UNIVERSITÀ DEGLI STUDI DI PADOVA

Sede amministrativa: Università degli Studi di Padova
Dipartimento di Studi Linguistici e Letterari

CORSO DI DOTTORATO DI RICERCA IN:
SCIENZE LINGUISTICHE, FILOLOGICHE E LETTERARIE
CICLO XXXII

Restraining and Encouraging the Use of Gestures: Exploring the Effects on Speech

Coordinatore: Prof. Rocco Coronato

Supervisore: Prof.ssa Maria Grazia Busà

Co-Supervisore: Prof.ssa Pilar Prieto

Dottoranda: Alice Cravotta

Alice Cravotta

PhD Thesis

Restraining and Encouraging the Use of Gestures: Exploring the Effects on Speech

Supervisors:	Maria Grazia Busà Associate Professor, Università degli Studi di Padova Italy Pilar Prieto Research Professor, ICREA - Universitat Pompeu Fabra, Barcelona Catalonia
Reviewers	Núria Esteve-Gibert Associate Professor, Universitat Oberta de Catalunya, Barcelona Catalonia Maria Graziano Researcher, Humanities Lab, Lund University Sweden
Committee Members	Emanuela Campisi Professor, Università per Stranieri Dante Alighieri, Reggio Calabria Italy Francesco Cutugno Professor, Università degli Studi di Napoli Federico II Italy Marc Swerts Full Professor, Tilburg University The Netherlands

Università degli Studi di Padova, December 2019
Doctoral School of Linguistic, Philological and Literary Sciences

Abstract

Humans speak and produce hand gestures. It has been shown that hand gestures are a key component of the process of speech production and that the two modalities interact at multiple levels (semantic, pragmatic, temporal and prosodic). The underlying mechanisms and functions of this interaction are still a debated question in the field. This dissertation aims to investigate some of these aspects by exploring **the effects of restraining and encouraging the use of gestures** on **speech production** in a **narration task**.

Previous studies have shown that speech and gesture production are controlled by a common motor control system involving a neural network connected to Broca's area; they are also interdependent at a biomechanical level. This suggests that speech acoustics should be interdependent to body movements/gestures in terms of motoric control and biomechanics as well as communicative and prosodic needs. However, the possible direct effects of both restraining and encouraging the use of gestures on the **acoustic features of speech** have not received much attention in previous studies. Previous research has investigated how the **inability to gesture** impacts speech production in terms of e.g., speech fluency, length, and content. However, the mixed results obtained do not allow to draw solid conclusions. Moreover, research is lacking as to how **encouraging the use of gesture** should affect speech in similar terms.

Two empirical studies were carried out with 40 Italian speakers telling short narratives to a listener. In Study 1 (**Chapter 2**), speakers were restrained from gesturing by sitting on their hands; in Study 2 (**Chapter 3**) they were encouraged to make gestures while speaking. In both studies,

the speech of the target narratives was assessed based on a set of acoustic, prosodic and textual features, and specifically by analyzing speech discourse length (number of words and discourse length in seconds), disfluencies (filled pauses, self-corrections, repetitions, insertions, interruptions), speech rate and acoustic properties (measures of F0 and intensity). Additional qualitative and quantitative analyses on the data collected for Studies 1 and 2 are reported in **Chapter 4**, which also discusses some aspects of the methods adopted in the studies.

The **results** show that (1) the inability to gesture does not affect speech, which does not become significantly longer, more disfluent or monotonous; however, (2) enhancing the gesture stream by encouraging speakers to gesture can affect speech length and acoustics, as evidenced by an increase in F0 and intensity metrics. Moreover, as shown in Chapter 4, encouraging the use of gesture leads speakers (3) to produce more gestures and in a higher (more salient) gesture space; (4) to make use of more representational gestures and (5) to bodily enact characters and actions more often by using various multimodal cues.

Overall, this thesis investigates the potential functions of gesture production in the speech planning and articulation phases and provides evidence that gesture production can enhance some prosodic features of speech in semi-spontaneous narratives. Also, it shows that the inability to gesture is not detrimental to fluent speech production and spoken prosody in narrative speech. Further investigations in this direction would contribute to shed light on if and how gesture and prosodic structures are jointly planned and produced.

Sommario

Quando le persone parlano fanno gesti. È stato infatti dimostrato che la gestualità è una componente fondamentale della produzione orale e che le due modalità interagiscono a diversi livelli linguistici (semantico, pragmatico, temporale e prosodico). Tuttavia, i meccanismi sottostanti e le cause di questa stretta interazione sono ancora poco noti e, per questo, sono oggetto di molti studi sperimentali. Il presente lavoro si propone di investigare alcuni di questi aspetti, focalizzandosi su come **l'inibire** o **l'incoraggiare** l'uso della gestualità nei parlanti abbia influenze sulla loro **produzione orale** nel **discorso narrativo**.

Studi precedenti hanno proposto che l'espressione orale e la produzione gestuale condividano un comune sistema di controllo motorio mediato dall'area di Broca, e che vi sia anche una dipendenza biomeccanica tra le due modalità. Questo suggerisce che le **caratteristiche acustiche** del parlato siano strettamente legate alla gestualità sia per motivi biomeccanici o riconducibili al controllo motorio, sia in termini di esigenze comunicative e prosodiche. Tuttavia, la questione non è stata sufficientemente investigata a livello empirico. A questo fine, si ritiene interessante investigare se inibire o incoraggiare i parlanti all'uso della gestualità possa avere degli effetti sulle caratteristiche acustiche del parlato. Gli effetti di **inibire** l'uso della gestualità sono stati esplorati in studi precedenti che hanno analizzato, ad esempio, come questo abbia un impatto sulla fluenza e sulla lunghezza e contenuto del parlato. Tuttavia, questi studi hanno portato a risultati a volte contrastanti dai quali è difficile trarre conclusioni univoche. D'altra parte, non sono stati condotti studi sperimentali sui possibili effetti derivanti dall'**incoraggiare** l'uso della gestualità sui medesimi aspetti del parlato.

In questo lavoro, sono stati condotti due studi sperimentali nei quali 40 partecipanti di madrelingua italiana hanno descritto delle brevi storie a fumetti a un ascoltatore. Nel primo studio (**Capitolo 2**) la gestualità dei parlanti è stata inibita (ai soggetti è stato chiesto di sedersi tenendo le mani sotto le gambe), mentre nel secondo studio (**Capitolo 3**) i parlanti sono stati incoraggiati all'uso della gestualità. In entrambi gli studi, il parlato è stato analizzato a livello prosodico e testuale. In particolare, si sono analizzate: la lunghezza delle produzioni orali (in termini di numero di parole e lunghezza in secondi), le disfluenze (pause piene, auto-correzioni, ripetizioni, inserimenti di suoni o segmenti di parole, interruzioni), la velocità del parlato (sillabe al secondo), e una serie di misure acustiche che riguardano la frequenza fondamentale (F0) e l'intensità (volume) della voce. Inoltre, ulteriori analisi qualitative e quantitative sono state condotte sui dati raccolti. Tali analisi sono descritte nel **Capitolo 4**, nel quale vengono affrontate anche alcune questioni metodologiche.

I **risultati** mostrano che l'impossibilità di spiegarsi con la gestualità non è di particolare impatto negativo sul parlato (che infatti non diventa più lungo in termini di durata, né meno fluente o più monotono a livello acustico). D'altro lato, lo studio dimostra che chiedere ai parlanti di usare la gestualità nel descrivere le sequenze narrative può avere effetti sulla lunghezza delle storie narrate (in termini di numero di parole) e può interagire con i parametri acustici del parlato (alcune misure di F0 e intensità risultano più elevate). Infine, nel Capitolo 4 si mostra che incoraggiare i parlanti a usare la gestualità li porta effettivamente a: (1) aumentare il numero di gesti che usano nel raccontare le storie; (2) gesticolare in modo più "saliente", ovvero in uno spazio più alto (ad es., in corrispondenza del petto) rispetto a quando non ricevono tali istruzioni. Inoltre, le stesse istruzioni portano i parlanti a: (3) fare uso di più gesti rappresentativi; (4) usare maggiormente strategie multimodali di *enactment* (impersonificazione) che coinvolgono anche, ad esempio, espressioni facciali e movimenti della testa e delle spalle.

In generale, questa tesi si propone di investigare la possibile funzione della gestualità nel processo di espressione verbale, in particolare nella sua pianificazione e articolazione. I risultati suggeriscono che la produzione della gestualità può contribuire a enfatizzare alcune caratteristiche prosodiche del parlato. Allo stesso tempo, i risultati mostrano che l'impossibilità di spiegarsi tramite la gestualità può essere ben compensata nel discorso narrativo semi-spontaneo. In futuro, ulteriori studi sperimentali potrebbero contribuire a fare luce sulla possibilità che gestualità e strutture prosodiche siano pianificate e prodotte insieme durante il processo di produzione verbale.

Contents

Figures and Tables

List of Publications

Part of the contents in this dissertation is based on the following:

Chapter 2 **Cravotta, A.**, Prieto, P., & Busà, M. G. (under review). Exploring the Effects of Restraining the Use of Gestures on Speech.

Chapter 3 **Cravotta, A.**, Busà, M. G., & Prieto, P. (2019). Effects of Encouraging the Use of Gestures on Speech, In *Journal of Speech, Language, and Hearing Research*, 62(9), 3204-3219.

Chapter 4 **Cravotta, A.,** Prieto, P., & Busà M. G. (2019), Encouraging Gesture Use in a Narration Task Increases Speakers' Gesture Rate, Gesture Salience and Production of Representational Gestures. In Proceedings of the 6th Conference on Gesture and Speech in Interaction (GeSpIN), Paderborn, 11-12 September 2019, pp. 21-26.

