding, *Semiotica, 1*(1), 49–98.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning. 54*(2), 227-275.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*(2), 141-172.

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221–246.

Elvin, J., and Escudero, P. (2019). Cross-linguistic influence in second language speech: implications for learning and teaching. In M. J. Gutierrez-Mangado, M. Martínez-Adrián, & F. Gallardo-del-Puerto (Eds.), *Cross-linguistic Influence:*

from Empirical Evidence to Classroom Practice (pp. 1–20). New York: Springer.

Engel, A. K., Maye, A., Kurthen, M., and Konig, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences, 17*(5), 202–209.

Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., and Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology, 110*(4), 659–675.

Erickson, K. I., Hillman, C. H., & Kramer, A. F. (2015). Physical activity, brain, and cognition. *Current Opinion in Behavioral Sciences, 4*, 27–32.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* [doctoral dissertation, Utrecht University]. LOT Publications.

Escudero, P., and Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition, 133(2)*, 408–413.

Eskildsen, S. W., & Wagner. J. (2013). Recurring and shared gestures in the L2 classroom: Resources for teaching and learning. *European Journal of Applied Linguistics, 1*(1), 139–161.

Eskildsen, S. W., & Wagner, J. (2015). Embodied L2 construction learning. *Language Learning*, *65*(2), 268–297.

Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., & Prieto, P. (2017). The timing of head movements: The role of prosodic heads and edges. *The Journal of the Acoustical Society of America, 141*(6), 4727–4739.

Esteve-Gibert, N., & Guellaï, B. (2018). Prosody in the auditory and visual domains: A developmental perspective. *Frontiers in Psychology, 9,* 338.

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research, 56*(3), 850–864.

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

Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: self-performing during recognition. *Memory & Cognition*, *22*(1), 34–39.

Esteve-Gibert, N., Suárez, M.M., Vasylets, O., Feijoo, S., Serrano, R. (2021). The children's use of tactile and visual information (vs. acoustic information) when learning non-native phonological contrasts. *Workshop "First and second language acquisition of phonology and its interfaces", 13th Old World Conference on Phonology*. Eivissa

Falk, S., & Dalla Bella, S. (2016). It is better when expected: aligning speech and motor rhythms enhances verbal processing. *Language, Cognition and Neuroscience*, *31*(5), 699–708.

Falk, S., Lanzilotti, C., & Schön, D. (2017). Tuning neural phase entrainment to speech. *Journal of Cognitive Neuroscience*, *29*(8), 1378–1389.

Falk, S., Volpi-Moncorger, C., & Dalla Bella, S. (2017). Auditory-motor rhythms and speech processing in French and German listeners. *Frontiers in Psychology, 8*, 395.

Farina, M. (2021). Embodied cognition: dimensions, domains and applications. *Adaptive Behavior*, 29(1), 73–88.

Fedewa, A. L., & Ahn, S. (2011). The effects of physical activity and physical fitness on children's achievement and cognitive outcomes: a meta-analysis. *Research Quarterly for Exercise and Sport, 82*(3), 521–535.

Ferré, G. (2010, May). Relations temporelles entre parole et gestualité co-verbale en français spontané. XXVIII[e] *Journées d'Étude Sur La Parole* (pp. 1–4). Mons, Belgique.

Field, J. (2005). Intelligibility and the listener: the role of lexical stress. *TESOL Quarterly, 39(3)*, 399–423.

Fincher-Kiefer, R. (2019). *How the body shapes knowledge: Empirical support for embodied cognition*. Washington: American Psychological Association.

Fischler, J. (2009). The rap on stress: Teaching stress patterns to English language learners through rap music. *MInneTESOL Journal*, *26,* 1–23.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*(1), 47–65.

Flege, J.E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders, a review* (pp. 224–401). New York: Ablex.

Flege, J. E. (1992). Speech learning in a second language. In C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.). *Phonological development: Models, research, implications* (pp. 565-604). Timonium, MD: York Press.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore: York Press.

Flege, J. E. (1999, August). Relation between L2 production and perception. In J. Ohala et al. (Eds.), *Proceedings of the XIVth International Congress of Phonetics Sciences* (pp. 1273–1276). Berkeley, CA, USA.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In Burmeister, P., Piske, T. & A. Rohde (Eds.). *An integrated view of language development : Papers in honor of Henning Wode* (pp. 217-243). Trier: Wissenschaftlicher Verlag.

Flege, J., and Bohn, O. (2021). The revised speech learning model (SLM-r). In Wayland, R. (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge: Cambridge University Press.

Flege, J., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*(4), 527-552.

Foote, J.A., Holtby, A.K., & Derwing, T.M. (2012). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal, 29*, 1-22.

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(3), 319–325.

Forster, B., Cavina-Pratesi, C., Aglioti, S., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research, 143*(4), 480–487.

Forsythe, J. L., & Kelly, M. M. (1989). Effects of visual-spatial added cues on fourth-graders' melodic discrimination. *Journal of Research in Music Education, 37*(4), 272-277.

Foster, N. E. V., & Zatorre, R. J. (2010). Cortical structure predicts success in performing musical transformation judgments. *NeuroImage, 53*(1), 26–36.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. J*ournal of Phonetics, 36*(2), 268–294.

Freed, B. F. (1995). Do students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Amsterdam: John Benjamins.

Frick-Horbury, D. (2002). The effects of hand gestures on verbal recall as a function of high- and low-verbal-skill levels. *Journal of General Psychology, 29*(2), 137–147.

Fu, Y., & Franz, E. A. (2014). Viewer perspective in the mirroring of actions. *Experimental Brain Research, 232*(11), 3665–3674.

Gabriel, C., & Kireva, E. (2014). Prosodic transfer in learner and contact varieties: Speech rhythm and intonation of Buenos Aires Spanish and L2 Castilian Spanish produced by Italian native speakers. *Studies in Second Language Acquisition, 36*(2), 257–281.

Gallese, V. (2005). Embodied simulation: from neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences, 4*, 23–48.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3–4), 455–479.

Gallimore, R., & Tharp, R. (1981). The interpretation of elicited imitation in a standardized context. *Language Learning, 31*(2), 369-392.

Gardner, R. C., (2010). *Motivation and second language acquisition: The socio-educational model*. New York: Peter Lang.

Gennari, S. P. (2012). Representing motion in language comprehension: lessons from neuroimaging. *Language and Linguistics Compass, 6*(2), 67–84.

Ghaemi, F., & Rafi, F. (2018). The impact of visual aids on the retention of English word stress patterns. *International Journal of Applied Linguistics & English Literature*, *7*(2), 225–231.

Gallagher, S. (2005). *How the body shapes the mind*. Oxford: Oxford University Press.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*(2), 593–609.

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly, 22*(3), 473–492.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to

fluency. *Canadian Modern Language Review, 61*(3), 325–353.

Gentilucci M, & Dalla Volta R. (2008). Spoken language and arm gestures are controlled by the same motor control system. *Quarterly Journal of Experimental Psychology, 61*(6), 944–957.

Gibbs, R. W. (2006). *Embodiment and cognitive science.* Cambridge: Cambridge University Press.

Gilbert, J. (2001a). *Clear speech from the start*. Cambridge: Cambridge University Press.

Gilbert, J. (2001b). Six pronunciation priorities for the beginning student. *The CATESOL Journal, 13*, 173-182.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context and consequence. In Giles, H., Coupland, J. & Coupland N. (Eds.), *Contexts of accommodation* (pp. 1–68). Cambridge University Press & Editions de la Maison des Sciences de l'Homme.

Gili Fivela, B. (2012). Testing the perception of L2 intonation. In M. Grazia Busà & A. Stella, *Methodological Perspectives on Second Language Prosody*. *Papers from ML2P 2012* (pp. 17-30). Padova: CLEUP.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology, 96*(3), 424–436.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558–565.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, *26*(1), 69–83.

Glenberg, A.M., & Robertson, D.A. (1999). Indexical understanding of instructions. *Discourse Processes, 28*(1), 1–26.

Glenberg, A.M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., & Buccino, G. (2008). Processing abstract language modulates motor system activity. *Quarterly Journal of Experimental Psychology, 61*(6), 905-919.

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

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesture lightens the load. *Psychological Science, 12*(6), 516–522.

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

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

Goldin-Meadow, S. (2010). When gesture does and does not promote learning. *Language and Cognition*, *2*(1), 1–19.

Goldin-Meadow, S. (2011). Learning through gesture. *Wiley Interdisciplinary Reviews. Cognitive Science*, *2*(6), 595–607.

Goldin-Meadow, S. (2014). Widening the lens: what the manual modality reveals about language, learning and cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(20130295), 1–11.

Goldin-Meadow, S. (2018). Taking a hands-on approach to learning. *Policy Insights from the Behavioral and Brain Sciences, 5*(2), 163–170.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, *64*, 257–283.

Goldin-Meadow, S., Cook, S.W., Mitchell, Z.A. (2009). Gesturing gives children new ideas about math. *Psychological Science, 20*(3), 267–272.

Goldin-Meadow, S., Kim, S., Singer, M. (1999). What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology, 91*(4), 720–730.

Goldin-Meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S. W, & Small, S. L. (2014). New evidence about language and cognitive development based on a longitudinal study: hypotheses for intervention. *American Psychologist, 69*(6), 588–599.

Goldin-Meadow, S., & Wagner, S. (2005). How our hands help us learn. *Trends in Cognitive Sciences, 9*(5), 234–240.

Goldman A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.

Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language switching makes pronunciation less nativelike. *Psychological Science*, *25*(4), 1031–1036.

Good, A. J., Russo, F. A., & Sullivan, J. (2015). The efficacy of singing in foreign-language learning. *Psychology of Music*, *43*(5), 627–640.

Goodin-Mayeda, E. (2019). The role of perception in learning Spanish pronunciation. In Rao, R. (Ed.), *Key issues in the teaching of Spanish pronunciation* (pp. 226–254). New York: Routledge.

Goodwin, C., & Goodwin, M. (1986). Gesture and co-participation in the activity of searching for a word. *Semiotica, 62*, 51–75.

Gordon, J. (2021). Pronunciation and Task-Based Instruction: Effects of a classroom intervention. *RELC Journal*, *52*(1), 94–109.

Gordon, J., Darcy, I., & Ewert, D. (2013). Pronunciation teaching and learning: Effects of explicit phonetic instruction in the L2 classroom. In J. M. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language*

Learning and Teaching Conference (pp. 194-206). Ames, IA: Iowa State University.

Gordon, R. L., Jacobs, M. S., Schuele, C. M., & Mcauley, J. D. (2015). Perspectives on the rhythm-grammar link and its implications for typical and atypical language development. *Annals of the New York Academy of Sciences*, *1337*(1), 16–25.

Gori, M., Del Viva, M. M., Sandini, G., Burr, D.C. (2008). Young children do not integrate visual and haptic form information. *Current Biology, 18*(9), 694–698.

Gorjian, B., Hayati, A., & Pourkhoni, P. (2013). Using Praat software in teaching prosodic features to EFL learners. *Procedia - Social and Behavioral Sciences*, *84*, 34–40.

Grabe, E., and Low, L. (2002). Acoustic correlates of rhythm class. In C. Gussenhoven and N. Warner (Eds.), *Laboratory phonology 7* (pp. 515–546). Berlin, New York: Mouton de Gruyter.

Graziano, M., & Gullberg, M. (2018). When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology*, *9*, 879.

Greenfield, K., Ropar, D., Themelis, K., Ratcliffe, N., & Newport, R. (2017). Developmental changes in sensitivity to spatial and temporal properties of sensory integration underlying body representation. *Multisensory Research*, *30*(6), 467–484.

Greenhead, K., & Habron, J. (2015). The touch of sound: Dalcroze eurhythmics as a somatic practice. *Journal of Dance and Somatic Practices*, *7*(1), 93–112.

Guberina, P. (1956). L'audiométrie verbo-tonale at son application. *Journal Français d'O.R.L., 6*, 23–42.

Guberina, P. (1961). La méthode audio-visuelle structuro-globale et ses implications dans l'enseignement de la phonétique. *Studia Romanica et Anglica Zagrabiensia, 11*, 12–40.