Introduction

Humans speak and produce hand gestures. These spontaneous movements that accompany speech across languages and cultures are a pervasive part of all human language. Gestures develop together with speech in children (e.g., Graziano, 2014; Gullberg, De Bot, & Volterra, 2008; Özçalişkan & Goldin-Meadow, 2005) and break down together in patients with language impairments (e.g., Mol, Krahmer, & van de Sandt-Koenderman, 2013). Also, people gesture while talking over the phone (e.g., Bavelas, Gerwing, Sutton, & Prevost, 2008) and congenitally blind individuals gesture while talking to each other just like people with no visual impairment do (e.g., Iverson & Goldin-Meadow, 1998). This suggests that **gestures** have some role in the **thinking and speaking processes** (e.g., Abner, Cooperrider, & Goldin-Meadow, 2015; Church, Alibali, & Kelly, 2017). Before discussing gesture in more detail, let us see what is commonly known about gestures.

At some point last year, I came across an Instagram post by the italian journalist and writer Roberto Saviano. The journalist was portrayed in a photo while speaking and gesturing, and the image was captioned as follows: *(EN) If I didn't gesture the words would choke in my throat. If I didn't* *gesture I couldn't accompany those words to the heart of the interlocutor. If I didn't gesture I would speak only halfway* (Roberto Saviano)[1]. What I found interesting was the presence of around 280 comments below it, many of them being some sort of **commonplace** about why we gesture when we speak and what is gesture for. It was interesting (and fun) to see how people feel about it. Though these comments may be considered trivial, they suggest a few intuitive things about gesture; for example, gestures are considered *necessary* and can hardly be avoided while speaking:

Ma perché, sul serio si può parlare senza gesticolare manco un po'?	Seriously. Can we even talk without gesturing at least a little?
Ah ah! te l'ho sempre detto! Parli con le mani.	Ah ah! I've always told you this: you speak with your hands.
Una lingua nella lingua che rende le parole corpo e colore.	A language in the language that makes the words body and color.
Una volta un'amica si è mostrata stupita perché parlavo tenendo le mani in tasca…	Once a friend of mine was surprised I was talking with the hands in the pockets…
Anche per me è difficilissimo non gesticolare. è parte di me!	Even for me, it's very difficult not to gesture. It's part of me!
è più forte di noi, non riusciamo a farne a meno!	It's stronger than us. We can't do without it!

Also, gestures are intuitively believed to be part of an individual's personality and style and also linked to one's cultural/linguistic background (e.g., Italy, Naples):

[1] Se non gesticolassi le parole mi si strozzerebbero in gola. Se non gesticolassi non riuscirei ad accompagnarle sino al cuore dell'interlocutore. Se non gesticolassi parlerei solo a metà (Roberto Saviano, September 2018).

Se non parlo muovendo le mani non sono io stessa.	If I don't speak moving my hands I'm not myself.
Se non gesticolassi non saresti italiano...	If you didn't gesticulate you wouldn't be Italian…
Per noi napoletani i gesti vengono prima delle parole poiché il concetto viene espresso prima con le mani. Siamo "gesticolanti culturali".	For us, Neapolitans, the gestures come before the words because the concept is expressed first with the hands. We are "cultural gesturers".

Some other comments showed that gestures are also commonly believed to aid to express concepts better, and to have a more effective impact on the listener:

È un gesticolare espressivo e comunicativo!	Yours is an expressive and communicative gesticulation!
Gesticola! Rende meglio ciò che dici.	Do gesture! It better explains what you say.
Sottolinea il pensiero, enfatizza un'emozione, rafforza un concetto!	It underlines a thought, emphasizes an emotion, strengthens a concept!
Se tu non gesticolassi, io non ti crederei…	If you didn't gesture I wouldn't believe you…
Gesticolare è l'eco delle parole, rimbombano piú forte nel cuore di chi ti osserva e ascolta.	Gesticulation is the echo of the words, which reverberate louder in the heart of those who watch and listen to you.
È che con i gesti le parole prendono forma, vita. Arrivano esattamente dove devono arrivare; al cuore, alla pancia, allo stomaco.	It is with gestures that words take shape and life. They arrive exactly where they are meant to. To the heart, to the stomach.

Some people also referred to the fact that gesture can have effects on the speakers themselves, and can also pair well with the **sound** of their **voice**:

Vero! è una vera e propria forma di estensione dei tuoi	True! They expand your thoughts.

pensieri.

È un modo per rendere pietra le parole... è un modo per sentire, modellare, scolpire le parole, proprio come uno scultore. È cosi per me.	It's a way to make words stone ... it's a way to feel, shape, sculpt words, just like a sculptor. It is so for me.
Sono come te, se non gesticolassi mi sentirei muta…	I'm like you. If I didn't gesture I would feel speechless…
I gesti sono una seconda voce, un controcanto, un contrappunto forse? alla musicalità della tua voce.	Gestures are a second voice, a counter-melody, a counterpoint perhaps? to the musicality of your voice.

Though all these comments are drawn from the Italian context and may not be fully representative of other non-Italian contexts, they show that people have some general ideas about gestures as part of the human need to communicate. At the same time, most of the comments seem to limit gestures to an "*add-on*" (Kendon, 2008) of language, their function being only to emphasize, integrate, and ornament speech. However, the integration between gesture and speech seems to be more profound: gestures are inseparable from language because they reflect the thinking process and the *imagery* that underlies speaking (McNeill, 2005, 2017). As McNeill (2017) puts it, gestures must be considered *part* of language and, for example, "even if, for some reason, the hands are restrained and a gesture is not externalized, the imagery it embodies can still be present, hidden but integrated with speech" (p. 78). In fact, when people are not speaking, but instead are in silent, non-communicative, problem-solving situations, they can still produce **co-thought gestures** (Chu & Kita, 2011, 2016). Possible explanations for this can be found in the idea that gestures are capable to support more general cognitive processes (Goldin-Meadow & Alibali, 2013; Hostetter & Alibali, 2019; Kita, Alibali, & Chu, 2017; Pouw, de Nooijer, van Gog, Zwaan, & Paas, 2014). Precisely, empirical findings have shown that gesture production influences thinking: problem solving (Beilock & Goldin-Meadow, 2010; Chu & Kita, 2011;

Pouw, Mavilidi, van Gog, & Paas, 2016), learning (Broaders, Cook, Mitchell, & Goldin-Meadow, 2007), and memory (Cook, Yip, & Goldin-Meadow, 2010, 2012; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Ping & Goldin-Meadow, 2010). The tight bond between gestures and speech may lie on the proposed common developmental origins of the two (Iverson & Thelen, 1999) and can also be explored in the evolutionary context, where gesture has been proposed as a potential starting point for human language (**Gesture First Hypothesis**; Arbib, 2012; Armstrong & Wilcox, 2007; Corballis, 2003; Stokoe, 2001; Tomasello, 2008).

In essence, gestures serve rich **cognitive functions** but also have a rich **linguistic dimension**. In fact, they integrate with language at all levels of linguistic structure (e.g., Abner et al., 2015). As a consequence, people cannot do without gesturing when thinking and speaking. Also, together with their cognitive and linguistic dimensions, gestures have a **communicative** dimension as well. However, disentangling the communicative functions of gestures and reconstructing the role of *intentionality* (i.e., communicative intention) is not an easy task. First, this would require an understanding of whether gestures are designed in a communicatively-efficient way for addressees, goals, and contexts; second, it would also require an understanding of the extent to which such adaptation to addressees, goals, and contexts is *intentional* (Campisi & Mazzone, 2016).

Over the past decades, these questions have gathered together scholars from different disciplines that have studied, for example, the relationship between gestures and speech in terms of semantics, pragmatics, syntax, phonology, temporal alignment; the role of gesture in social interaction and human cognition; the development of gesture and language in children; the decay of gesture in language impairments; the creation of codified/shared gestural forms from spontaneous gestures; the relationship between gesture and signs; the role of gesture in language origins and evolution. The interest on these topics has led to the flourishing of experimental studies that have had an impact on different fields such as cognitive science, psychology, psycholinguistics, cognitive linguistics, developmental psychology and linguistics, speech therapy, neuroscience, primatology, human communication studies, computational multimodal research. In this wide context, this dissertation focuses on co-speech gestures and their role for the person who produces them while speaking.

Mechanisms and functions of gesture

Previous research has contributed a clearer idea of how people gesture when speaking (see Sections 1.1 and 1.2) and has highlighted some patterns of gesture use in language production and cognitive processes (Section 1.3). Following the reasoning in Church, Alibali, & Kelly (2017)'s milestone book, understanding gesture has historically required to approach the question from two perspectives. The first is to examine the **functions** of gesturing, its purpose, the why rather than the how (Novack & Goldin-Meadow, 2017). Speculating about possible functions/purposes for gesture requires exploring their **effects** on producers and observers, and also accounting for these effects in cognitive, neural and social terms. The second way to gain further understanding of gesture is to focus on the gesture **mechanisms**, that is, exploring their **cause**: What makes us gesture? Where does gesture arise from? What types of events makes gesture likely to occur? (e.g., Novack & Goldin-Meadow, 2017). In essence, both perspectives (e.g., function and mechanism) are critically related and are needed for the development of comprehensive models.

Goals: Restraining and encouraging the use of gestures

The aim of this dissertation is to explore the *effects* of gestures on narrative speech production. By doing so, we expect to gain further insight on the role of gesture production for speech fluency and acoustic modulation.

To explain the mechanisms of gesture and speech production, researchers have developed psycholinguistic models that integrate the gesture stream in the process of speech production (Section 1.3). Moreover, research has shown that speech and gesture production are both controlled by a common motor control system engaging a neural network connected to the Broca's area (e.g., Bernardis & Gentilucci, 2006; Marstaller & Burianová, 2015) and that the two modalities can be interdependent at a biomechanical level (e.g., Pouw, Harrison, & Dixon, 2019).