Guberina, P. (2008). *Retrospección* (J. Murillo, ed. & trans.). Éditions du CIPA - Asociación Española Verbotonal.

Guenther, F., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105*(4), 611–633.

Gullberg, M. (1998). *Gesture as a communication strategy in second language discourse: a study of learners of French and Swedish*. Lund: Lund University Press.

Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (Hommage to Adam Kendon). *IRAL - International Review of Applied Linguistics in Language Teaching, 44*(2), 103–24.

Gullberg, M., & McCafferty, S. G. (2008). Introduction to gesture and SLA: Toward an integrated approach. *Studies in Second Language Acquisition*, *30*(2), 133–146.

Gullberg, M., Roberts, L., & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal

exposure to a new language? *IRAL - International Review of Applied Linguistics in Language Teaching*, *50*(4), 239–276.

Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., & Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning*, *60*(2), 5–24.

Gut, U., & Pillai, S. (2014). Prosodic marking of information structure by Malaysian speakers of English. *Studies in Second Language Acquisition, 36*(2), 283–302.

Hama, M., Leow, R. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition, 32*(3), 465-491.

Hamada, Y. (2018). Shadowing for pronunciation development: Haptic-shadowing and IPA-shadowing. *Journal of Asia TEFL*, *15*(1), 167–183.

Hannah, B., Wang, Y., Jongman, A., Sereno, J. A., Cao, J., & Nie, Y. (2017). Cross-modal association between auditory and visuospatial information in Mandarin tone perception in noise by native and non-native perceivers. *Frontiers in Psychology*, *8*, 2051.

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics, 40*(2), 269–279.

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

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, *8*(1), 34–52.

Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics, 26*(4), 579–596.

Harwood, E., (1993). Content and context in children's playground songs. *Applications of Research in Music Education*, *12*(1), 4-8.

Haught, J. R., & McCafferty, S. G. (2008). Embodied language performance: Drama and the ZPD in the second language classroom. In J. P. Lantolf & M. E. Poehner (Eds.), *Sociocultural theory and the teaching of second languages* (pp. 139– 162). Sheffield: Equinox.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307.

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

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America, 119*(3), 1740–1751.

He, X. L. , van Heuven, V. J. , & Gussenhoven, C. (2012). The selection of intonation contours by Chinese L2 speakers of Dutch: Orthographic closure vs. prosodic knowledge. *Second Language Research, 28*(3), 283–318.

Helfer, K. S. Helfer, & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *The Journal of the Acoustical Society of America 117*(2), 842–849.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology, 63*(3), 289–293.

Heyes, C. (2011). Automatic imitation. *Psychological Bulletin, 137*(3), 463–483.

Hillman, C. H., Erickson, K. I., & Kramer, A. F. (2008). Be smart, exercise your heart: exercise effects on brain and cognition. *Nature Reviews Neuroscience, 9*(1), 58–65.

Hincks, R., & Edlund, J. (2009). Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning and Technology*, *13*(3), 32–50.

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

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

Hoetjes, M., van Maastricht, L., & van der Heijden, L. (2019). Gestural training benefits L2 phoneme acquisition: Findings from a production and perception perspective. *Proceedings of the 6th Gesture and Speech in Interaction Conference* (pp. 50-55). Paderborn, Germany.

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

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137*(2), 297–315.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514.

Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin and Review*, *26*(3), 721–752.

Huang, B. H. & Jun, S-A (2011). The effect of age on the acquisition of second language prosody. *Language and Speech, 54*(3), 387–414.

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

Hübscher, I., & Prieto, P. (2019). Gestural and prosodic development act as sister systems and jointly pave the way for children's sociopragmatic development. *Frontiers in Psychology*, *10*, 1259.

Hudson, N. (2011). *Teacher gesture in a post-secondary English as a second language classroom: A sociocultural approach* [doctoral dissertation, University of Nevada]. UNLV Theses, Dissertations, Professional Papers, and Capstones.

Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: Saccades versus manual responses. *Journal of Experimental Psychology: Human Perception and Performance, 20*(1), 131–153.

Hutto, D., Kirchhoff, M., & Abrahamson, D. (2015) The enactive roots of STEM: Rethinking educational design in mathematics. *Educational Psychology and Review, 27*(3), 371–389.

Ibáñez, A., Manes, F., Escobar, J., Trujillo, N., Andreucci, P., & Hurtado, E. (2010). Gesture influences the processing of figurative language in non-native speakers: ERP evidence. *Neuroscience Letters*, *471*(1), 48–52.

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

Iizuka, T., Nakatsukasa, K., & Braver, A. (2020). The efficacy of gesture on second language pronunciation: An exploratory study of hand-clapping as a classroom instructional tool. *Language Learning*, *70*(4), 1054–1090.

Inceoglu. S. (2016). Effects of perceptual training on L2 vowel perception and production. *Applied Psycholinguistics, 37*(5), 1175–1199.

Intravaia, P. (2000). *Formation des professeurs de langue en phonétique corrective. Le système verbo-tonal*. Paris: Didier Érudition.

Ionescu, T., & Vasc, D. (2014). Embodied Cognition: Challenges for psychology and education. *Procedia - Social and Behavioral Sciences*, *128*, 275–280.

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*(2), 135–159.

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility. *Studies in Second Language Acquisition*, *34*(3), 475–505.

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

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies, 6*(11–12), 19–40.

James, K. H., and Maouene, J. (2009). Auditory verb perception recruits motor systems in the developing brain: an fMRI investigation. *Developmental Science, 12*(6), F26–F34.

Jaques-Dalcroze E. (1920). *Le Rythme, la musique et l'éducation*. Paris: Librairie Fischbacher

Jenner, B. (1989). Teaching pronunciation: The common core. *Speak Out! 4*, 2–4.

Joseph, A. (1982). *A Dalcroze eurhythmics approach to music learning in kindergarten through rhythmic movement, ear-training and improvisation* [doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA]. ProQuest Dissertations and Theses.

Jakonen, T. (2020). Professional Embodiment: Walking, re-engagement of desk interactions, and provision of instruction during classroom rounds. *Applied Linguistics*, *41*(2), 161–184.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*(4), 548–567.

Jun, S.–A. (2005). Prosodic typology. In S. A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430–458). Oxford: Oxford University Press.

Jung, Y., Kim, Y., & Murphy, J. (2017). The role of task repetition in learning word-stress patterns through auditory priming tasks. *Studies in Second Language Acquisition*, *39*(2), 319–346.

Jungheim, N.O. (1995). *Assessing nonverbal ability as a component of language learners' communicative competence* [doctoral dissertation, Temple University]. ProQuest Dissertations and Theses.

Juntunen, M.-L. (2016). The Dalcroze approach : Experiencing and knowing music through the embodied exploration. In C. R. Abril & B. Gault (Eds.), *Approaches to teaching general music: Methods, issues, and viewpoints* (pp. 141–167). Oxford: Oxford University Press.

Juntunen, M.-L., & Hyvönen, L. (2004). Embodiment in musical knowing: how body movement facilitates learning within Dalcroze Eurhythmics. *British Journal of Music Education*, *21*(2), 199–214.

Kääntä L. (2015). The Multimodal organisation of teacher-led classroom interaction. In: Jenks C.J., & Seedhouse P. (Eds).

International perspectives on ELT classroom interaction. London: Palgrave Macmillan.

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and language, 192*, 15–24.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, *38*(2), 301–315.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal*, *94*(4), 554–566.

Keating, P., Baroni, M., Mattys, S., Scarborough, R., Alwan, A., Auer, E. T., et al. (2003, August). Optical phonetics and visual perception of lexical and phrasal stress in english. In *Proceedings of the ICPHS conference* (pp. 2071–2074). Barcelona, Spain.

Keetman, G., & Orff, C. (1963). *Music for children (Orff-Schulwerk)*. Mainz: Schott Musik International.

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

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

Kelly, S. D., & Hirata, Y. (2017). What neural measures reveal about foreign language learning of Japanese vowel length contrasts with hand gestures. In S. Tanaka (Ed.), *New development in phonology research: Festschrift in honor of Haruo Kubozono* (pp. 278–294). Tokyo: Kaitakusha.

Kelly, S. D., Hirata, Y., Manansala, M., & Huang, J. (2014). Exploring the role of hand gestures in learning novel phoneme contrasts and vocabulary in a second language. *Frontiers in Psychology*, *5*, 673.

Kelly, S. D., Manning, S. M., & Rodak, S. (2008). Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education. *Language and Linguistics Compass, 2*(4), 569– 588.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21*(2), 260–267.

Kempe, V., Bublitz, D., & Brooks, P.J. (2015). Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. *British Journal of Psychology, 106*(2), 349–366.

Kendon, A. (1980). Gesticulation and speech: two aspects of the process of utterance. In M. R. Key (Ed.), *The Relationship of verbal and nonverbal communication* (pp. 207–227). Berlin, New York: De Gruyter Mouton.

Kendon, A. (1982). The study of gesture: some observations on its history. *Recherches Sémiotiques/Semiotic Inquiry, 2*(1), 45–62.

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

Kendon, A. (2017). Pragmatic functions of gestures. *Gesture*, *16*(2), 157–175.

Kennedy, S., Blanchet, J., Trofimovich, P. (2014). Learner pronunciation, awareness, and instruction in French as a second language. *Foreign Language Annals, 47*(1), 79-96.

Kennedy, S., & Trofimovich, P. (2010). Language awareness and second language pronunciation: a classroom study. *Language Awareness*, *19*(3), 171–185.

Kern, F. (2018). Clapping hands with the teacher: What synchronization reveals about learning. *Journal of Pragmatics*, *125*, 28–42.

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

Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanism of

phonetic imitation. *Language, Cognition, and Neuroscience, 34*(6), 769–786.

Kiriloff. (1969). On the auditory perception of tones in Mandarin. *Phonetica*, *20*(2-4), 63–67.

Kirk, E., & Lewis, C. (2017). Gesture facilitates children's creative thinking. *Psychological Science, 28*(2), 225–232.

Kita, S. (1993). Japanese adults' development of English speaking ability: Change in the language-thought process observed through spontaneous gesture. *Paper presented at the Second language research forum*. University of Pittsburgh.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture: Window into thought and action* (pp. 162–185). Cambridge: Cambridge University Press.

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review, 124*(3), 245–266.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*(1), 16–32.

Kivistö-de Souza, H. (2016). *Phonological awareness and pronunciation in a second language* [doctoral dissertation, Universitat de Barcelona]. Tesis Doctorals en Xarxa.

Klaver, P., Fell, J., Dietl, T., Schür, S., Schaller, C., Elger, C. E., & Fernández, G. (2005). Word imageability affects the hippocampus in recognition memory. *Hippocampus, 15*(6), 704–712.

Klein, L. (2010). Phonetic correction in class with verbo-tonal method. *Studies in Language and Literature, 30*(1), 35–55.

Kohler, K. (2009). Rhythm in speech and language. A new research paradigm. *Phonetica, 66*(1-2), 29–45

Kondo, A. (2012). Phonological memory and L2 pronunciation skills. *In A. Stewart & N. Sonda (Eds.), JALT2011 Conference Proceedings* (pp. 535–541). Tokyo: JALT.

Kontra, C., Goldin-Meadow, S., and Beilock, S. L. (2012). Embodied learning across the lifespan. *Topics in Cognitive Science, 4*, 731–739.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*(2), 145–164.

Kotz, S. A., & Gunter, T. C. (2015). Can rhythmic auditory cuing remediate language-related deficits in Parkinson's disease? *Annals of the New York Academy of Sciences*, *1337*(1), 62–68.

Krahmer, E., Ruttkay, Z., Swerts, M., and Wesselink, W. (2002, April). Pitch, eyebrows and the perception of focus. In *Proceedings of the Speech Prosody 2002,* (pp. 443–446). Aix- en-Provence, France.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*(3), 396–414.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.

Krashen, S. D. and T. D. Terrell (1983). *The natural approach*. Phoenix: Phoenix Elt.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*(2), 54-60.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: a process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge: Cambridge University Press.

Kreiner, H., & Eviatar, Z. (2014). The missing link in the embodiment of syntax: Prosody. *Brain and Language*, *137*, 91–102.