One way to explore the dynamic interaction between speech and gesture during thinking and speaking has been by manipulating the presence of gestures, by either preventing their use (e.g., Hoetjes, Krahmer, & Swerts, 2014; Rimé, Schiaratura, Hupet, & Ghysselinckx, 1984) or encouraging it (e.g., Chu & Kita, 2011; Hostetter, Alibali, & Kita, 2007). Precisely, as it will be reviewed in Section 2.1, previous studies have investigated the direct effects of **restraining gestures** on speech fluency, content, and length. These studies have used heterogeneous methodologies and tasks and have yielded mixed results. For example, some detrimental effects of the inability to gesture on fluency have been found in visual objects description (Morsella & Krauss, 2004) and low codability abstract lines drawings descriptions (Graham & Heywood, 1975). By contrast, two studies which elicited speech by asking participants to describe motor tasks (Hoetjes et al., 2014; Hostetter et al., 2007) did not find effects of restraining gestures on fluency nor speech length. In studies using story retellings, unclear results have been found on both fluency and speech content and planning (Finlayson et al., 2003; Jenkins et al., 2018; Rauscher et al., 1996). From these studies, it is not possible to draw solid conclusions. Perhaps, the inability to gesture impacts speech differently depending on the different tasks. Also, from the heterogeneous types of analysis proposed in the different studies, it is hard to draw clear commonalities. Moreover, other possible effects of the inability to gesture on speech can be explored; for example, the question of how it could affect speech on the **acoustic features** has thus far received little attention (but see Hoetjes et al., 2014).

On the other hand, participants have been instructed to gesture in previous studies that explored the effects of **encouraging the use of gesture** on activities such as problem solving (Beilock & Goldin-Meadow, 2010; Chu & Kita, 2011), learning math (Broaders et al., 2007), second language pronunciation (Baills, Suárez-González, González-Fuente, & Prieto, 2019; Llanes-Coromina, Prieto, & Rohrer, 2018), or retelling tasks (Vilà-Giménez & Prieto, 2018). However, as it will be discussed in Section 3.1, the possible direct effects of actively encouraging the use of gestures on connected/semi-spontaneous speech have been overlooked. Precisely, there is a lack of studies on how the instruction should affect speech on levels such as fluency, content, and length and also its **acoustic properties** (but see Krahmer & Swerts, 2007; Pouw, Harrison, & Dixon, 2019).

Though encouraging the use of gesture is not exactly the polar opposite of restraining their use (i.e., both conditions can lead to side effects that can interfere differently with thinking and speaking), our view is that the two types of instructions can in principle provide complementary evidence on the potential functions of gestures for speech production.

The present dissertation reports on two twin studies exploring the **effects** of **restraining** and **encouraging** the use of gestures on **speech**. The two studies elicit narrative speech in an ecologically valid way to investigate the speech production process in its spontaneous full functioning (without taxing/challenging specific "modules" of the process, e.g., memory, lexical access, etc.). The analysis of speech is based on features that have been previously investigated in relation to hand gesture restriction and only marginally in relation to hand gesture encouragement. These are: speech discourse length, fluency (number of filled pauses, self-corrections, repetitions, insertions, interruptions, silent pauses, speech rate), as well as acoustic features of speech (fundamental frequency and intensity measures).

Outline of the present dissertation

This dissertation starts with a chapter that is devoted to the general theoretical background (Chapter 1) including the definition of gesture and the main questions that revolve around gesture and speech integration. The dissertation is then composed of two empirical studies (described in Chapters 2 and 3) that are self-standing contributions with their own abstract, introduction, methods, results and discussion sections. The two studies explore the effects of gesture restriction (Chapter 2) and encouragement (Chapter 3) on narrative speech. Additional qualitative and quantitative analyses on the data collected for Study 1 and 2 are reported in Chapter 4, together with some methodological remarks. Chapter 5 is devoted to a general discussion of the results and proposes some open questions for future research.

1 Theoretical Background

1.1 What is gesture

The field of gesture studies flourished in the last decades. The core questions addressed by linguists, psychologists, and neuroscientists have been *how* and *why* people gesture while speaking (Kendon, 1997; McNeill, 1992; for a review, see Abner et al., 2015). To answer these questions, gestures have been defined and classified in many ways. In general, they can be intended as body movements produced in communication exchanges (Kendon, 2004). This broad definition can be further specified in more detailed categorizations. The **Kendon's continuum** (McNeill, 1992, 2005) classifies gesture as follows, depending on different dimensions:

gesticulation	>>	**pantomimes**	>>	**emblems**	>>	**sign language**
speech present		speech absent		speech present or absent		speech absent
no linguistic properties		no linguistic properties		some linguistic properties		linguistic properties
not conventionalized		not conventionalized		partly conventionalized		fully conventionalized

Moving from left to right of the continuum the obligatory presence of speech decreases and the stability of the meaning, standardization and linguistic properties of the hand movement increases. **Gesticulation** is the most pervasive type of gesture, in that it is any motion produced spontaneously in the context of speaking, and it embodies "meaning relatable to the accompanying speech" (McNeill, 2005, p. 5). It mainly involves hands and arms movements, but it is not limited to these body parts (shoulders, head, face, legs can be part of it). Gesticulations can be regarded as "symbols of action, movement and space" (McNeill, 1992, p.2). They can be also referred to as **co-speech gestures**, or often, for brevity, simply as gestures. Gestures are connected with the concurrent speech at the semantic, pragmatic, and discourse levels (see Section 1.2). Along the continuum, a **pantomime** is a (sequence of) gestures produced without speech that can convey a whole narrative without being conventionalized or having any linguistic properties. Next to it, **emblems** (or "quotable gestures", Kendon, 1992) are conventionalized hand movements whose meaning is shared in a (linguistic) community (e.g., the thumb up gesture that often stands for OK). They can be listed (e.g., Munari, 1963) and "glossed" (e.g., Poggi, 2002) because they have meaning, rules of well-formedness, contexts of use, synonyms, grammatical and pragmatic classification that are not present in co-speech gesture. A famous example of emblems is the Italian *hand purse* or *mano a borsa*, that can be glossed as a gesture of *question* (Poggi, 1983). From the Italian repository, a second example can be given: the *pistol-twist* gesture. It is performed with the hand having thumb and forefinger extended at right angles, and rotated back and forth and it stands for "nothing" (Kendon, 1992). On the rightmost side of the continuum, there are **sign languages**. As for *signs*, elements such as handshape, movement, location, orientation, and non-manual elements are the main building blocks of the sign language phonological and morphosyntactic structure. Importantly, the Kendon's continuum excludes movements such as **self-adaptors** (like scratching one's nose or touching one's hair). In the present dissertation, from now on, any mention of **gesture** will refer to gesticulations or co-speech gesture.

Gestures have been observed and described in various ways over the decades, and examined on different levels (e.g., the extent to which they are iconic or carry symbolic meaning, their semantic properties, their level of conventionalization, etc.). In ancient times (first century), Quintilian described gestures as important oratory components and devoted to their description a whole chapter of the rhetoric compendium *Institutio Oratoria* (book XI, chapter 3). In modern times, different proposals have contributed a more fine-grained description of gestures as communicative movements and have referred to their semantics, shape and communicative/pragmatic functions (e.g., Bavelas, Chovil, Lawrie, & Wade, 1992; De Iorio, 1832; Efron, 1941; Ekman & Friesen, 1969; McNeill, 1992). The next section will give a general overview of how gestures can be classified and described and will focus on the most influential proposals.

1.2 Co-speech gestures: Types and minimal units

Co-speech gestures can be grouped, classified and described along different dimensions, based on their form, semantics, pragmatic functions and temporal coordination with speech (Kendon, 2004; McNeill, 2005). For example, in terms of their form, hand gestures can be described as **iconics** when they depict images of concrete entities or actions, via the shape of the gesture (hand shape) or manner of execution (trajectory, direction). Iconic gestures are imagistically relatable to the ongoing speech. For example, making climbing movements with the hands or fingers when expressing verbally a "climbing" action, or making a rounded shape with the two hands to describe a "tuna can". **Metaphoric** gestures are those that depict the abstract in terms of the concrete (McNeill, 1992, p. 14) or rather they engage the cognitive process of understanding something in terms of something else via cross-domain mapping (Cienki & Müller, 2008). For example, space can be used metaphorically to represent time (e.g., pointing back to refer to "past" or moving an open hand over the shoulder to refer to "yesterday"). Moreover, one of the possible semiotic components of gesture is deixis. **Deictic gestures** (or pointing gestures) can be performed with any body part, and are used to indicate a certain direction, location or object in the space, though objects or entities pointed at are often not physically present (see Kita, 2003); these can be referred to as **abstract pointings**. Iconics, metaphorics and abstract pointings can be referred to as **representational gestures** (McNeill, 1992).

However, hand gestures can also be non imagistically relatable to the ongoing speech. These can be referred to as **non-referential** or **non-representational gestures**. They can appear in a great variety of shapes and forms and have several functions. For example, **interactive** gestures do not *represent* the content of the co-occurring speech but instead are useful to frame it into the discourse (Abner et al., 2015). In fact, non-representational gestures ensure turn-taking exchanges and dialogic interaction (e.g., Bavelas, Chovil, Coates, & Roe, 1995; Cooperrider, Abner, & Goldin-Meadow, 2018); or express speakers' stance, parsing/punctuational and marking focus and information status (Kendon, 2004, 2017). Such gestures with pragmatic functions can be used by children too, and develop together with the latter's ability to structure discourse and use of meta-discourse connectives (Graziano, 2014); also, gestures can pave the way to children's pragmatic development of knowledge state and politeness (Hübscher, 2018; Hübscher, Garufi, & Prieto, 2019).

In the literature, some types of gestures such as flicks of the hands or back-and-forth/up-and-down movements have been variously defined as **beats** (Kendon, 2004; McNeill, 2005), or *batons* (Efron, 1941; Ekman & Friesen, 1969) since they seem to have a more predominant temporal highlighting function (McNeill, 2005, p. 41). However, any type of gesture can in principle being built upon the *rhythmical pulse* underlying speaking (Tuite, 1993). Precisely, all gesture types, including representational and non-representational ones, can have both rhythmic and prosodic components and express a range of pragmatic and interactive functions (e.g., Prieto, Cravotta, Kushch, Rohrer, & Vilà-Giménez, 2018; Shattuck-Hufnagel & Prieto, 2019).

Gesture temporal structure

It has been observed that co-speech gestures unfold by passing through a series of phases (Kendon, 1980, 2004; McNeill, 2005) which constitute their basic *anatomy* (McNeill, 2005). These phases are organized around the **stroke** phase. The stroke is the part of the gesture that carries the meaning and takes on the gesture's communicative role. It is the phase of the excursion in which the movement dynamics of effort and shape are manifested with greatest clarity (Kendon, 1980). The exact point in time in which the movement reaches its 'peak of the peak', that is, the kinetic 'goal' of the stroke can be referred to as *apex (or gesture peak)* (Loehr, 2012). According to Loehr (2012), apexes can be identified as either the moments in which the arms/hands maintain their maximum extension or any other relevant dynamic points within the stroke (e.g., changes of direction).

Any prototypical gesture, as described by Kendon (1980), starts with a **preparation** phase. In the preparation phase, the hands start departing from a *rest* position to reach the *stroke* phase. The hands can then return to a rest position again (**retraction** phase). There might not be a retraction phase if the speaker moves directly from a stroke to a new *stroke*. Together these phases constitute a **gesture phrase**. Figure 1 represents a short gesture phrase from a video included in the data collected for this dissertation. The *stroke* appears in the central frame and it is preceded and followed by two rest positions. The text reported above the figure is a transcription of the speech context in which the gesture appeared. The square brackets enclose the speech co-occurring with the three images, and bold delimits the speech accompanying specifically the *stroke* phase.

(IT) C'è un gatto con il suo padrone davanti a una scatola che sembra vuota [… **ee a una ciotola** di cibo]

(EN) There's a cat with a guy in front of a box that seems empty [… **aand of a bowl** of food]

Figure 1 Example of a short gesture phrase The first picture shows the start time of the preparation phase, the second picture shows the stroke, and the third one represents the end of the retraction phase.

This is a simple example, as gesture phrases can be longer and combined with each other, and include, for example, gesture holds (i.e., the momentary suspension of a movement). Holds can include either a held/interrupted preparation or any pre- and post-stroke holds phases (Kita, van Gijn, & van der Hulst, 1997). For example, the stroke can be held in place until the related co-expressive speech is over (i.e., post-stroke hold). Conversely, a pre-stroke hold can occur when the stroke is ready to be completed but it "waits" for its related linguistic segments to be uttered.

As originally proposed by McNeill (1992), the co-occurrence between gesture strokes and certain speech segments suggests that the two relate at least at three levels: (1) co-occurring gestures and speech relate to the same idea unit (**semantic synchrony rule**), (2) they have the same pragmatic functions (**pragmatic synchrony rule**) and that (3) gesture strokes (or apexes) occur in proximity of the most prominent syllables (**phonological synchrony rule**). These three levels of synchrony have been the focus of many empirical studies that have increasingly shown the complexity of the temporal interaction and coordination between gesture and speech. The dynamic co-production of speech and gestures has also been explored to better understand the role of gestures for speech production and their communicative and cognitive functions. The next section provides an overview of the main theoretical accounts that give potential explanations as to how and why speech and gesture are co-produced and are synchronous on different levels. Also, the next section will zoom into some empirical findings on how gestures temporally interact with **prosodic features** and **disfluent speech** specifically, since these two aspects are relevant for the two studies presented in Chapters 2 and 3.

1.3 Gesture and Speech Production

Speech and gesture production models

Many of the proposed cognitive models accounting for gesture production draw upon **Levelt** (1989)'s **'blueprint for the speaker'** framework, that explains the process of speech production from conceptualization to overt speech production. In brief, the model distinguishes three stages in the speech production process: **conceptualization** (construction of communicative intention from procedural knowledge through working memory), **formulation** (preverbal message is the input to the formulator and it is encoded in grammar, phonology, lexicon), and **articulation** (message is overtly articulated resulting in auditory signal). The whole process is carried out by different modules interacting with each other. First of all, in the **conceptualization phase**, the *conceptualizer* converts the communicative intention into a *preverbal message* with semantic structure. The preverbal message is then encoded (*grammatical encoding*): preverbal lexical concepts activate the corresponding lemmas from the *mental lexicon*. This is where the **formulation phase** starts. When a lemma is selected, the speaker gets access to the morpho-phonological forms of the lemma corresponding to the lexical concept (*morpho-phonological encoding*). The phonological encoding consists of the building of the *phonological score* of the utterance i.e., syllabified words, phrases and intonation patterns. Finally, the *phonetic encoding* ensures the generation of the *articulatory score* of the utterance: in this phase, each syllable in the phonological score activates articulatory gestures from the *syllabary*. Here is where articulatory gestures are generated and triggered, for the final **articulation phase.** Overt speech results from the execution of the articulatory score by speech articulators (articulation phase).

In **Levelt-inspired models** (e.g., Kita & Özyürek, 2003; Rauscher et al., 1996) a new gesture stream is proposed to be integrated into this process. A matter of theoretical debate is where and how the two streams interact. Though the study presented in this dissertation is not aimed at taking a particular stance towards one or the other models, a short overview of the main existing speech-gesture hypotheses and models will be given below as background information.

According to the **Lexical Retrieval hypothesis** (Krauss, Chen, & Gottesman, 2000; Rauscher et al., 1996), (iconic) gestures (termed *lexical gestures*) aid lexical access in speech production via cross-modal priming: Gestures activate spatial-dynamic features of concepts, which then activate the lexical items to be retrieved from the mental lexicon (see also Hadar & Butterworth, 1997). Therefore, according to this view, producing a gesture helps speech production in a **late stage** of the speech production process (*formulation* stage, where words are retrieved from the lexicon and the phonological form is generated).

By contrast, other models propose that gestures play a role at the **early stage** of *conceptualization*. An influential theory related to this idea is the **Growth Point Theory** (McNeill, 2005; McNeill & Duncan, 2000). According to it, the utterance grows from a holistic representation that constitutes the initial "seed" of the utterance (the growth point, GP). This seed includes both *imagistic* and *symbolic* information. As speech unpacks, the **imagistic part** of a GP is expressed globally and synthetically into a gesture; the **symbolic part** of it needs to be turned into speech. This requires meanings to be distributed analytically into a linear series of morphemes, words, and phrases. According to the GP Theory, because gesture and speech come from the same GP, they are synchronized from a temporal and semantic point of view.

The GP's idea of the unity and dialectics of linguistic expressions and spatio-motoric (imagistic) representations is incorporated in the **Interface Hypothesis (IH)** (Kita & Özyürek, 2003) and its cognitive architecture. The cross-linguistic empirical findings by Kita & Özyürek (2003) showed that the lexical possibilities and clausal structure of a language can influence gesture production. They found that, for example, if in a language there is no word for "swing", the gesture will not often represent an arc-trajectory. Explaining these findings, Kita & Özyürek (2003) introduced the Interface Hypothesis. In this model, Levelt's conceptualizer is split into two modules: (1) the *Communication Planner*, where the general communicative intention is generated and where the modality of expression (speech vs gesture) is selected; (2) the *Message Generator*, which formulates the exact proposition to be verbally formulated. In the model, the speech generating module (*formulator*) and the gesture/action generating module (*action generator*) are two independent streams that constantly interact bidirectionally so that the gestural content can be "shaped on-line by linguistic formulation possibilities" (Kita & Özyürek, 2003).

The **Asymmetric Redundancy Sketch Model** (de Ruiter, 2017), and its original previous version (De Ruiter, 2000), also incorporates the foundations of the GP theory. It assumes that the Leveltian's *conceptualizer,* (a module that converts the communicative intention into a *preverbal message* with semantic structure) also generates an abstract *sketch* containing the imagistic information to be converted into gesture by the *gesture planner*. The *formulator*, in synchrony with the gesture planner, encodes the preverbal message into verbal utterance and this ensures the gesture-speech synchrony.

However, the process of how speech is capable to express both symbolic and imagistic information can be explained by assuming that speech and gestures arise from different (cognitive) sources: gestures are generated by *spatio-motoric thinking,* while speech arises from *analytical thinking* (Alibali, Yeo, Hostetter, & Kita, 2017). Spatio-motoric thinking provided by gesture can be "an alternative informational organization that is not readily accessible to analytic thinking" (p. 63) (Kita, 2000), therefore, gesturing helps to organize rich spatio-motor information into packages suitable for speaking, as claimed by the **Information Packaging Hypothesis (IPH)** (Kita, 2000). This proposal is further elaborated in the more recent **Gesture-for-Conceptualization Hypothesis** (Kita et al., 2017). This framework is based on the assumption that gestures are produced by a general-purpose *action generator* that also generates practical actions. The proposal holds that gesture's capability of *schematizing* information affects not only speaking but cognitive processes in general, by activating, manipulating, packaging and exploring spatio-motoric information. This idea is compatible with the embodied cognition framework (Glenberg, Witt, & Metcalfe, 2013) and, specifically, with the **Gestures as Simulated Action framework (GSA)** (Hostetter & Alibali, 2008, 2019). The GSA framework is based on the assumption that any *thinking* involving visuospatial or motor imagery requires activation of the motor system (sensorimotor simulation) and that gestures arise from and reflect this motor activity. That is to say, speakers gesture because they *simulate* action and perceptual states as they think. Therefore, the GSA model predicts that the likelihood of a gesture appearing at a particular moment depends on:

- the extent to which producers evoke a **mental simulation** of an action or perceptual state while speaking. For example, when speakers rely on a stored verbal or propositional code as the basis for speaking or thinking about an idea, a gesture is less likely to occur (or it will be a non-representational gesture);
- the **co-occurrent activation of the motor system** for speech production. In other words, a gesture is more likely to occur when the motor system is also engaged in producing speech;
- the height of the producer's **gesture threshold** at a given moment. In other words, speakers have their own *resistance* to overtly producing a gesture. This *resistance* can be seen as a variable threshold that depends on the single individuals and the communicative context.