Krivokapić,, J. (2014). Gestural coordination at prosody boundaries and its role for prosodic structure and speech planning processes. *Philosophical Transactions of the Royal Society London B, Biological Sciences, 369*(1658), 20130397.

Krivokapić, J., Tiede, M. K., & Tyrone, M. E. (2017). A kinematic study of prosodic structure in articulatory and manual gestures: Results from a novel method of data collection. *Laboratory Phonology, 8*(1), 1-36.

Krönke, K. M., Mueller, K., Friederici, A. D., & Obrig, H. (2013). Learning by doing? The effect of gestures on implicit retrieval of newly acquired words. *Cortex*, *49*(9), 2553–2568.

Kuhl, P. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics, 50*, 93–107.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception : The native language magnet theory. In de Boysson-Bardies, B. (Ed.). *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–74). Dordrecht: Kluwer Academic Publishers.

Kuhl, P. K., & Iverson, P. 1995. Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121-154). York: York Press.

Kushch, O. (2018). *Beat gestures and prosodic prominence: Impact on learning* [doctoral dissertation, Universitat Pompeu Fabra]. Tesis Doctorals en Xarxa.

Kushch, O., & Prieto, P. (2016). The effects of pitch accentuation and beat gestures on information recall in contrastive discourse. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of the 8th Speech Prosody Conference* (pp. 922–925). Boston, USA.

Kushch, O., Igualada, A., & Prieto, P. (2018). Gestural and prosodic prominence favor second language novel word acquisition. *Language and Cognitive Processes*, *33*(8).

Kutas, M., & Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 4*(12), 463–470.

Laakso, A. (2011). Embodiment and development in cognitive science. *Cognition, Brain, Behavior, 15*, 409–425.

Ladd, D. R. (2008). *Intonational phonology* (2nd edition). Cambridge: Cambridge University Press.

Lakin, J. L., Jefferis, V. E., Cheng, C. M., &, Chartrand, T. L., (2003). The chameleon effect as social glue: evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior, 27(3)*, 145–162.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Larson-Hall, J. (2008) Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24*, 35–63.

Lazaraton, A. (2004). Gesture and speech in the vocabulary explanations of one ESL teacher: A micro-analytic inquiry. *Language Learning, 54*, 79–117.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. Language Learning, 40, 387–417.

Levantinou, E. I., & Navarretta, C. (2015). An investigation of the effect of beat and iconic gestures on memory recall in L2 speakers. In *Proceedings from the 3rd European Symposium on Multimodal Communication* (pp. 32–37). Dublin.

Lee, B. J. (2020). Enhancing listening comprehension through kinesthetic rhythm training. *RELC Journal*, *OnlineFirst*. https://doi.org/10.1177/0033688220941302

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345–366.

Lee, B., Plonsky, L., & Saito, K. (2020). The effects of perception- vs. production-based pronunciation instruction. *System*, *88*, 182185.

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes, 26*(10), 1457–1471.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369*(1651): 20130302.

Levis, J.M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*, 369–378.

Levis, J.M. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation.* Cambridge University Press.

Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences, 13*(11), 470-478.

Li, P., Baills, F. & Prieto, P. (2020). Observing and producing durational hand gestures facilitates the pronunciation of novel vowel-length contrasts. *Studies in Second Language Acquisition, 42*(5), 1015–1039.

Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition*, *39*(4), 593–620.

Li, A., and Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm. *Studies in Second Language Acquisition, 36*(2), 223–255.

Li, Y., & Somlak, T. (2017). The effects of articulatory gestures on L2 pronunciation learning: A classroom-based study. *SAGE Journals, 23*(3), 325-337.

Li, P., Xi, X., Baills, F. & Prieto, P. (2021). Training non-native aspirated plosives with hand gestures: Learners' gesture performance matters. *Language Cognition and Neuroscience, FirstOnline*. https://doi.org/10.1080/23273798.2021.1937663

Liang, J. & van Heuven, V. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. In C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes, Experimental Studies in Word and Sentence Prosody* (pp. 27-61). Berlin, New York: Mouton de Gruyter.

Liu-Ambrose, T., Nagamatsu, L. S., Voss, M. W., Khan, K. M., & Handy, T. C. (2012). Resistance training and functional plasticity of the aging brain: a 12-month randomized controlled trial. *Neurobiology of Aging, 33*, 1690–1698.

Liszkowski, U. (2008). Before L1: a differentiated perspective on infant gestures. *Gesture, 8*, 180–196.

Llanes-Corominas, J., Prieto, P., & Rohrer, P. L. (2018). Brief training with rhythmic beat gestures helps L2 pronunciation in a reading aloud task. *Proceedings of the 9th Speech Prosody Conference* (pp. 498–502). Poznań, Poland.

Llanes-Coromina, J., Vilà-Giménez, I., Kushch, O., Borràs-Comes, J., & Prieto, P. (2018). Beat gestures help preschoolers

recall and comprehend discourse information. *Journal of Experimental Child Psychology, 172*(8), 168– 188.

Llorca, R. (2001). Jeux de groupe avec la voix et le geste sur les rythmes du français parlé. In : J. Johnston (Ed.), *L'enseignement des langues aux adultes, aujourd'hui : une pratique de la pédagogie pour une pédagogie de la pratique* (pp. 141-150). Université de Saint-Étienne.

Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture, 7*(2), 179–214.

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

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign Language Research in Cross-Cultural Perspective* (pp. 39-52). Amsterdam: John Benjamins Publishing Company.

López-Ozieblo, R. (2020). Proposing a revised functional classification of pragmatic gestures. *Lingua*, *247*, 102870.

Lorås, H., Stensdotter, A. K., Öhberg, F., & Sigmundsson, H. (2013). Individual differences in motor timing and its relation to cognitive and fine motor skills. *PLoS ONE*, *8*(7), e69353.

Lord, G. (2005). (How) Can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania, 88*, 557-567.

Loukina, A., Kochanski, G., Rosner, B., Keane, E., and Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech. *Journal of the Acoustical Society of America, 129*(5), 3258–3270.

Lucero, C., Zaharchuk, H., & Casasanto, D. (2014). Beat gestures facilitate speech production. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 898–903). Austin, TX. Cognitive Science Society.

Ludke, K. M. (2018). Singing and arts activities in support of foreign language learning: An exploratory study. *Innovation in Language Learning and Teaching*, *12*(4), 371–386.

Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. *Memory & Cognition*, *42*(1), 41–52.

Luo, J., Li, V. G., & Mok, P. P. K. (2020). The Perception of Cantonese Vowel Length Contrast by Mandarin Speakers. *Language and Speech*, *63*(3), 635–659.

Lyster, R. (2007). *Learning and Teaching Languages Through Content*. John Benjamins Publishing.

MacDonald, S. (2002). Pronunciation: views and practices of reluctant teachers. *Prospect, 17*(3), 3–18.

Macedonia, M. (2014). Bringing back the body into the mind: Gestures enhance word learning in foreign language. *Frontiers in Psychology*, *5*, 1467.

Macedonia, M. (2019). Embodied learning: Why at school the mind needs the body. *Frontiers in Psychology*, *10*, 2098.

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

Macedonia, M., & Knosche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education, 5*(4), 196–211.

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*(6), 982–998.

Macedonia, M., & von Kriegstein, K. (2012). Gestures enhance foreign language learning. *Biolinguistics*, *6*(3–4), 393–416.

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

Madan, C. R., & Singhal, A. (2012). Using actions to enhance memory: Effects of enactment, gestures, and exercise on human memory. *Frontiers in Psychology*, *3*, 507.

Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken

french: behavioral and electrophysiological evidence. *Cerebral Cortex*, *17*, 2659–2668.

Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris, 102*, 59–70.

Majlesi, A. R. (2014). *Matching gestures-Teachers' repetitions of students' gestures in second language learning classrooms*.

Marley, S. C., Levin, J. R., and Glenberg, A. M. (2010). What cognitive benefits does an activity-based reading strategy afford young Native American readers? *Journal of Experimental Education, 78*, 395–417.

Marsh, K. (1995). Children's singing games: Composition in the playground? *Research Studies in Music Education, 4*, 2–11.

Martínez-Fernández, A. (2008). Revisiting the involvement load hypothesis: Awareness, type of task and type of item. In M. Bowles, R. Foote, S. Perpiñán, & R. Bhatt (Eds.), *Selected proceedings of the 2007 second language research forum* (pp. 210-228). Somerville, MA: Cascadilla Proceedings Project.

Mather, S. M. (2005). Ethnographic research on the use of visually based regulators for teachers and interpreters. In Metzger, M. & Fleetwood, E. (Eds.) *Attitudes, innuendo, and regulators (pp.* 136–161). Gallaudet University Press.

Matsumoto, Y., & Dobs, A. M. (2017). Pedagogical gestures as interactional resources for teaching and learning tense and aspect in the ESL grammar classroom. *Language Learning, 67*(1), 7–42.

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

Mavilidi, M. F., Okely, A. D., Chandler, P., Cliff, D. P., & Paas, F. (2015). Effects of integrated physical exercises and gestures on preschool children's foreign language vocabulary learning. *Educational Psychology Review*, *27*(3), 413–426.

McAllister, R., Flege, J., & Piske, T. (1999). The acquisition of Swedish long vs. short vowel contrasts by native speakers of English, Spanish and Estonian. *Proceedings of the 14th International Congress of Phonetic Sciences*, 751–754.

McAndrews, M. (2019). Short periods of instruction improve learners' phonological categories for L2 suprasegmental features. *System*, *82*, 151–160.

McCafferty, S. G. (1998). Nonverbal expression and L2 private speech. *Applied Linguistics*, *19*(1), 73–96.

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

McCafferty, S. G. (2008). Mimesis and second language acquisition: A sociocultural perspective. *Studies in Second Language Acquisition*, *30*(2), 147–167.

McClave, E. (1998). Pitch and manual gestures. *Journal of Psycholinguistic Research, 27*, 69– 89.

McCulloch, W. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics, 7*, 115–133.

McGregor, K.K., Rohlfing, K.J., Bean, A., & Marschner, E. (2009). Gesture as a support for word learning: The case of under. *Journal of Child Language, 36*(4), 807–828.

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

McKinnon, S. (2017). TBLT instructional effects on tonal alignment and pitch range in L2 Spanish imperatives versus declaratives. *Studies in Second Language Acquisition*, *39*(2), 287–317.

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

McNeill, D. (Ed.). (2000) *Language and gesture*. Cambridge: Cambridge University Press.

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

McNeill, D. (2016). *Why we gesture: The surprising role of hand movements in communication*. Cambridge: Cambridge University Press.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, *32*(4), 543–563.

Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In J. Trouvain & U. Gut (Eds.), *Non-native Prosody: Phonetic Descriptions and Teaching Practice* (pp. 53 - 76). Berlin, New York: Mouton De Gruyter.

Mennen, I. (2015). Beyond segments: towards a L2 intonation learning theory (LILt). In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations* (pp. 171–188). Berlin: Springer Verlag.

Merleau-Ponty, M. (1945). *Phénoménologie de la Perception*. Paris: Gallimard.

Meteyard, L., Cuadrado, S. R., Bahrami, B., and Viglicco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex, 48*, 788–804.

Miller, R.M., Sanchez, K., & Rosenblum, L. D. (2013). Is speech alignment to talkers or tasks? *Attention, Perception, & Psychophysics, 75*, 1817–1826.

Milovanov, R., Pietilä, P., Tervaniemi, M., & Esquef, P.A. (2010). Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. *Learning and Individual Differences, 20*, 56–60.

Minogue, J. & Jones, M. (2006). Haptics in education: Exploring an untapped sensory modality. *Review of Educational Research, 76*(3), 317-348.

Missaglia, F. (1999). Contrastive prosody in SLA: An empirical study with Italian learners of German. *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 551–554).San Francisco.

Missaglia, F. (2008). Prosodic training for adult Italian learners of German: the Contrastive Prosody Method. In J. Trouvain & U. Gut (Eds.), *Non-native Prosody* (pp. 237–258). Berlin, New York: De Gruyter Mouton.

Mora, J. C., and Levkina, M. (2017). Task-based pronunciation teaching and research. *Studies in Second Language Acquisition, 39*, 381–399.

Mora, J. C. & Nadeu, M. (2012). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism, 16*(4), 484–500.