To conclude, the theoretical models reviewed above account for most of the empirical findings on speech and gesture production, predicting that gestures serve a functional role for speaking, by alternatively helping speakers in the **lexical retrieval phase** (Krauss et al., 2000; Rauscher et al., 1996), or in the **conceptualization phase** (de Ruiter, 2017; Kita et al., 2017; Kita, 2000; Kita & Özyürek, 2003), and they also explain how gestures are capable to support speech production by in fact supporting more general internal **cognitive processes** (e.g., GSA model) (Goldin-Meadow & Alibali, 2013; Hostetter & Alibali, 2019; Kita et al., 2017; Pouw et al., 2014).

The next two sections will serve as further background information for the two studies presented in Chapters 2 and 3. They will focus on gestures and speech production by specifically referring to gesture and its relation with (1) **prosodic features** and (2) **disfluent speech**.

Gesture and speech integration: Temporal alignment and acoustics

Behavioural and neuroimaging studies have suggested that spoken language and arm gestures are controlled by the same **motor control system** (Bernardis & Gentilucci, 2006) and that both speech and co-speech gesture production engage a neural network connected to Broca's area (Marstaller & Burianová, 2015; see Gentilucci & Dalla Volta, 2008 for a review). This is in line with findings and theoretical models described so far, claiming that gesture and speech constitute a single and integrated process. Krivokapić, Tiede, & Tyrone (2017) recorded concurrent speech and body movements using electromagnetic articulometry for vocal tract movements and a motion capture system for body movements. They showed that the prosodic structure itself can be similarly expressed in both modalities: e.g., final lengthening at prosodic boundaries are accompanied by manual gestures lengthening, which also occurs during prominence. This suggests that gesture production and its relation to speech needs to be explored at the prosodic and phonological level. This idea was somehow present in Kendon (1972, 1980)'s studies which analyzed a 2-minutes video excerpt from a film made in a London pub (where a single speaker was talking during a group discussion with an American anthropologist). He observed that "each level of organization distinguished in the speech stream was matched by a distinctive pattern of bodily movement" (Kendon, 1980, p. 210). In other words, he described how prosodic structures (named *tone units*, *locutions*, *locution groups,* and *locution clusters*, or "paragraphs" of the *discourse*) are **hierarchically co-organized** with body postures, co-speech gestures structure and kinematics. For example, each locution within a locution cluster was characterized by a particular type of movement/posture.

More recent research has shown that gesture strokes, prominent parts of gestures (peaks) or gestural "hits" can, in turn, align with prominent part of speech, and prosodic structure of the spoken utterance (e.g., Esteve-Gibert, Borràs-Comes, Asor, Swerts, & Prieto, 2017; Esteve-Gibert & Prieto, 2013; Loehr, 2012; Shattuck-Hufnagel & Ren, 2018). However, a question to be addressed is whether gestures and prosodic structures are jointly planned by speakers to occur simultaneously and coordinate with each other (Shattuck-Hufnagel, 2019; Shattuck-Hufnagel & Ren, 2018). In effect, the result of this **temporal alignment** is that the two modalities (speech and gesture) can work in tandem: for example, speakers can employ acoustic cues (e.g., pitch accents) to mark a word, or information status (Swerts, Krahmer, & Avesani, 2002), and they may also use visual cues like gestures, head nods, eyebrow movements for the same purpose. However, the precise nature of this temporal (and functional) coordination seems to bear many open questions (Wagner, Malisz, & Kopp, 2014).

Recently, it has been claimed that this gesture-speech synchrony is mediated by a more direct **biomechanical interdependence** between speech and body movements (Pouw, Harrison, et al., 2019). That is, the repetition of hand movements directly affects the action of expiration-related muscles and this can directly affect some of the (prosodic) features of speech, like F0 and amplitude. In their study, Pouw and colleagues found that beat-like movements with high physical impetus affect phonation properties in terms of periodicity, that is, a downbeat to upward movement phase of the beat aligns temporally with a peak in amplitude envelope and a peak in F0.

Also, there is evidence of a direct effect of asking speakers to produce a gesture on the **spectral properties of speech**. Krahmer & Swerts (2007) showed that producing a beat gesture leads to changes in how prosodic prominence is realized in speech (in particular on the duration and on the higher formants, F2 and F3 of the target word where the gesture is produced). This means that gestures can have effects on the accentual strength (prominence) of the co-occurring word. What is unclear is from which mechanisms these effects originate. For example, if they are related to direct physical impetus that goes from arm movements to oral articulators (Pouw, Harrison, et al., 2019), or if they are to be explained in terms of the neural-cognitive mechanisms of a common underlying system that controls arms and oral articulators (Hammond, 1990; Flanagan, Ostry, & Feldman, 1990). Bernardis & Gentilucci (2006)'s study on the influence that gestures and speech have on each other when simultaneously produced showed that speech and symbolic gestures (not just meaningless arm movements) influence each other when produced simultaneously. For example, waving bye-bye accompanying "Hello" has effects on the spectrum (F2) of the co-produced speech, while meaningless hand movements do not have comparable effects. At the same time, they found that gesture kinematics is reduced when co-produced with meaningful/related words and that this does not happen with pseudo-words. Therefore, they claim that spoken words and symbolic gestures are *functionally* related and are coded by a **common communication system** involving Broca's area. This is in line with the language evolution perspective that sees speech as evolving from a primitive communication system based on gestures (see, among others, Corballis, 1999; Hewes, 1973).

The abovementioned proposals on gesture-speech interaction are not mutually exclusive, and our view is that the mechanisms underlying the integration of gestures and speech need to be explained in this wide context. As Shattuck-Hufnagel (2019) proposes, issues related to speech and gesture synchronization and mutual interactions need to be taken into account for the development of a comprehensive model that, while accounting for the speech planning processes, also includes a mechanism explaining the speech-gesture alignment, the higher level prosody that governs surface phonetic variability, and the functional relationship between gestures and speech structure and meaning.

Gesture and disfluent speech

As speech unfolds disfluencies are pervasive. Speakers hesitate in many ways by making pauses, cutting-offs, repairs with a fresh start or a phoneme correction (e.g., rephrasing wrong sentences, substituting a speech segment), filled pauses (*uh*, *uhm*). A *self-monitoring system* ensures that speakers detect the problem and deal with it (Levelt, 1999). The speaker must decide whether to correct it in some way (e.g., by interrupting speech or not) and when and how to do it (for a review of speech monitoring theories see Postma, 2000).

The question of what happens to gesture when speech is disfluent has been explored as a way to gain more insight into the speech-gesture production mechanisms and the self-monitoring mechanism itself.

Empirical studies on speech and gesture production during disfluency have suggested that when speech stops, so does gesture (e.g., Graziano & Gullberg, 2018). In fact, gesture can be highly sensitive to speech disfluencies: as speech is suspended and resumed, gestures can also be suspended and resumed in systematic temporal coordination with speech (Seyfeddinipur, Kita, & Indefrey, 2008). Precisely, there is evidence that gesture can fore-shadow speech problems to be solved that is, gestures are often interrupted before speech is (Seyfeddinipur & Kita, 2001).

In general, the relationship between gesture production and disfluent speech has been explained in at least two ways: (1) if gestures have mainly a (lexical) compensatory role and contribute to the **lexical retrieval phase** of speech production (Krauss et al., 2000; Rauscher et al., 1996), they should occur more frequently during speech disfluencies; (2) if they are at play in the **conceptualization** phase (de Ruiter, 2017; Kita et al., 2017; Kita, 2000; Kita & Özyürek, 2003), gestures should rather co-occur with fluent speech (as in Graziano & Gullberg, 2018). To disentangle whether gesture is at play in the speech formulation phase (*retrieval* of lexicon) or rather in the conceptualization phase, some previous studies have tested participants' performance in different tasks while **restraining** them from gesturing. Preventing speakers from gesturing should, in fact, result in less efficient speech production (e.g., inducing disfluencies) or to worse performance in linguistic tasks (e.g., picture naming, word recall from definition). However, the studies exploring the effects of gesture restriction on language performance (e.g., fluency) have yielded mixed results. For example, some detrimental effects of the inability to gesture on fluency have been found in a visual objects description task (Morsella & Krauss, 2004) and in the description of low codability abstract lines drawings (Graham & Heywood, 1975). By contrast, two studies which elicited speech by asking participants to describe motor tasks (Hoetjes et al., 2014; Hostetter et al., 2007) did not find effects of restraining gestures on fluency or speech length. In studies using story retellings, mixed effects were found on both fluency and speech content and planning (Finlayson et al., 2003; Jenkins et al., 2018; Rauscher et al., 1996). This issue will be further explored in the first study presented in Chapter 2.

On the other hand, the potential effects of encouraging participants to gestures on speech and fluency have not been directly investigated before. Yet, there is some evidence that tapping or producing beat gestures can facilitate lexical retrieval and word recall from definition (Lucero, Zaharchuk, & Casasanto, 2014; Ravizza, 2003) and that producing gestures in speech rehabilitation can enhance intelligibility of patients with acquired dysarthria (in terms of interword intervals, speaking time, total sentence duration, speaking rate, and phrasing strategy by triggering a more natural speech chunking) (Garcia, Cannito, & Dagenais, 2000; Garcia & Dagenais, 1998; Hustad & Garcia, 2005). Also, in educational settings, there is evidence that training children by asking them to produce beat gestures in a pre-test can help their narrative abilities and fluency in a post-test retelling task (Vilà-Giménez & Prieto, 2018); Llanes-Coromina et al. (2018) have shown that asking Catalan learners of English to produce beat gestures in a training phase, improves their reading in English in the post-test. The potential effects of encouraging the use of gestures on L1 speech fluency will be explored in the second study (Chapter 3).

1.4 Aim of the thesis

All in all, more empirical studies are needed to shed light on the role of gesture in the process of speech production. While the theoretical models briefly reviewed above provide some predictions for the potential effects of gesture restriction/encouragement on **fluency**, their scope is not to make direct predictions for the effects on **speech acoustics and underlying prosodic patterns**. For example, in the speech production process, it is not clear whether and how gesture production plays a role in the planning and formulation of prosodic patterns and in the final articulation phase. Therefore, further evidence on the possible effects of gesture restriction and encouragement on fluency and speech prosodic features would contribute to clarifying this issue.

The **goal** of the present dissertation is to investigate the effects of gesture restriction/encouragement in semi-spontaneous narrative speech. Crucially, evidence of these effects in semi-spontaneous speech production would help provide a better understanding of the speech production process in his full unfolding. The next two chapters will report on two twin investigations aimed at exploring how **restraining** and **encouraging** the use of gesture can impact on fluency and other prosodic features (Chapters 2 and 3). In addition, Chapter 4 will report on some qualitative and quantitative analyses carried out on the data collected for the two studies and will discuss some aspects of the methodological approaches used in the investigations. Finally, in Chapter 5, the results reported in this dissertation will be discussed in light of previous studies and hypotheses.

2 Exploring the Effects of Restraining the Use of Gestures on Speech

This chapter is based on the following paper:

Cravotta, A., Prieto, P., & M.G., Busà, (*under review*). Exploring the Effects of Restraining the Use of Gestures on Speech

Abstract

Purpose: Research in gesture studies has investigated whether the inability to gesture is detrimental to speech, as suggested by the main theoretical accounts. However, this research has yielded mixed results. To our knowledge, only one study analyzed whether restraining the use of gestures impacts on acoustic measures like pitch or intensity, but did not find effects (Hoetjes et al., 2014). Moreover, previous studies have used very controlled tasks and evidence is lacking about more spontaneous speech. Thus, further research is needed to assess the effects of gesture restriction on more naturalistic speech tasks that also takes into account a complete set of acoustic prosodic measures, including F0 and intensity.

Method: The present study investigates the effects of restraining hand gestures on narrative speech. Twenty native Italian speakers described the content of short comic strips to a listener in two conditions: Non-Restraining gestures (N); Restraining gestures (R) (i.e., the speakers had to sit on their hands). The following correlates of speech were examined: speech discourse length (number of words and discourse length in seconds), disfluencies (filled pauses, self-corrections, repetitions, insertions, interruptions) and prosodic properties related to speech rate, F0 and intensity.

Results: No evidence was found that the speakers' inability to gesture affects semi-spontaneous narrative speech in terms of discourse length, fluency and acoustic features.

Conclusion: This result expands Hoetjes et al.'s (2014) results and shows that speech does not become longer, more disfluent or monotonous when participants cannot gesture compared to when they can gesture. Further work is needed to shed more light on the direct influence of gesture on speech.

2.1 Introduction

Research in the last decades has investigated the self-directed role of gestures in the process of speech production. Theoretical models for speech-gesture production have proposed that gestures contribute to utterance planning and conceptualization (Gesture-for-conceptualization-hypothesis, Kita, Alibali, & Chu, 2017; Interface Model, Kita and Özyürek, 2003), facilitate lexical access (Krauss, Chen, & Gottesman, 2000), provide additional spatial information (de Ruiter, 2017), express the speakers' mental simulation of motor actions and perceptual states during speech production (Gestures as simulated action framework, Hostetter & Alibali, 2019), and reduce cognitive load (Cook, Yip, & Goldin-Meadow, 2012; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Ping & Goldin-Meadow, 2010). These effects are still under investigation from different perspectives and disciplines.

Gesture production has also been shown to be strongly interconnected with speech production at the **prosodic level**. Specifically, gestures and prosodic units are synchronized from a temporal point of view. For example, gestural strokes or prominent parts of gestures (or gesture 'hits') tend to align with prosodically prominent parts of speech, e.g. pitch accents (e.g., among many others, (Esteve-Gibert et al., 2017; Esteve-Gibert & Prieto, 2013; Loehr, 2012; Shattuck-Hufnagel & Ren, 2018). Also, recent recordings of concurrent speech and body movements (using electromagnetic articulometry for vocal tract movements and a motion capture system for body movements) have shown that final lengthening at prosodic boundaries extends to body movements, as manual gestures have been shown to lengthen during speech prominence and at boundaries (e.g., Krivokapić et al., 2017). In fact, there is evidence that language and action are closely related on a **motoric level** and co-speech gesture production engages brain areas that are functionally connected to Broca's area (Marstaller and Burianová, 2015; see Gentilucci & Dalla Volta, 2008 for a review). [For a general overview on the interaction between gestures and speech, see also Wagner et al. (2014)].

Investigating how people speak when they cannot use their hands (e.g., through gesture restriction during speech production, for example, by sitting on their hands or folding their arms) has been one of the methods used to test the predictions of some of the theoretical models mentioned in Section 1.3 and to explore the interrelation between gesture and speech production. However, in general, how the inability to gesture impacts semi-spontaneous speech is still unclear, and the effects of inhibiting gestures on acoustic features of speech (e.g., F0 and intensity) have not received much attention. In the present study, we aim to examine the impact of gesture restriction on narrative speech, with respect to fluency and speech length. The study also explores how this can impact acoustic features such as F0 and intensity.

Previous studies on restraining gestures

The potential effects of restraining gestures' use on speech have been assessed in relation to fluency, speech length, as well as speech content (i.e., semantic richness, spatial relations expression, imagery content) and, with the exception of one study (Hoetjes et al., 2014), the effects on acoustic features of speech such as F0 and intensity has been overlooked. One of the first studies to claim that restraining the use of gestures directly affects the expressiveness and richness of speech, as well as its fluency, was Dobrogaev (1929). This study is often reported in the literature (e.g., Krauss et al., 2000; McClave, 1998; Rauscher, Krauss, & Chen, 1996; Wagner et al., 2014), though it does not provide specific details about the methodology (e.g., participants, procedure, etc.), or any quantitative analysis. The participants were asked to speak while trying to avoid all possible body movements while talking (i.e., head, face, hands);

however, as observed by Dobrogaev, they were still showing rhythmic gestures and movements in different body parts (i.e., fingers, eyes, head). The main findings were the following: when speakers were asked to try to exclude all body movements (a) speech lost expressiveness and richness; and (b) speakers had difficulties with word retrieval resulting in short and disconnected sentences[2]. After this study, more recent empirical investigations have assessed the effects of restraining gestures on speech production. These are described below and summarized in Table 1. In general, the studies were aimed to test different hypotheses and thus used heterogeneous designs and methodology (e.g., between or within-subject designs; different types of task and different gesture inhibition methods). This makes it hard to compare the results and to draw a generalization.

With regard to the effects of restraining gestures on **fluency**, studies have addressed the issue by either focusing on connected speech (typically using very focused description tasks), or by directly testing lexical retrieval, a key component of successful fluent speech production (Hagoort & Indefrey, 2014; Indefrey, 2011; Kearney & Guenther, 2019), using tasks such as picture naming or word recall from definitions. As for the studies that examined fluency in connected speech, results are mixed. Morsella and Krauss (2004) investigated the effects of restraining hands use during an object description task. The study showed that the participants who were prevented from gesturing, in the description of both visible and absent objects, produced more disfluent speech. By contrast, in Graham and Heywood (1975), in which participants were asked to describe abstract lines drawings (with both high and low verbal

[2] We thank Mariia Pronina for providing a detailed summary of Dobrogaev (1929) that, to our knowledge, is only available in Russian.

codability), there was no difference between the restraining and non-restraining conditions on any of the measures of fluency (e.g., hesitations, filled pauses, etc.); however, the proportion of total speech time spent pausing was significantly higher in the restraining condition. Rauscher et al. (1996) found that preventing speakers from gesturing during oral descriptions of animated action cartoons increased the relative frequency of non-juncture filled pauses in speech with spatial content, while Finlayson et al. (2003), exploring the effects of hands' immobilization on a similar task (i.e., animated cartoon retelling), found that when gestures were restrained, speech was overall more disfluent in terms of pauses, repetitions and reformulations. More recently, Hostetter, Alibali, and Kita (2007) analyzed the spoken productions of participants that were asked to describe how to complete three motor tasks (e.g., wrapping a package), with half of them being prohibited from gesturing during the descriptions. The participants whose hands were restrained did not produce more filled pauses or a higher percentage of non-juncture filled pauses than the participants whose hands were not restrained. This was confirmed by Hoetjes et al., (2014), in which speakers had to give instructions on how to tie a tie, with half of the participants performing the task while sitting on their hands (other factors such as mutual visibility and previous experience were also tested). The study did not find effects of the inability to gesture on fluency (in terms of speech rate and filled pauses).

With reference to the effects of hand gesture restriction on **lexical retrieval tasks**, Frick-Horbury and Guttentag (1998) found that speakers were more likely to generate target words from definitions when they were free to gesture than when they were prevented from gesturing. In a **picture-naming task**, Pine, Bird and Kirk (2007) found that children named more words correctly and resolved successfully more Tip-of-the-Tongue states (TOTs)[3] when they were free to gesture vs when they were not. However, children did not experience more TOTs when restricted from gesturing, compared to when they were free to move. Beattie and Coughlan (1999) found that restraining gesture use does not affect **word recall from definition**: participants with their arms folded had, in general, a more fluent retrieval process than participants that were free to gesture. Precisely, the results showed that the participants who were prevented from gesturing experienced fewer TOTs than those who were free to gesture, and that free-to-gesture participants had to resolve proportionally more TOTs than participants in the folded-arms group.