Mora, J. C., Rochdi, Y., Kivistö-de Souza, H. (2014). Mimicking accented speech as L2 phonological awareness. *Language Awareness, 23*, 57-75.

Mora-Plaza, N., Mora, J. C., & Gilabert, R. (2018). Learning L2 pronunciation through communicative tasks. In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference* (pp. 174–184). Ames, IA: Iowa State University.

Morett, L. M. (2014). When hands speak louder than words: The role of gesture in the communication, encoding, and recall of words in a novel second language. *Modern Language Journal*, *98*(3), 834–853.

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

Morett, L. M. (2018). In hand and in mind: Effects of gesture production and viewing on second language word learning. *Applied Psycholinguistics*, *39*(2), 355–381.

Morford, M., and Goldin-Meadow, S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language, 23*, 559–580.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, *66*(4), 911–936.

Mori, J., & Hasegawa, A. (2009). Doing being a foreign language learner in a classroom: Embodiment of cognitive states as

social events. *IRAL - International Review of Applied Linguistics in Language Teaching*, *47*(1), 65–94.

Morley, J., (1991). The pronunciation component of teaching English to speakers of other languages. *TESOL Quarterly, 25,* 481–520.

Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics, 35*, 418–440.

Muñoz, C. (2014). Starting age and other influential factors: Insights from learner interviews. *Studies in Second Language Learning and Teaching, 3*, 465–484.

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

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, *34*(4), 520–531.

Murphy, J. (1991). Oral communication in TESOL. Integrating speaking, listening and pronunciation. *TESOL Quarterly, 25*, 51–75.

Myung, J. Y., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition*, *98*(3), 223–243.

Nagle, C. (2019). Perception, imitation, and production: Exploring a three-way perception-production link. In Calhoun, S.,

Escudero, P., Tabain, M & Warren, P. (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1248-1252). Melbourne, Australia.

Nagle, C., Sachs, R., & Zárate–Sández, G. (2018). Exploring the Intersection Between Teachers' Beliefs and Research Findings in Pronunciation Instruction. *Modern Language Journal*, *102*(3), 512–532.

Nakatsukasa, K. (2016). Efficacy of recasts and gestures on the acquisition of locative prepositions. *Studies in Second Language Acquisition*, *38*(4), 771–799.

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(39), 17041–17046.

Nathan, M., & Walkington, C. (2017). Grounded and embodied mathematical cognition: Promoting mathematical insight and proof using action and language. *Cognitive Research: Principles and Implications, 2*, 9.

Naucodie, O. (2019). *Imitation et contrôle prosodique dans l'entraînement à la remédiation phonétique : évaluation, mesure et applications pour l'enseignant de langues étrangères* [doctoral dissertation, Toulouse le Mirail]. TEL Archives Ouvertes.

Nava, E., & Zubizarreta, M. L. (2010). Deconstructing the nuclear stress algorithm: Evidence from second language speech. In

N. Erteschik-Shir & L. Rochman (Eds.), *The sound patterns of syntax* (pp. 291–316). Oxford University Press.

Neil, P.A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D.J. and Shimojo, S. (2006), Development of multisensory spatial integration and perception in humans. *Developmental Science, 9*, 454-464.

Newcombe, N., & Weisberg, S. (Eds.), (2017). Embodied cognition and STEM learning. *Cognitive Research: Principles and Implications, 2*, 28.

Nicoladis, E. (2007). The effect of bilingualism on the use of manual gestures. *Applied Psycholinguistics, 28*(3), 441–454.

Nibert, H. J. (2006). The Acquisition of the phrase accent by beginning adult learners of Spanish as a second language. In M. Díaz-Campos (Ed.), *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology* (pp. 131-148). Somerville, MA: Cascadilla Proceedings Project.

Nielsen, K., 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*(2), 132–142.

Nobe, S. (1993). *Cognitive processes of speaking and gesturing: A comparison between first language speakers and foreign language speakers* [unpublished master's dissertation, University of Chicago].

Nolan, F., and Asu, E. (2009). The pairwise variability index and coexisting rhythms in language. *Phonetica, 66*(1-2), 64–77.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: using the hands to learn math. *Psychological Science, 25*(4), 903–910.

Núñez, R., Edwards, L., & Matos, J. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics, 39*, 45–65.

Odden, D. (2011). The representation of vowel length. In M. van Oostendorp, C. J. Ewen, E. H. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 465–490). Oxford, UK: Wiley-Blackwell.

Odisho, E. Y. (2007). A Multisensory, multicognitive approach to teaching pronunciation. *Linguística - Revista de Estudos Linguísticos Da Universidade Do Porto, 2*, 3–28.

Olson, D. J. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania*, *97*(1), 47–68.

Ong, J. H., Burnham, D., and Escudero, P. (2015). Distributional learning of lexical tones: a comparison of attended vs. unattended listening. *PLoS One, 10*(7), e0133446.

Ong, J. H., Burnham, D., Escudero, P., and Stevens, C. J. (2017). Effect of linguistic and musical experience on distributional learning of nonnative lexical tones. *Journal of Speech, Language and Hearing Research, 60*(10), 2769–2780.

Ordin, M., and Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System 42*, 244–257.

Ordin, M., & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America*, *138*(2), 533–544.

Ortega-Llebaria, M. , & Colantoni, L. (2014). L2 English intonation, relations between form-meaning association, access to meaning, and L1 transfer. *Studies in Second Language Acquisition, 36*, 331–353.

Ortega-Llebaria, M., Nemogá, M., & Presson, N. (2015). Long-term experience with a tonal language shapes the perception of intonation in English words: How Chinese–English bilinguals perceive “rose?” vs. “rose.” *Bilingualism: Language and Cognition, 11*, 1–17.

Ozakin, A. S., Xi, X., Li, P. & Prieto, P. (2021, June). Thanks or tanks: training with tactile cues facilitates the pronunciation of English interdental consonants. *Poster presentation at*

The 12th Annual Pronunciation in Second Language Learning and Teaching Conference (PSLLT 2021). Boston University: Ontario (Canada).

Özçalışkan, S., and Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition, 96, B101–B113*.

Özçalışkan, Ş., Lucero, C., & Goldin-meadow, S. (2016). Is seeing gesture necessary to gesture like a native speaker? *Psychological Science, 27*(5), 737–747.

Özyürek, A., & Kelly, S. D. (2007). Gesture, brain, and language. *Brain and Language, 101*(3), 181–184.

Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience, 19*(4), 605-616.

Pamies Bertran, A. (1999). Prosodic typology: On the dichotomy between stress-timed and syllable-timed languages. *Language Design, 2*, 103–130.

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., &, Lewandowski, E. (2013). Phonetic convergence in shadowing speech: the relation between acoustic and perceptual measures. *Journal of Memory and Language, 69*(3), 183–195.

Pardo, J.S., Urmanche, A., Wilman, S.,, & Wiener, J. (2017). Phonetic convergence across multiple measures and model

talkers. *Attention, Perception, & Psychophysics. 79*, 637–659.

Parrell, B., Goldstein, L., Lee, S., & Byrd, D. (2014). Spatiotemporal coupling between speech and manual motor actions. *Journal of Phonetics, 42*, 1–11.

Peck, K. K., Galgano, J. F., Branski, R. C., Bogomolny, D., Ho, M., Holodny, A. I., & Kraus, D. H. (2009). Event-related functional MRI investigation of vocal pitch variation. *Neuroimage, 44*(1), 175–181.

Peeters, D., Snijders, T. M., Hagoort, P., & Özyürek, A. (2017). Linking language to the visual world: Neural correlates of comprehending verbal reference to objects through pointing and visual cues. *Neuropsychologia*, *95*, 21–29.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America, 116*(4), 2338–2344.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., … Guenther, F. H. (2004). The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language, and Hearing Research, 47*(6), 1259–1269

Pesce, C., Crova, C., Cereatti, L., Casella, R., & Bellucci, M. (2009). Physical activity and mental performance in preadolescents: Effects of acute exercise on free-recall memory. *Mental Health and Physical Activity*, *2*(1), 16–22.

Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: Perceptual illusion of body swapping. *PLoS ONE 3*(12), e3832.

Piaget, J. (1952). *The Origins of Intelligence in Children*. Norton & Co.

Piccinini, G., & Bahar, S. (2013). Neural computation and the computational theory of cognition. *Cognitive Science*, *37*(3), 453–488.

Pierrehumbert, J. B. 1980. *The phonology and phonetics of English intonation* [doctoral dissertation MIT Cambridge University]. MIT Libraries.

Pine, K., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science, 10*(6), 747–754.

Ping, R. M., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology, 44*(5), 1277–1287.

Ping R, & Goldin-Meadow S. (2010). Gesturing saves cognitive resources when talking about non-present objects. *Cognitive Science, 34*, 602–619.

Ping, R. M., Goldin-Meadow, S., & Beilock, S. L. (2014). Understanding gesture: Is the listener's motor system

involved? *Journal of Experimental Psychology, 143*(1), 195–204.

Piske, T. (2008). Phonetic awareness, phonetic sensitivity and the second language learner. In J. Cenoz & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education Vol. 6: Knowledge about Language* (pp. 155-166). New York: Springer Science.

Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*, 191–215.

Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 564–573.

Polyanskaya, L., Ordin, M., and Busa, M. G. (2017). Relative salience of speech rhythm and speech rate on perceived foreign accent in a second language. *Language and Speech, 60*(3), 333–355.

Pontifex,M. B., Hillman, C. H., Fernhall, B., Thompson, K.M., & Valentini, T. A. (2009). The effect of acute aerobic and resistance exercise on working memory. *Medicine & Science in Sports & Exercise, 41*(4), 927–934.

Porter, A. (2012). A helping hand with language learning: Teaching French vocabulary with gesture. *Language Learning Journal, 44*, 236–256.

Poss, N., Hung, T.-H., & Will, U. (2008). The effects of tonal information on lexical activation in Mandarin. *Proceedings of the 20th North American Conference on Chinese Linguistics (pp.* 205–211).

Post, L. S., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of simultaneously observing and making gestures while studying grammar animations on cognitive load and learning. *Computers in Human Behavior*, *29*(4), 1450–1455.

Pouw, W., & Dixon, J. A. (2018). Entrainment and modulation of gesture-speech synchrony under delayed auditory feedback. *Cognitive Science, 43*(3), e12721.

Pouw, W., Harrison, S. J., & Dixon, J. A. (2020). Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony. J*ournal of Experimental Psychology, General, 149*(2), 391-404.

Pouw, T., van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulations. *Educational Psychology Review, 26*, 51–72.

Prator, C. H. (1971). Phonetics vs. phonemics in the ESL classroom: When is allophonic accuracy important? *TESOL Quarterly, 5,* 61-72.

Prieto, P., Cravotta, A., Kushch, O., Rohrer, P., & Vilà- Giménez, I. (2018). Deconstructing beat gestures: a labelling proposal. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D.

Śledziński (Eds.), *Proceedings of the 9th International Conference on Speech Prosody* (pp. 201–205). Poznań, Poland.

Prieto, P., del Mar Vanrell, M., Astruc, L., Payne, E., and Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication, 54*(6), 681–702.

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences, 17*(9), 458–470.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience, 11*, 351–360.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793–797.

Ramírez Verdugo, D. (2006). A study of intonation awareness and learning in non-native speakers of English. *Language Awareness*, *15*(3), 141–159.

Ramus, F., and Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustic Society of America, 105*(1), 512–521.

Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265–292.

Rankin, T. (2017). (Working) memory and L2 acquisition and processing. *Second Language Research*, *33*(3), 389–399.

Ravizza, S. (2003). Movement and lexical access: Do non iconic gestures aid in retrieval? Psychonomic Bulletin & Review, 10(3), 610–615.

Rebuschat, P., Hamrick, P., Sachs, R., Riestenberg, K., Ziegler, N. (2013). Implicit and explicit knowledge of form-meaning connections: evidence from subjective measures of awareness. In J. Bergsleithner, S. Frota, & J. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 249-270). Honolulu: University of Hawai'i, National Foreign Language Resource Center.

Rebuschat, P., Williams, J. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics. 33*, 829-856.

Reichle, R. V., Tremblay, A., & Coughlin, C. (2016). Working memory capacity in L2 processing. *Probus*, *28*(1), 29–55.