A number of studies have investigated the potential effects of the inability to gesture on **speech length** and **speech content**, testing the idea that the inability to gesture can in principle affect the speaker's selection of information and the words and structure used to convey it (Church, Alibali, & Kelly, 2017; Kita et al., 2017). A study by Rimé et al. (1984) found that restraining speakers' gestures in 50-minute **spontaneous conversations** led to a lowered imagery level in the words used, as well as a reduction of content related to activity/movement (both measures were obtained via a dictionary-based computer program). By contrast, Walkington, Woods, Nathan, Chelule, & Wang (2019) found that gesture restriction does not impact language use in **math explanations**. Participants had to assess and explain whether 8 geometry conjectures were true or false and why. The transcriptions of the speech produced by the participants were analyzed through dictionary-based text analysis

[3] Type of problematic lexical accessing event, experienced as 'being sure that the information is in memory but (...) temporarily unable to access it' (Brown, 1991, p. 204).

tools (Coh-metrix, McNamara, Louwerse, Cai, & Graesser, (2013); LIWC, Pennebaker, Booth, Boyd, & Francis, 2015)). The results showed that there was no difference between the speech produced in the gesture-inhibited vs gesture-free trials in any of 148 different language measures assessed. These measures included kind and number of words used and other speech patterns. In a study using a **story retelling task** to test the effects of gesture inhibition on the speech produced, Finlayson et al. (2003) reported a higher number of words used and higher number of spatial content phrases in the speech produced in the gesture-restricted condition than in the hands-free condition. Similarly, an increase in the number of words used for expressing spatial content was found in Graham and Heywood (1975) in **abstract lines descriptions**. With regard to speech length as well as speech structure and content, Jenkins, Coppola and Coelho (2018) provide evidence that **narrative speech**, elicited via a story retelling task, does not change in terms of speech length or content (e.g., number of novel propositions, episode length) when gestures are restrained, but it is negatively affected by gesture restriction in terms of grammatical complexity (i.e., number of subordinated clauses in each narrative) and organization. Both Hostetter et al. (2007) and Hoetjes et al. (2014) did not find differences in the amount of speech produced (i.e., number of words) when gestures were restrained during **motor task descriptions**. However, Hostetter et al. (2007) found that speakers, if unable to gesture, produced less detailed (semantically rich) speech when describing spatio-motor events (e.g., putting one lace over the other vs. crossing the laces over one another); but no difference in the number of spatio-motor terms used was found. Emmorey and Casey (2001) explored whether gesture restriction affects spatial and motor expressions and found that in **giving commands to solve a spatial problem** (e.g., filling a puzzle grid with blocks), free-to-gesture speakers produced more verbal references to object orientation, while speakers that were prevented from gesturing were more likely to also lexically specify the direction of the rotation. Özer et al. (2017) found that, in the **descriptions of routes on a map**, gesture restriction did not impact the duration of routes descriptions in two groups of young vs elderly adults speakers; nonetheless, elderly adults produced more spatial information (i.e., street name, landmark or direction) when sitting on their hands than when free to gesture, whereas young adults expressed comparable spatial information in both conditions.

Table 1 provides a summary of the studies reviewed above, giving detailed information about their research design.

Table I Summary of previous findings on the effects of restraining gestures on speech (from more recent to older).

Study	**Fluency; Lexical retrieval performance**	**Content; Speech length**	**Design, task**
Walkington, Woods, Nathan, Chelule, Wang (2019)	-	No effects on language use in math justifications. Based on 148 speech measures (from text analysis software), including N of words per sentence, word concreteness.	**T**: true/false statements on 8 geometry conjectures + justifications; **L**: English **Add**: interviewer **S**: 108 (f, m) **R**: Hands in oven mitts attached to a table; **R Condition**: Within-subjects (WS)
Jenkins, Coppola & Coelho (2018)	-	Speech less grammatically complex and worse organized. No effects on speech length and content.	**T**: Retell story (pictures sequence) **L**: English **Add**: Unfamiliar listener. **S**: 10 (f, m) **R**: Gripping bottom of the seat. **R Condition**: WS
Özer,Tansan, Özer,Malykhina, Chatterjee & Göksun (2017)	-	Elderly speakers produced more spatial content in R.	**T:** Description of routes on a map. **L:** English. **Add:** participants **S**: 20 (young) + 19 (elderly), f, m. **R:** sitting on hands. **R Condition:** WS
Hoetjes, Krahmer & Swerts (2014)	No effects on fluency (nor F0 and intensity)	No effects on number of words	**T**: Description of motor task (i.e., tie a tie). **L**: Dutch **Add**: participant. **S**: 38 pairs (f, m): instruction givers + addressees. **R**: Sitting on hands. **R condition:** WS

Pine, Bird & Kirk (2007)	less TOTs* resolved =worst performance	-	**T**: Picture naming **L**: English **S**: 65 children (f, m). **R**: Gloves with Velcro on table. **R condition**: WS
Hostetter, Alibali, & Kita (2007)	No effects on fluency (e.g., filled pauses)	Less semantically rich verbs. No effects on number of spatial motor terms. No effect on amount of speech	**T:** Description of motor task (e.g., wrap a package) **L:** English **Add:** confederate **S:** 26 (f, m) **R:** Velcro cotton gloves on wooden board. **R condition:** Between-subjects (BS)
Morsella & Krauss (2004)	Speech more disfluent	-	**T:** Description of visual objects **L:** English **Add:** Offline **S:** 79 (f, m) **R:** Dummy electrodes on arms. **R condition:** BS
Finlayson, Forrest, Lickley & Beck (2003)	Speech more disfluent	More spatial content phrases Longer speech (number of words)	**T:** Retelling a cartoon (video) **L:** English **Add:** Participants **S:** 6 (f) **R:** Armchair with velcro strips on arms. **R condition:** WS
Emmorey & Casey, (2001)	-	Speakers more likely to lexically specify rotation direction	**T:** Give command to solve a spatial problem (filling a puzzle grid with blocks); **L:** English **Add:** Experimenter; **S:** 30 (15 f, 15m); **R:** sitting on hands; **R condition:** BS
Beattie & Coughlan (1999)	less TOTs* experienced = better performance	-	**T**: Retrieval of lexical items from definition. **L**: English. **S**: 60 (f, m) **R:** Folded arms. **R condition**: BS
Frick-Horbury & Guttentag (1998)	Fewer lexical items retrieved (no effect on number of TOTs and number of resolved TOTs). Different Restraining methods of EXP1 and 2 have similar results.	-	**T:** Retrieval of lexical items from definition**. L:** English. **Add:** - S: 36 (f, m) (EXP1) + 18 (f, m) (EXP2) **R:** Holding a rod (EXP1) and wearing an apron with Velcro (EXP2). **R condition:** BS
Rauscher, Krauss & Chen (1996)	Spatial content (only) more disfluent	-	**T:** Description of animated cartoon L: English **Add:** confederate. S: 41 (f, m) **R:** Dummy electrodes on hands. **R condition:** WS

Rimé, Schiaratura, Ghysselinckx & Hupet (1984)	Marginal reduction of speech rate	Decrease in imagery content and movement/ action content.	**T:** 50-minutes spontaneous conversation **L:** English. **Add:** Experimenter. **S:** 16 (m) **R:** Armchair devised to restrain movements (head, limbs). **R Condition:** WS
Graham & Heywood (1975)	Increased total speech time spent pausing In low-codability items descriptions=more hesitations, more pauses, more words.	More spatial content words and phrases Less use of demonstratives ('there', 'like this', 'like so')	T: Description of abstract lines drawings. **L:** English **Add:** Audience S: 6 (m). **R:** Folded arms. **R condition:** WS
Dobrogaev (1929)	Speech more disfluent (i.e., short, disconnected sentences); reduction of expressiveness.	Reduction of vocabulary size (i.e., richness).	(not mentioned)

*__TOTs:__ Tip-of-the-Tongue states; **T:** Task; **L:** Language; **Add:** Addressee; **S:** number of subjects; **f:** females; **m:** males; **R:** Restraining gestures' method; **R condition:** Restraining Gesture condition; **WS:** within subjects; **BS:** between subjects. This review excludes studies on restraining gestures where the speech produced was not directly analyzed.

Influence of gesture production on speech acoustics

While a number of studies have investigated the potential effects of the inability to gesture on fluency, speech length and speech content, the effects of restraining gestures on **acoustic features** of speech like F0 and intensity have not received much attention. Hoetjes et al. (2014) is, to our knowledge, the only one study that investigated whether speech becomes more monotonous (in terms of pitch range) when speakers cannot gesture. The study found that there were no effects of restraining gestures on the speakers' pitch range nor on any other acoustic measure (i.e., max, min and mean pitch and mean intensity). Also, the speech data was tested perceptually and showed that listeners were not able to tell, by hearing the speakers' voice only, whether someone was gesturing or not while speaking. However, Pouw, Paxton, Harrison & Dixon (2019) claim that arm movements of speakers can be predicted from their speech recordings. The study is based on the finding that arm movements affect voice acoustics of speakers producing steady vowels. The study showed that such effects on the acoustics allow listeners to reproduce speakers' arm movements in a synchronized manner just by listening to the recordings.

This line of research, exploring the tight relationship between gesture production and prosodic modulation, is related to the evidence that spoken language and arm gestures are controlled by the same **motor control system**, as suggested by both behavioral and neuroimaging studies (Marstaller and Burianová (2015); see Gentilucci & Dalla Volta (2008) for a review). For example, Bernardis and Gentilucci (2006) showed that when words were co-produced with meaningful/semantically related gestures, F0 and spectral properties of vowels were enhanced (F0 and F2 increase). Also, Krahmer and Swerts (2007) showed that producing a visual beat (head nod, eyebrow movement or hand gesture) on a given target word led to changes in the **acoustic realization of prominence** (in terms of vowel duration - longer durations - and spectral properties - lower F2, F3). They proposed that visual beats have a similar emphasizing function as pitch accents. Furthermore, in an experiment where speakers were asked to phonate while performing movements of different strengths, Pouw, Harrison, et al. (2019) showed that speakers' merely moving arms affected the acoustics of phonation at particular moments in time (i.e., a downbeat to upward movement phase of the beat gesture temporally aligned with a peak in amplitude envelope and a peak in F0). This provides evidence for a **biomechanical interdependence** between gestures and the acoustic realization of co-occurring speech. That is, hand gesture movements could affect the actions of the muscles involved in expiration, and this could directly affect prosodic metrics of speech.