Reiterer, S. M., Hu, X., Sumathi, T. A., & Singh, N. C. (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Frontiers in Psychology, 4*, 782.

Renard, R. (1979). *Introduction à la Méthode Verbo-Tonale de Correction Phonétique*. CIPA.

Renard, R. (Ed.). (2002). *Apprentissage d'une langue étrangère/seconde. La phonétique verbotonale*. Bruxelles: De Boek Université.

Rescorla, M. (2020). The computational theory of mind. In E. N. Zalta (Ed.) The *Stanford Encyclopedia of Philosophy (Fall 2020 Edition)*. https://plato.stanford.edu/archives/fall2020/entries/computational-mind/

Richter, K. (2018). Factors affecting the pronunciation abilities of adult learners of English: A longitudinal group study. In S. M. Reiterer (Ed.), *Exploring language aptitude: Views from psychology, the language sciences, and cognitive neuroscience* (pp. 339–361). Cham: Springer.

Riddel, C. (1990). *Traditional singing games of elementary school children in Los Angeles* [doctoral dissertation, University of California-Los Angeles]. Proquest.

Rieser, J. J., Garing, A. E., & Young, M. F. (1994). Imagery, action, and young children's spatial orientation: It's not being there that counts, it's what one has in mind. *Child Development*, *65*(5), 1262. https://doi.org/10.2307/1131498

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes, 14*, 423–441.

Rinne, T., Koistinen, S., Salonen, O., & Alho, K. (2009). Task-dependent activations of human auditory cortex during pitch discrimination and pitch memory tasks. *The Journal of Neuroscience, 29*(42), 13338–13343.

Risko, E. F., & Gilbert, S.J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology*, *210*(5–6), 419–421.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience, 21*(5), 188–194

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey: II. Area F5 and the control of distal movements. *Experimental Brain Research, 71*, 491–507.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192.

Roach P. (1982). On the distinction between 'stress-timed' and 'syllable- timed' languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73–79). London: Edward Arnold.

Roberge, C., Kimura, M., & Kawaguchi, Y. (1996). *Pronunciation training for Japanese: Theory and practice of the VT method [Nihongo no hatsuon shidoo: VT-hoo no riron to jissai]*. Tokyo: Bonjinsha

Robinson, P. (1995). Aptitude, awareness and the fundamental similarity of implicit and explicit foreign language learning. In R. Schmidt (ed.), *Attention and Awareness in Foreign Language Learning* (pp. 303–357). Honolulu, HI: University of Hawai'i Press.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27–57.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25*, 46–73.

Rogers, H. (2000). *The sounds of language: An introduction to phonetics*. New York: Routledge.

Rohrer, P. L., Delais-Roussarie, E. & Prieto, P. (2020). Beat gestures for comprehension and recall: Differential effects of language learners and native listeners. *Frontiers in Psychology, 11*, 2836.

Rohrer, P. L., Prieto, P. & Delais-Roussarie, E. (2019). Beat gestures and prosodic domain marking in French. In Calhoun, S., Escudero, P., Tabain, M. & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 1500-1504.

Romero Naranjo, F. J. (2013). Science & art of body percussion: A review. *Journal of Human Sport and Exercise*, *8*(2), 442–457.

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech Rhythm Facilitates Syntactic Ambiguity Resolution: ERP Evidence. *PLoS ONE*, *8*(2), 1–9.

Rosa, E., Leow, R. (2004). Awareness, different learning conditions, and second language development. *Applied Psycholinguistics, 25*, 269-292.

Rosa, E., O'Neill, M. (1999). Explicitness, intake, and the issue of awareness. *Studies in Second Language Acquisition, 21*, 511-556.

Rose, S. E. (1995). *The effects of Dalcroze eurhythmics on beat competency performance skills of kindergarten, first-, and second-grade children* [doctoral dissertation, University of North Carolina at Greensboro]. ProQuest Dissertations and Theses.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review, 65*, 395–412.

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2011). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia, 50*(2), 232–244.

Rowe, M. L., Silverman, R. D., & Mullan, B. E. (2013). The role of pictures and gestures as nonverbal aids in preschoolers' word learning in a novel language. *Contemporary Educational Psychology, 38*(2), 109–117.

Rueschemeyer, S. A., Lindemann, O., Rooij, D. Van Dam, W. Van, & Bekkering, H. (2010). Effects of intentional motor actions on embodied language processing. *Experimental Psychology*, *57*(4), 260–266.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences, 11*(6), 251-257.

Rusiewicz, H. L., & Esteve-Gibert, N. (2018). Set in time: Temporal coordination of prosodic stress and gesture in the development of spoken language production. In P. Prieto & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language Acquisition* (pp. 103–124). Amsterdam: John Benjamins Publishing Company.

Safronova, E. (2016). *The role of cognitive ability in the acquisition of second language perceptual phonological competence* [doctoral dissertation, Universitat de Barcelona]. Tesis en Xarxa.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. TESOL Quarterly, 46, 842-854.

Saito, K. (2013a). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English /ɹ/ to adult Japanese learners. *Language Learning, 63*, 499-529.

Saito, K. (2013b). Reexamining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition, 35*, 1-29.

Saito, K. (2014). Experienced teachers' perspectives on priorities for improved intelligible pronunciation: The case of Japanese learners of English. *International Journal of Applied Linguistics, 24*, 250–277.

Saito, K. (2015). Communicative focus on second language phonetic form: Teaching Japanese learners to perceive and produce English /π/ without explicit instruction. *Applied Psycholinguistics, 36*(2), 337–409.

Saito, K. (2017). Effects of sound, vocabulary and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning, 67*, 665–693.

Saito, K., & Hanzawa, K. (2016). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics, 37*, 813–840.

Saito, K., & Hanzawa, K. (2018). The role of input in second language oral ability development in foreign language classrooms: A longitudinal study. *Language Teaching Research, 22*, 398–417.

Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics, 39*, 593–617.

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 52–708.

Saito, Y., & Saito, K. (2017). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, *21*(5), 589–608.

Saito, K., Trofimovich, P. & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics, 37*(2), 217–240.

Saito, K., Sun, H., & Tierney, A. (2019). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental and suprasegmental sensitivity and performance via behavioral and neurophysiological measures. *Bilingualism: Language and Cognition, 22*(5), 1123–1140.

Saito, K., Sun, H., & Tierney, A. (2020). A longitudinal investigation of explicit and implicit auditory processing in

L2 segmental and suprasegmental acquisition. *Studies in Second Language Acquisition, 41*, 1083–1112.

Saito, K., Suzukida, Y., & Sun, H. (2019). Aptitude, experience, and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition, 41*, 201–225.

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics, 38*, 439–462.

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

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*, 357–357.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review, 11*, 11–26.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3–30). Cambridge University Press.

Schwab, S., & Goldman, J. P. (2018). MIAPARLE: Online training for discrimination and production of stress contrasts. *Proceedings of the 9th Speech Prosody Conference* (pp. 572–576). Poznań, Poland.

Schwenzer, M., & Mathiak, K. (2011). Numeric aspects in pitch identification: an fMRI study. *BMC Neuroscience, 12*(1), 26–35.

Segalowitz, N. (2010). *Cognitive bases of Second Language Fluency*. New York: Routledge.

Seo, M. S. (2021). Multimodally enhanced opportunities for language learning: Gestures used in word search sequences in ESL tutoring. *Journal of Language Teaching and Research, 12*(1), 44–56.

Shapiro, L. (2019). *Embodied Cognition* (2nd ed.). New York: Routledge.

Shapiro, L., & Stolz, S. A. (2019). Embodied cognition and its significance for education. *Theory and Research in Education*, *17*(1), 19–39.

Sharwood Smith, M. A. (1981). Consciousness-raising and the second language learner. *Applied Linguistics, 2*, 159-168.

Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology, 9*, 1514.

Siakaluk, P., Pexman, P., Aguilera, L., Owen, W., & Sears, C. (2008). Evidence for the activation of sensorimotor information during visual word recognition: The body-object interaction effect, *Cognition, 106*, 433–443.

Sibley, B. A., & Etnier, J. L. (2003). The relationship between physical activity and cognition in children: a meta- analysis. *Pediatric Exercise Science, 15*, 243–256.

Sime, D. (2006). What do learners make of teachers' gestures in the language classroom? *International Review of Applied Linguistics, 44*, 209–228.

Slevc, L.R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science, 17*, 675–681.

Smith, C. P., King, B., & Hoyte, J. (2014). Learning angles through movement: Critical actions for developing understanding in an embodied activity. Journal of Mathematical Behavior, 36, 95–108.

Smotrova, T. (2014). *Instructional functions of speech and gesture in the L2 classroom* [doctoral dissertation, Pennsylvania State University]. Penn State Electronic Theses and Dissertations for Graduate School.

Smotrova, T. (2017). Making Pronunciation Visible: Gesture In Teaching Pronunciation. *TESOL Quarterly*, *51*(1), 59–89.

Smotrova, T. and Lantolf, J.P. (2013), The Function of Gesture in Lexically Focused L2 Instructional Conversations. *Modern Language Journal, 97*, 397-416.

So, W., Chen-Hui, C. S., & Wei-Shan, J. L. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes, 27*, 665–681.

Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2017). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition*, *39*(2), 347–380.

Soulaine, S. (2013). *Les effets du geste sur l'apprentissage du rythme en anglais : couplage des dynamiques vocale et corporelle* [unpublished doctoral dissertation, Université du Maine].

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

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. The MIT Press.

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2009). The Neural Basis of Multisensory Integration in the Midbrain: Its

Organization and Maturation. *Hearing Research*, *258*(1–2), 4–5.

Stockmal, V., D. Markus, and Bond, D. (2005). Measures of native and non-native rhythm in a quantity language. *Language and Speech, 48*(1), 55–63.

Sulkin, I., & Brodsky, W. (2007) The effects of hand-clapping songs training on temporal-motor skills among elementary school children. In K. Overy (Ed.), *Proceedings of the Summer Workshop on Music, Language, and Movement*, Edinburgh: Institute for Music in Human and Social Development, University of Edinburgh.

Sullivan, J. V. (2018). Learning and embodied cognition: A review and proposal. *Psychology Learning and Teaching*, *17*(2), 128–143.

Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*(1), 143–167.

Suzukida, Y. (2021). The contribution of individual differences to L2 pronunciation learning: Insights from research and pedagogical Implications. *RELC Journal*, *52*(1), 48–61.

Suzukida, Y., & Saito, K. (2021). Which segmental features matter for successful L2 comprehensibility? Revisiting and

generalizing the pedagogical value of the functional load principle. *Language Teaching Research, 25*(3), 431–450.

Tanner, M. W., & Landon, M. M. (2009). The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning and Technology*, *13*(3), 51–65.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 239–276). Amsterdam: John Benjamins.

Tellier, M. (2006). *L'impact du geste pédagogique sur l 'enseignement /apprentissage des langues étrangères: Étude sur des enfants de 5 ans* [doctoral dissertation, Université Paris-Diderot - Paris VII]. TEL Archives Ouvertes.

Tellier, M. (2008a). Dire avec des gestes. *Le Français Dans Le Monde, Recherche et Application*, *44*, 1–8.

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

Theakston, A., Coates, A., & Holler, J. (2014). Handling agents and patients: Representational co-speech gestures help children comprehend complex syntactic constructions. *Developmental Psychology, 50*(7), 1973–1984.

Thelen, E., Schöner, G., Scheier, C., & Smith, L. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. B*ehavioral and Brain Sciences, 24*(1), 1-34.

Thomson, R. (2018). High Variability Pronunciation Training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208–231.

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326–344.

Thorson, J. C. (2018). The role of prosody in early word learning: Behavioral evidence. In P. Prieto & N. Esteve-Gibert (Eds.), *The development of prosody in first language acquisition* (pp. 59–77). Amsterdam: John Benjamin Publishing Company.

Toth, P. D., & Moranski, K. (2018). Why haven't we solved instructed SLA? A sociocognitive account. *Foreign Language Annals, 51*,73–89.

Tomlin, R.S. and Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16*, 183–203

Tomporowski, P. D. (2003). Effects of acute bouts of exercise on cognition. *Acta Psychologica, 112*, 297–324.

Tomporowski, P. D., Davis, C. L., Miller, P. H., & Naglieri, J. A. (2008). Exercise and children's intelligence, cognition, and academic achievement. *Educational Psychology Review, 20*, 111–131.

Trimble, J. C. (2013). Perceiving intonational cues in a foreign language: Perception of sentence type in two dialects of Spanish. In C. Howe et al. (Eds.), *Selected Proceedings of the 15th Hispanic Linguistics Symposium* (pp. 78-92). Somerville, MA: Cascadilla Proceedings Project.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(01), 1–30.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition, 15*(4), 905- 916.

Troubetzkoy, N.S. (1964). *Principes de Phonologie*. Klincksieck.

Tsunemoto, A. Trofimovich, P., & Kennedy, S. (2020). Pre-service teachers' beliefs about second language pronunciation teaching, their experience, and speech assessments. *Language Teaching Research*, *FirstOnline*. http://doi.org/10.1177/1362168820937273

Tuninetti, A., Mulak, K. E., and Escudero, P. (2020). Cross-situational word learning in two foreign languages:

effects of native language and perceptual difficulty. *Frontiers in Communication, 5*, 109.

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

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology, 28*(2), 187–204.

van Compernolle, R. A., & Williams, L. (2011). Thinking with your hands: Speech–gesture activity during an L2 awareness-raising task. *Language Awareness, 20*, 203–219.

van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology, 58*, 251–273.

van Leussen, J. W., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology, 6*, 1000.

van Maastricht, L., Hoetjes, M., & van Drie, E. (2019, July). Do gestures during training facilitate L2 lexical stress acquisition by Dutch learners of Spanish? In *15th Conference on Auditory-Visual Speech Processing* (pp. 6–10). Melbourne, Australia.

van Maastricht, L., Krahmer, E., & Swerts, M. (2016a). Prominence patterns in a second language: Intonational

transfer From Dutch to Spanish and vice versa. *Language Learning, 66*(1), 124–158.

van Maastricht, L., Krahmer, E., & Swerts, M. (2016b). Native speaker perceptions of (non-)native prominence patterns: Effects of deviance in pitch accent distributions on accentedness, comprehensibility, intelligibility, and nativeness. *Speech Communication*, *83*, 21–33.

Van Maastricht, L., Krahmer, E., Swerts, M., & Prieto, P. (2019). Learning direction matters: A study on L2 rhythm acquisition by Dutch learners of Spanish and Spanish learners of Dutch. *Studies in Second Language Acquisition, 41*(1), 87-121.

van Maastricht, L., Zee, T., Krahmer, E., & Swerts, M. (2020). The interplay of prosodic cues in the L2: how intonation, rhythm, and speech rate in speech by Spanish learners of Dutch contribute to L1 Dutch perceptions of accentedness and comprehensibility. *Speech Communication, 133*, 81-90.

VanPatten, B. (1996). *Input Processing and Grammar Instruction in Second Language Acquisition*. Norwood: Ablex Publishing

VanPatten, B. (2002). Processing instruction: An update. *Language Learning, 52*, 755-803.

Verdugo, M. D. R. (2002). Non-native interlanguage intonation systems: A study based on a computerized corpus of Spanish learners of English. *ICAME Journal, 26*, 115–132.

Vilà-Giménez, I., Dowling, N., Demir-Lira, Ö. E., Prieto, P., & Goldin-Meadow, S. (in press, 2021). The predictive value of non-referential beat gestures: Early use in parent-child interactions predicts narrative abilities at 5 years of age. *Child Development.* doi: 10.1111/cdev.13583.

Vilà-Giménez, I., Igualada, A., & Prieto, P. (2019). Observing storytellers who use rhythmic beat gestures improves children's narrative discourse performance. *Developmental Psychology, 55*(2), 250-262.

Vilà-Giménez, I., & Prieto, P. (2020). Encouraging kids to beat: Children's beat gesture production boosts their narrative performance. *Developmental Science, 23(6)*, e12967.

Vilà-Giménez, I., & Prieto, P. (2021). The value of non-referential gestures: A systematic review of their cognitive and linguistic effects in children's language development. *Children, 8*(2), 148.

Vinther, T. (2002). Elicited imitation: a brief overview. *International Journal of Applied Linguistics*, *12*(1), 54–73.

Volterra, V., Bates, E., Benigni, L., Bretherton, I., and Camaioni, L. (1979). First words in language and action: a qualitative look. In E. Bates (Ed.), T*he Emergence of Symbols: Cognition and Communication in Infancy* (pp. 141–222). New York Academic Press.

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

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language, 50*(4), 395–407.

Wang, X. (2020). Segmental versus suprasegmental: Which one is more important to teach? *RELC Journal*, *OnlineFirst*. https://doi.org/10.1177/0033688220925926

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

Wang, W., & Loewen, S. (2016). Nonverbal behavior and corrective feedback in nine ESL university-level classrooms. *Language Teaching Research*, *20*(4), 459–478.

Wanrooij, K., Escudero, P., and Raijmakers, M. E. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics, 41*(5), 319–102.

Warren, P., Elgort, I., and Crabbe, D. (2009). Comprehensibility and prosody ratings for pronunciation software development. *Language Learning and Technology, 13*(3), 87–102.

Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology*, *5*, 1–10.

Weltens, B., & de Bot, K. (1984a). Visual feedback of intonation II: Feedback delay and quality of feedback. *Language and Speech, 271*, 79–88.

Weltens, B., & de Bot, K. (1984b). The visualization of pitch contours: some aspects of its effectiveness in teaching foreign intonation. *Speech Communication, 3*(2), 157–163.

White, L., and Mattys, S. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*(4), 501–522.

Willems, R. M., and Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: a review. *Brain and Language, 101*(3), 278–289.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636.

Wrembel, M. (2005). *Phonological metacompetence in the acquisition of second language phonetics* [doctoral dissertation, Adam Mickiewicz University].

Xi, X., Li, P., Baills, F. & Prieto, P. (2020). Hand gestures facilitate the acquisition of novel phonemic contrasts when they appropriately mimic target phonetic features. *Journal of Speech, Language, and Hearing Research, 63*, 3571–3585.

Yang, N. (2016). *Improving the English speaking skills and phonological working memory of Chinese primary EFL learners with a verbotonal-based approach* [doctoral dissertation, Suranaree University of Technology]. Suranaree University of Technology Intellectual Repository.

Yang, J., & Shu, H. (2016). Involvement of the motor system in comprehension of non-Literal action language: A meta-analysis study. *Brain Topography*, *29*(1), 94–107.

Xu, H. M. (2009). A survey study of Chinese EFL learners' acquisition of English intonation: A Functional perspective [unpublished master's dissertation, Jiangsu University].

Yazawa, K., Whang, J., Kondo, M., & Escudero, P. (2020). Language-dependent cue weighting: an investigation of perception modes in L2 learning. *Second Language Research, 36*(4), 557–581.

Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of spanish intonation by mandarin speakers. *Studies in Second Language Acquisition*, *41*(1), 5–32.

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *Journal of Cognitive Neuroscience, 22*(4), 775–789.

Zhang, F. Z. (2006). *The teaching of Mandarin prosody: A somatically-enhanced approach for second language*

learners [doctoral dissertation, university of Canberra]. University of Canberra.

Zhang, Y., Baills, F., & Prieto, P. (2020a). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words. *Language Teaching Research, 24*(5), 666–689.

Zhang, Y., Baills, F., & Prieto, P. (2020b, September). Good hand-clapping performance is related to foreign language imitation skills. *Proceedings of the 7th Gesture and Speech in Interaction Conference*. KTH Royal Institute of Technology, Stockholm.

Zhang, R., & Yuan, Z. (2020). Examining the effects of explicit pronunciation instruction on the development of L2 pronunciation. *Studies in Second Language Acquisition*, *42*(4), 905–918.

Zhen, A., Van Hedger, S., Heald, S., Goldin-Meadow, S., & Tian, X. (2019). Manual directional gestures facilitate cross-modal perceptual learning. *Cognition*, 187, 178–187.

Zheng, A., Hirata, Y., & Kelly, S. D. (2018). Exploring the effects of imitating hand gestures and head nods on L1 and L2 mandarin tone production. *Journal of Speech, Language, and Hearing Research*, *61*(9), 2179–2195.

Zheng, C., Saito, K., & Tierney, A. (2020). Successful second language pronunciation learning is linked to domain-general auditory processing rather than music aptitude. *Second*

Language Research, FirstOnline. http://doi.org/10.1177/0267658320978493

Zielinski, B. (2008). The listener: No longer the silent partner in reduced intelligibility. *System, 36*, 69–84.

Appendices of Chapter 3

Appendix A. Dialogues

A1. Sample dialogue for the training session

Text of one of the dialogues trained during the first training session (left-hand box) with English translation (right-hand box).

Title: À la poste Personnages: L'employée (E) et un client (C) E: Bonjour, Monsieur. C: Bonjour. Je viens chercher une lettre recommandée. E: Vous avez une pièce d'identité? C: Oui, voilà. E: Je suis désolée, votre lettre n'est pas là. C: Mais j'ai reçu cet avis dans ma boîte tout à l'heure. E: Oui, mais le facteur n'est pas rentré de sa tournée. Repassez dans deux heures! C Alors, j'ai attendu vingt minutes pour rien? E: Désolée. Au suivant! Vocabulaire La poste: Correus	Title: At the post office Characters: the post office employee (E) and a client (C) E: Good morning, sir. C: Good morning. I came to pick up a registered letter. E: Do you have some identification? C: Yes, here it is. E: I'm very sorry, your letter isn't here. C: But I received this notice in my mail box earlier today. E: Yes, but the postman is not back from his rounds yet. Come back in two hours! C: So I just waited twenty minutes for nothing? E: Sorry. Next! Glossary

Une lettre recommandée: Una carta certificada Une pièce d'identité: El DNI Le facteur: El carter La tournée: La ronda	La poste: French national postal service Une lettre recommandée: A registered letter Une pièce d'identité: A form of identification Le facteur: The postman La tournée: The postman route

A2. Organization of the dialogues across sessions and dialogues' transcription

Session	Dialogues			
pretest	A1	B2	C3	Untrained
training 1	A1	A2	A3	
training 2	B1	B2	B3	
training 3	C1	C2	C3	
posttest	A1	B2	C3	Untrained

A3. Transcription of the 10 dialogues

The target sentences used as the audiovisual training stimuli are underlined.

A1 - À la poste

Personnages: L'employée (en bleu) et un client (en noir)

- Bonjour, Monsieur
- Bonjour. Je viens chercher une lettre recommandée.
- Vous avez une pièce d'identité?
- Oui, voilà.
- Je suis désolée, votre lettre n'est pas là.
- Mais j'ai reçu cet avis dans ma boîte tout à l'heure.
- Oui, mais le facteur n'est pas rentré de sa tournée. Repassez dans deux heures!
- Alors, j'ai attendu vingt minutes pour rien?
- Désolée. Au suivant!

A2 - Rendez-vous chez le coiffeur

Personnages: La coiffeuse (en bleu) et Mr Ladurie (en noir)

- Espace coiffure, bien le bonjour.
- Bonjour, mademoiselle. Ici madame Ladurie. Je voudrais prendre un rendez-vous.
- Oui, madame. Qui vous coiffe normalement?
- C'est Jean-Pierre.
- 10h30, ça vous va?
- Je préférerais un peu plus tard.
- 11h30?
- C'est parfait.
- Vous pouvez me rappeler votre nom?
- Madame Ladurie.
- Bien, Madame Ladurie, mercredi, 11h30. C'est noté. Au revoir, madame. À mercredi.
- Au revoir, mademoiselle.

A3- Invitation refusée

Personnages: Sylvie (en bleu) et Daniel (en noir)

- Si on allait à la piscine?

- Tous les deux?
- Oui.
- Ça ne me dit pas grand-chose.
- Tu veux aller te promener?
- Écoute, j'ai du travail en ce moment. On verra bien après les cours.

B1 - Le nouveau chef

Personnages: M. Dumas (en bleu) et M. Hugon (en noir)

- Alors, le nouveau chef, vous l'avez vu?
- Oui, je sors maintenant de son bureau. Elle a l'air compétente.
- Vous dites "elle"? C'est une femme?
- Oui, elle s'appelle Myriam Duchemin.
- D'où vient-elle?
- De l'agence de Rennes. Elle est bretonne, à mon avis.
- Et quel âge a-t-elle?
- Elle est plutôt jeune pour le poste. La quarantaine.
- Et comment est-elle physiquement?
- Oh! Elle est brune, de taille moyenne, avec des yeux verts. Que dire d'autre? Elle semble être très dynamique.
- Hum! Hum! Merci bien. Je vais aller faire sa connaissance immédiatement.

B2 - Portrait-robot - Au commissariat

Personnages: L'inspecteur (en bleu) et Mme Thomas (en noir)

- Alors, madame, décrivez-moi votre agresseur.
- C'est un homme d'un certain âge, à l'allure bizarre.
- Quel âge a-t-il environ?
- Je ne sais pas, une soixantaine d'années. Il n'était pas très grand.
- De quelle couleur sont ses cheveux?
- Il était un peu chauve. Avec des cheveux blancs. Il avait aussi une longue barbe blanche.
- Et ses yeux?
- Il avait les yeux bleus.

- Bien. Et comment était-il habillé?
- Il portait un long manteau rouge.
- Un manteau rouge? Vous en êtes sûre?
- Oui, il était déguisé en Père Noël. Je ne vous l'avais pas dit?

B3 - Les retrouvailles

Personnages: Philippe (en bleu) et Carole (en noir)

- Eh, Carole. C'est toi?
- Excusez-moi, on se connaît?
- Mais oui! c'est moi, Philippe Langon. Tu ne me reconnais pas?
- Philippe! Ça fait longtemps! Tu as changé! dis donc. Tu as maigri, non?
- Ah! Tu as divorcé? Je ne savais pas…
- Et toi, tu n'as pas changé. Toujours la même? Célibataire?
- Oui, mais je vis avec mon ami, Pierre…
-: Ah! Ah! Tu vas me raconter ça. Allez, viens. On va prendre un verre pour fêter nos retrouvailles.

C1 - Déprime

Personnages: Pascale (en bleu) et Brigitte (en noir)

- Salut, Brigitte. Comment ça va?
- Ça pourrait allait mieux.
- Qu'est-ce qu'il t'arrive ?
- J'ai le cafard depuis que Marc est en stage à Londres.
- Mais ce n'est pas la fin du monde. Il revient quand ?
- Le mois prochain.
- Allez courage ! Un mois, c'est rien! C'est vite passé.

C2 - Vacances

Personnages: Aline (en bleu) et Paul (en noir)

- Alors tes vacances?

- Mes vacances? Pas terribles. Je me suis ennuyé: je me suis cassé la jambe le premier jour.
- Oh, ma pauvre! Pas de chance!
- J'ai donc dû rester au chalet sans rien faire.
- Aíe! Aïe! Aïe! Ça ne devait pas être génial.
- Des vacances comme ça, j'en veux plus jamais. Et les tiennes, au fait?
- De mon côté, c'était très chouette! J'en suis ravi!
- On dirait!
- Moi qui aie peur de tout, j'ai fait du saut à l'élastique et, encore mieux, j'ai rencontré la femme parfaite. Tu imagines?
- Oui, je vois! Et c'est bientôt que tu me la présentes?

C3 - Sortie

Personnages: Caroline (en noir) et le père de Caroline (en bleu)

- Papa, je peux aller au cinéma?
- Je regrette mais tu as cours demain. Je ne veux pas que tu te couche tard.
- Mais Sylvie a eu la permission.
- Pas question! Sylvie, c'est Sylvie. Toi, c'est toi.
- Mais papa…
- Ça suffit. C'est comme ça.
- Y en a marre. C'est toujours la même chose.
- J'ai dit non et c'est non. Et parle-moi autrement.

Untrained - L'interrogatoire

Personnages: Gilles (en bleu) et la mère de Gilles (en noir).

La mère: On peut savoir à quelle heure tu es rentré cette nuit?
Gilles: À deux heures du matin, je crois.
La mère: Et qu'est-ce que tu faisais dehors à une heure pareille?
Gilles: Je revenais de la discothèque.
La mère: Et tu étais avec qui?
Gilles: Avec Sophie.
La mère: Sophie? Qui est-ce?

Gilles: Une collègue de travail.
La mère: Ah, bon! Et que font ses parents?
Gilles: Mais maman, tu exagères! J'ai trente-deux ans, quand-même!

Appendix B. Audiovisual training stimuli: recording process

Before recording, the three instructors reached consensus on how they would utter the target sentences in terms of intonation and rhythm, and then practised reading them together, checking that the sentences were produced in a clear and natural manner appropriate to the context of the dialogue. They then jointly practised saying the nonsense syllable logatomes that corresponded to each target sentence, both with and without accompanying gestures. Finally, each instructor was individually video-recorded performing each target sentence in the three modes (saying the sentence, uttering the logatome, and uttering the logatome while making hand movements) consistently in this order to maintain a high degree of uniformity in their performance. During this recording process, the instructors monitored each other's performance for naturalness and inter-instructor consistency both in terms of speech and gesture and repeated the performance if this seemed desirable.

After all these materials were recorded, the first author used Praat software (www.praat.org) to compare the pitch contours of the target sentence across the three conditions (as captured in the audio track of the recordings) to ensure consistency within each instructor's performance. Also, for each embodied logatome stimulus, the performance of hand movements was checked to make sure it appropriately matched the pitch contour and rhythm of the target sentence to which it corresponded.

This process yielded video clips showing the three instructors each performing 45 target sentences (5 sentences × 9 dialogues) in three conditions. Note that not all of this raw material was necessary,

since only one performance of each sentence (in three conditions) would be needed in the final stimulus material. However, the fact that the three instructors had all performed made it possible to use recordings from different instructors to represent the different speakers in a dialogue (for example, for the dialogue shown in Table 1, one instructor would perform the sentences spoken by the post office employee and a different instructor would perform the sentences spoken by the client), thus producing a more naturalistic final stimulus.

Appendix C. Inferential statistics: effect of type of training and pairwise contrasts

C1. Comprehensibility

Comprehensibility	Fixed coefficients					Fixed effects			
Term	*b**	*SE*	*t*	*p*	95% CI	*F*	*df1*	*df2*	*p*
Intercept	7.68	0.11	66.97	< .001	[7.45, 7.90]	3.54	6	1793	.002
Group = speech only	- 0.18	0.15	-1.14	.25	[-0.48, -0.13]	0.24	2	1793	.78
Group = non-embodied logatome	-0.14	0.16	-.82	.41	[-0.46, -0.19]				
Session = pretest	-0.37	0.10	-3.54	<.001	[-0.57, -0.16]	18.63	1	1793	< .001
Group (=speech) X Session (=pretest)	0.16	0.14	1.15	.25	[-0.11, 0.44]	.81	2	1793	.44
Group (=lotatome) X Session (= pretest)	0.16	0.15	1.05	.29	[-0.14, 0.45]				
Familiarity (= new)	-0.041	0.07	-0.59	.55	[-0.176, 0.09]	.35	1	1793	.55

Term	Level	Contrast	Est.	*SE*	*t*	*df*	*p*	95% CI
Group	-	speech only - non-emb. logatome	-.04	.14	-0.273	1793	1	[-.34, .26]
	-	speech only - embodied logatome	-.10	.14	-0.69	1793	1	[-.43,.24]
	-	non-emb. logatome – embodied logatome	-.06	.15	-0.388	1793	1	[-.37, 0.25]
Session	-	pretest - posttest	-.26	.06	-4.32	1793	1	[-.38, -.14]
Group x Session	Pretest	speech only - non-emb. logatome	-.01	.17	-0.04	1793	1	[-.35, .34]
		speech only - embodied logatome	-.01	.17	0.049	1793	1	[-.33, .34]
		non-emb. logatome – embodied logatome	.02	.18	0.09	1793	1	[-.41, .45]
	Posttest	speech only - non-emb. logatome	-.07	.17	-0.41	1793	.68	[-.41, .27]
		speech only - embodied logatome	-.42	.17	-2.52	1793	.04	[-.83, -.02]
		non-emb. logatome – embodied logatome	-.353	.18	-1.97	1793	.099	[-.76, .05]
	Speech only	pretest - posttest	-.20	.10	-2.07	1793	.04	[-.39,-.01]
	Non-emb. logatome	pretest - posttest	-.21	.11	-1.89	1793	.06	[-.42, .01]
	Embodied logatome	pretest - posttest	-.37	.10	-3.54	1793	<.001	[-.57, -.16]
Familiarity	-	trained- untrained	-.04	.07	-0.59	1793	.55	[-.18, .09]

C2. Fluency

Fluency	Fixed coefficients					Fixed effects			
Term	*b**	*SE*	*t*	*p*	95% CI	*F*	*df1*	*df2*	*p*
Intercept	7.13	.12	57.21	< .001	[6.88, 7.38]	16.70	6	1793	< .001
Group = speech	- 0.15	.17	-0.88	.38	[-0.48, 0.18]	0.18	2	1793	.83
Group = logatome	-0.17	.18	-0.93	.35	[-0.52, 0.19]				
Session = pretest	-0.75	.11	-6.68	<.001	[-0.97, -0.53]	96.50	1	1793	< .001
Group (=speech) X Session (=pretest)	0.19	.15	1.22	.22	[-0.11, 0.49]	0.78	2	1793	.46
Group (=lotatome) X Session (= pretest)	0.14	.16	0.86	.39	[-0.18, 0.46]				
Familiarity (= new)	-0.09	.07	-1.17	.24	[-0.06, 0.23]	1.37	1	1793	.24

Term	Level	Contrast	Est.	*SE*	*t*	*df*	*p*	95% CI
Group	-	speech only - non-emb. logatome	.04	.16	0.26	1793	1	[-.28 , .36]
	-	speech only - embodied logatome	-.05	.15	-0.36	1793	1	[-.37,.26]
	-	non-emb. logatome – embodied logatome	-.10	.16	-0.60	1793	1	[-.48, .29]
Session	-	pretest - posttest	-.64	.06	-9.82	1793	<.001	[-.76, -.51]
Group x Session	Pretest	speech only - non-emb. logatome	.65	.18	0.37	1793	1	[-.36, .48]
		speech only - embodied logatome	.38	.17	0.23	1793	1	[-.31, .38]
		non-emb. logatome – embodied logatome	-.03	.18	0.14	1793	1	[-.34, .39]
	Posttest	speech only - non-emb. logatome	.02	.17	0.10	1793	1	[-.55, .25]
		speech only - embodied logatome	-.15	.17	-0.88	1793	1	[-.55, -.25]
		non-emb. logatome – embodied logatome	-.17	.18	-0.93	1793	1	[-.60, .26]
	Speech only	pretest - posttest	-.56	.11	-5.29	1793	< .001	[-.77 , .35]
	Non-emb. logatome	pretest - posttest	-.61	.12	-5.03	1793	<.001	[-.84, -.37]
	Embodied logatome	pretest - posttest	-.75	.11	-6.68	1793	<.001	[-.97, -.53]
Familiarity		trained- untrained	-.01	.06	-0.20	1793	.84	[-.12, .10]

C3. Accentedness

Accentedness	Fixed coefficients					Fixed effects			
Term	*b**	*SE*	*t*	*p*	95% CI	*F*	*df1*	*df2*	*p*
Intercept	6.61	.12	53.45	< .001	[6.37, 6.85]	14,110	6	1793	< .001
Group = speech	- 0.42	.17	-2.52	.01	[-0.76, -0.09]	.95	2	1793	.387
Group = logatome	-0.35	.18	-1.97	.05	[-0.70, -0.01]				
Session = pretest	-0.68	.09	-7.87	<.001	[-.85, -.51]	68,138	1	1793	< .001
Group (=speech) X Session (=pretest)	0.43	.12	3.63	<.001	[0.20, 0.67]	7,384	2	1793	.001
Group (=lotatome) X Session (= pretest)	0.37	.13	2.91	.004	[0.12, 0.62]				
Familiarity (= new)	-0.01	.06	-0.20	.84	[-0.12, 0.10]	.04	1	1793	.838

Term	Level	Contrast	Est.	*SE*	*t*	*df*	*p*	95% CI
Group	-	speech only - non-emb. logatome	-.40	.16	-0.24	1793	.81	[-.36, .28]
	-	speech only - embodied logatome	-.21	.16	-1.32	1793	.56	[-.59,.17]
	-	non-emb. logatome – embodied logatome	-.17	.16	-1.003	1793	.63	[-.54, 0.21]
Session	-	pretest - posttest	-.42	.05	-8.25	1793	<.001	[-.51, -.32]
Group x Session	Pretest	speech only - non-emb. logatome	-.01	.17	-0.04	1793	1	[-.35, .34]
		speech only - embodied logatome	-.01	.17	0.049	1793	1	[-.33, .34]
		non-emb. logatome – embodied logatome	.02	.18	0.09	1793	1	[-.41, .45]
	Posttest	speech only - non-emb. logatome	-.07	.17	-0.41	1793	.68	[-.41, .27]
		speech only - embodied logatome	-.42	.17	-2.52	1793	.04	[-.83, -.02]
		non-emb. logatome – embodied logatome	-.35	.18	-1.97	1793	.099	[-.76, .05]
	Speech only	pretest - posttest	-.25	.08	-3.05	1793	.002	[-.41 ,-.10]
	Non-emb. logatome	pretest - posttest	-.31	.09	-3.37	1793	.001	[-.47, -.13]
	Embodied logatome	pretest - posttest	-.68	.09	-7.87	1793	<.001	[-.85, -.51
Familiarity		trained- untrained	-.01	.06	-0.20	1793	.84	[-.12, .10]

C4. Segmental accuaracy

Segmental accuracy	Fixed coefficients					Fixed effects			
Term	*b**	*SE*	*t*	*p*	95% CI	*F*	*df1*	*df2*	*p*
Intercept	6.63	.12	52.80	< .001	[6.38, 6.88]	4,75	6	1793	< .001
Group = speech	- 0.14	.17	-0.85	.40	[-.48, .19]	.10	2	1793	.90
Group = logatome	-0.03	.18	-0.17	.86	[-0.39, 0.33]				
Session = pretest	-0.36	.10	-3.64	<.001	[-0.55, -0.16]	27.44	1	1793	< .001
Group (=speech) X Session (=pretest)	0.15	.13	1.15	.25	[-0.11, 0.42]	.77	2	1793	.46
Group (=lotatome) X Session (= pretest)	0.02	.14	0.16	.87	[-0.26, 0.30]				
Familiarity (= new)	0.01	.06	0.10	.92	[-0.12, 0.35]	.01	1	1793	.92

Term	Level	Contrast	Est.	*SE*	*t*	*df*	*p*	95% CI
Group	-	speech only - non-emb. logatome	-.05	.16	-0.30	1793	1	[-.39, .29]
	-	speech only - embodied logatome	-.07	.16	-0.43	1793	1	[-.45,.31]
	-	non-emb. logatome – embodied logatome	-.02	.16	-0.12	1793	1	[-.35, 0.32]
Session	-	pretest - posttest	-.30	.06	-5.24	1793	<.001	[-.41, -.19]
Group x Session	Pretest	speech only - non-emb. logatome	-.02	.18	0.09	1793	1	[-.41, .44]
		speech only - embodied logatome	.01	.17	0.05	1793	1	[-.33, .35]
		non-emb. logatome – embodied logatome	-.01	.18	-0.04	1793	1	[-.37, .35]
	Posttest	speech only - non-emb. logatome	-.11	.18	-0.64	1793	1	[-.51, .28]
		speech only - embodied logatome	-.14	.17	-0.85	1793	1	[-.56, .26]
		non-emb. logatome – embodied logatome	-.03	.18	-0.17	1793	1	[-.40, .34]
	Speech only	pretest - posttest	-.20	.09	-2.19	1793	.029	[-.38 ,-.02]
	Non-emb. logatome	pretest - posttest	-.33	.10	-3.19	1793	.001	[-.54, -.13]
	Embodied logatome	pretest - posttest	-.36	.10	-3.64	1793	<.001	[-.55, -.16
Familiarity		trained- untrained	-.01	.06	-0.10	1793	.92	[-.13, .12]

C5. Suprasegmental accuracy

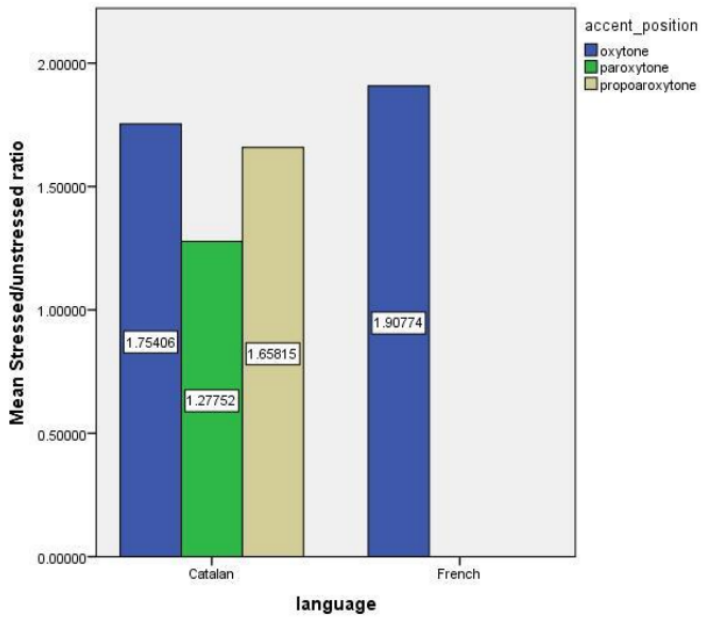
Suprasegmental accuracy	Fixed coefficients					Fixed effects			
Term	*b**	*SE*	*t*	*p*	95% CI	*F*	*df1*	*df2*	*p*
Intercept	7.57	.11	69.26	< .001	[6.37, 6.85]	35.53	6	1793	< .001
Group = speech	- 0.37	.15	-2.49	.01	[-0.66, -0.08]	1.03	2	1793	.36
Group = logatome	-0.30	.16	-1.92	.05	[-0.61, 0.07]				
Session = pretest	-0.88	.08	-10.78	<.001	[-1.04, -0.72]	197.89	1	1793	< .001
Group (=speech) X Session (=pretest)	0.39	.11	3.52	<.001	[0.17, 0.61]	6.26	2	1793	.002
Group (=lotatome) X Session (= pretest)	0.25	.12	2.07	.038	[0.14, 0.48]				
Familiarity (= new)	-0.08	.05	-1.52	.13	[-0.19, 0.02]	2.33	1	1793	.127

Term	Level	Contrast	Est.	*SE*	*t*	*df*	*p*	95% CI
Group	-	speech only - non-emb. logatome	-.007	.14	0.05	1793	.96	[-.27, .28]
	-	speech only - embodied logatome	-.17	.14	-1.26	1793	.63	[-.50,.16]
	-	non-emb. logatome – embodied logatome	-.18	.15	-1.23	1793	.63	[-.53, 0.17]
Session	-	pretest - posttest	-.67	.05	-14.07	1793	<.001	[-.76, -.57]
Group x Session	Pretest	speech only - non-emb. logatome	-.08	.15	0.52	1793	1	[-.29, .45]
		speech only - embodied logatome	.02	.15	0.16	1793	1	[-.28, .32]
		non-emb. logatome – embodied logatome	-.06	.16	-0.36	1793	1	[-.39, .28]
	Posttest	speech only - non-emb. logatome	-.07	.15	-0.43	1793	.68	[-.37, .24]
		speech only - embodied logatome	-.37	.15	-2.49	1793	.04	[-.73, -.01]
		non-emb. logatome – embodied logatome	-.30	.16	-1.92	1793	.11	[-.66, .05]
	Speech only	pretest - posttest	-.48	.08	-6.29	1793	< .001	[-.64 ,-.33]
	Non-emb. logatome	pretest - posttest	-.63	.09	-7.27	1793	< .001	[-.80, -.46]
	Embodied logatome	pretest - posttest	-.88	.8	-10.78	1793	<.001	[-1.04, -.72]
Familiarity		trained- untrained	-.08	.05	-1.52	1793	.13	[-.19, .02]

Appendices of Chapter 4

Appendix A. Exploratory acoustic analysis to compare French and Catalan cognate words pronounced by children.

To compare the difference in duration of the stress syllable between French and Catalan, two French eight-years-old children and two Catalan eight-years-old children were recorded pronouncing 20 cognate words in their native language (see Table 1). For each word, the accented vowel and the preceding unaccented vowel were manually labeled in Praat (for biberó and biberon, the unstressed vowel /i/ was labeled because /ə/ is reduced in French. For Catalan words ending with the diphthong /jə/ and French words ending with the diphthong /jɔ̃ / the full diphthongs were labeled). The duration of the vowels (in seconds) was extracted using a Praat script by Lennes (2003) and the ratio accented vowels/unaccented vowels was calculated. The results show as follows:



Appendix B. Lists of words employed in the memory span task

Words used in this span memory task appear in the Spanish-language MacArthur–Bates Communicative Development Inventories (S-CDIs) (López-Ornat et al., 2005).

1 word
- menjar

2 words
- pa, gos
- abric, galeta
- pajama, mar

3 words
- barba, suc, cosí
- bolígraf, plàtan, porta
- cadira, coll, paper
- nuvi, pastís, ull

4 words
- casa, llengua, poma, telèfon
- sabata, guitarra, sol, cuina
- ungla, llapis, nebot, maduixa
- abella, peu, pallasso, llibre

5 words
- clau, aigua, dent, nina, dutxa
- lluna, bruixa, germà, cocodril, banyador
- ratolí, guants, tren, vestit, mandarina
- barret, grua, autobus, jardí, llum

6 words
- lloro, raqueta, cotxe, poma, sandalia, neu
- tisores, dibuix, nit, groc, tovallola, cuina
- aixeta, moneda, sabó, marieta, nas, llaminadura
- got, sofà, crema, pantalons, ou, conill

Reference:
López-Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S., & Martínez, M. (2005). *Inventario del desarrollo MacArthur: Versión española*. Madrid: TEA.

Appendix C. Words for the speech Imitation Talent task in the six languages with their translations into English

Russian

- milo ‘soap’
- shelushenie ‘exfoliation’

German

- Haarphön ‘hairdryer’
- Küchenschrank ‘cupboard'

Tagalog

- totoo ‘true’
- naghanda ‘prepared’

Hebrew

- mechonit ‘car’
- tsaharaim ‘midday’

Turkish

- üzgün ‘sad’
- bgrti ‘shouting’

Chinese

- zhuozi ‘desk’
- shangwu ‘morning’

List of publications

The following co-authored and authored publications are associated with this dissertation:

Baills, F., Rohrer, P. -L., & Prieto, P. (in press). Le geste et la voix pour enseigner la prononciation en langue étrangère. *Mélanges Crapel.*

Li, P., Xi, X., Baills, F. & Prieto, P. (in press, 2021). Training non-native aspirated plosives with hand gestures: Learners' gesture performance matters. *Language Cognition and Neuroscience. First Online.* https://doi.org/10.1080/23273798.2021.1937663

Li, P., Xi, X., Baills, F., Baqué, L., & Prieto, P. (under review). Embodied prosodic training helps improve not only accentedness but also vowel accuracy. *Second Language Research.*

Baills, F., Zhang, Y., Cheng, Y., Bu, Y. & Prieto, P. (2021). Listening to songs and singing benefit initial stages of L2 pronunciation. *Language Learning, 71*(2), 369-413.

Li, P., Baills, F. & Prieto, P. (2020). Observing and producing durational hand gestures facilitates the pronunciation of

novel vowel-length contrasts. *Studies in Second Language Acquisition, 42*(5): 1015 - 1039.

Xi, X., Li, P., Baills, F. & Prieto, P. (2020). Hand gestures facilitate the acquisition of novel phonemic contrasts when they appropriately mimic target phonetic features. *Journal of Speech, Language, and Hearing Research, 63*, 3571–3585.

Zhang, Y., Baills, F., & Prieto, P. (2020). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words. *Language Teaching Research, 24*(5), pp.666-689.

Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition, 41*(1), 5-32.