Summary & research question

In sum, as shown in Table 1, previous studies have investigated the effects of restraining gestures on speech using different kinds of tasks. Some of these are focused and controlled tasks, such as the description of visual objects, the description of low codability abstract drawings or routes on a map. These tasks are purposefully designed to investigate speech production in relation to, for example, spatial memory or lexical retrieval. Other tasks, such as story retelling or the description of an animated cartoon are designed to elicit semi-spontaneous speech in more ecologically valid settings and involve more comprehensive speech planning and production mechanisms without challenging specific speech production processes. The different kinds of tasks may interact with the speakers' inability to gesture and yield different outcomes. For example, some detrimental effects of the inability to gesture on fluency have been found in visual objects description (Morsella & Krauss, 2004) and low codability abstract lines drawings descriptions (Graham & Heywood, 1975). By contrast, two studies which elicited speech by asking participants to describe motor tasks (Hoetjes et al., 2014; Hostetter et al., 2007) did not find effects of restraining gestures on fluency nor speech length. In studies using story retellings mixed effects have been found on both fluency and speech content and planning (Finlayson et al., 2003; Jenkins et al., 2018; Rauscher et al., 1996); in semi-spontaneous narrative speech the specific impact of the inability to gesture might be less strong and less evident. Nonetheless, such tasks can be useful to investigate the speech production process in a more ecologically valid setting.

Our study builds upon previous research investigating the effects of the inability to gesture by eliciting narrative speech as a way to investigate the speech production process in its full functioning. The study focuses on the acoustic features of speech and is particularly aimed to test whether the inability to gesture affects speech production with respect to fluency, speech length and speech prosody which, to our knowledge, has only been explored by Hoetjes et al. (2014).

2.2 Method

The present study used a **narration task** in which participants had to watch and describe a set of comic strips in two different conditions: **Non-restraining gesture condition (N)** in which speakers were free to gesture when narrating; and **Restraining gesture condition (R)** in which participants were asked to sit on their hands while telling the story. The experiment has a within-subject design (with a within subject factor: Condition) in order to control for the unavoidable presence of individual differences in gesture production (Briton & Hall, 1995; Chu, Meyer, Foulkes, & Kita, 2014; Goksun, Goldin-Meadow, Newcombe, & Shipley, 2013; Hostetter & Hopkins, 2002; Hostetter & Potthoff, 2012; Kita, 2009; Nicoladis, Nagpal, Marentette, & Hauer, 2018; O'Carroll, Nicoladis, & Smithson, 2015).

The analysis will focus on: speech discourse length (in terms of number of words and story length), fluency (in terms of number of filled pauses, self-corrections, repetitions, insertions, interruptions, silent pauses, and speech rate) as well as a set of speech features related to fundamental frequency and intensity.

Participants

Twenty female native speakers of Italian participated in the experiment. They were all from the Veneto region (age M = 24.1; SD = 5.5). Nineteen of them were undergraduate students at the University of Padova and 1 of them was former student from the same university. As compensation for their participation they were either given partial fulfillment of course credits or a free breakfast. Only female participants were recruited in the study for two main reasons, namely (a) to control for gender-related differences in F0 values; and (b) to control for potential gender differences in gesture production, as it might be the case that females are more expressive and produce more gestures when speaking than males (Briton & Hall, 1995; Hostetter & Hopkins, 2002).

Materials

Sixteen 4-scene comic strips adapted from Simon's Cat by Simon Tofield were used for the narration task (see Figure 2 for an example, and the Appendix for the complete set). The comic strips were carefully selected

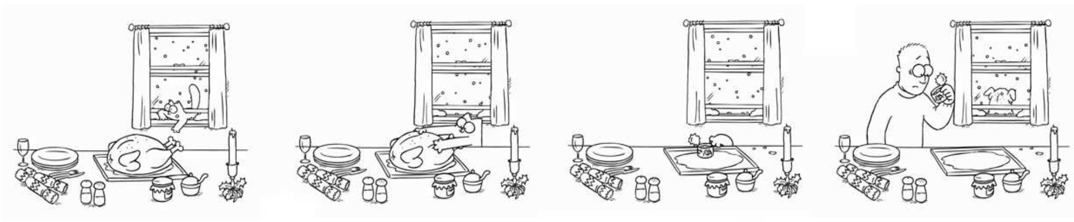


Figure 2 Example of a 4-scene comic strip used for the experiment (from Simon's Cat by Simon Tofield, reproduced with permission).

and adapted so that they were considered equivalent in terms of complexity and length (4-scene narration). Moreover, Simon's Cat comic strips do not contain text but feature a variety of characters and represent many motion events. Presumably, this property of the comic strips would make participants describe the events and spatial relations using gestures. To control for potential item effects, the target comic strips were shown in two orders of presentations which were counterbalanced across conditions (see next section).

Procedure

The participants were tested individually in a quiet room at the University of Padova. Each session was recorded with a HD video camera (JVC GZ-HD7E Everio) and speech was recorded (16 bit .wav files, 44.1kHz sampling rate) as a separate audio track using a MIPRO wireless headmounted microphone with a body-pack transmitter connected to a Zoom R16 digital audio mixer. All levels were set prior to the first participant and remained consistent throughout data collection.

The camera was set in front of the participant (at 2.50 m distance) recording her upper body and face. As shown in Figure 3, the participant sat on an office armchair and interacted with a listener (a confederate research assistant) that sat in front of her at a distance of 1.50 m. Distances were kept consistent across data collection. A second video camera was placed in front of the listener and recorded the listener's upper body and face during the whole session. The experimenter sat at the participant's side for the entire experiment.

Figure 3 Set-up: the image shows the speaker (left) and the addressee who is sitting in front of the speaker (right).

Each participant entered the room and was first given an informed consent form to sign. She was introduced to the listener as if he was also a fellow participant. Both the participant and the listener were given written instructions. The participant received the following **instructions** (translated from Italian): "You will be shown a set of short-sequence comic strips. A cat and its friends are the protagonists. Take your time to look at each of the short strips. When you think you understand the story they depict, the comic strip will be covered up. Then you will have to describe the story in sufficient detail so that your partner (who does not know the story) is able to reconstruct it by placing four comic cards that make up the strip in the correct order". The reason why we made participants believe the confederate addressee was a fellow participant who did not know the stories in advance was to avoid potential effects of common ground (Holler & Wilkin, 2009) as well as to give ecological validity to the narration task. In this way, the participants felt an obligation to explain the story clearly and fully because their "fellow participant" was dependent on them to understand it in order to finish the comprehension task. The confederate listener was instructed to provide basic backchannel and feedback cues to the speaker while listening to the stories (e.g., nodding when he felt it was natural to do so, while avoiding asking for clarifications and showing either amusement or boredom). In fact, it has been shown that gestures can be adapted depending on the addressee's feedback (e.g., lower gesture rate when addressees are less attentive (Jacobs & Garnham, 2007)). By contrast, to ensure that the interaction between participant and confederate was natural, he was allowed to interact more with the participant after the narration task and while solving his part of the task, i.e., when he was reconstructing the story.

Each participant had to retell a total of 16 stories. To make sure the written instructions were clear, the experiment started with a set of 2 initial familiarization trials to show the participant how the task should be performed and to make them confident with the camera. Specifically, each trial consisted of a **three-step sequence**: (1) the participant examined a four-scene comic strip to learn the story it depicted (for approximately a minimum of 5 seconds to a maximum of 40 seconds); (2) the comic strip was then concealed and the subject told the story to the confederate addressee in a face-to-face interaction; (3) the listener was then given four cards, each showing one scene of the comic, and had to reconstruct it by putting the four images in the correct order based on the speaker's story.

After the two-story familiarization phase, the participants had to tell the first half of the comic strips set (2 extra familiarization stories + 5 target trials) in the **Non-restraining gesture condition** (i.e., speakers were free to gesture while narrating; hence, **N condition**), and the second half (2 extra familiarization stories + 5 target trials) in the **Restraining gesture condition** (i.e., participants are asked to sit on their hands while narrating; hence, **R condition**). Asking speakers to sit on their hands (as in Emmorey & Casey, 2001; Hoetjes et al., 2014; Özer et al., 2017) rather than simply asking them not to gesture, was meant to limit the risk of imposing additional cognitive load resulting from the need of consciously remembering not to gesture.

The order of the two conditions was kept the same (N, R) for all participants: this is because we believed that telling participants to "come back" to a N condition after having restrained their gestures' use was not natural and it would lead to carryover effects between R and N. On the other hand, we are aware that this experimental set-up cannot exclude possible order effects due to the fact that the R condition is always produced after the N condition. For example, participants in the R condition could be more familiar with the task, more comfortable with the setting/the listener than in the N condition, with possible effects on their productions. However, the presence of two initial general familiarization trials plus other two familiarization trials before each condition excludes the argument that the N condition was not trained enough to be comparable with the R condition.

In the R condition participants were asked to sit on their hands while narrating via written instructions and an illustration of a person sitting on her hands. Even though the comic strips were carefully selected and adapted so that they were equivalent in terms of complexity and length, in order to avoid potential item effects half of the participants explained half of the comic strips in the N condition, while the same comic strips were explained in the R condition by the other half of the participants. By this we made sure that comic strip materials were counterbalanced across conditions.

The experiment lasted approximately 30 minutes. Audiovisual recordings of a total of 200 short narratives were obtained (20 participants × 10 target trials) lasting a total of 77.5 minutes (37.7 minutes in the N condition and 39.8 in the R condition).

Data Analysis: Transcriptions, fluency and acoustics

Speech discourse length

The recordings were edited so that a separate short audio file was created for each story told. Each audio file starts at the moment the participant starts telling the story until the moment the utterance ends (i.e., silences are excluded both at the beginning and at the end). A measure of audio file duration in seconds was included as a measure of speech discourse length (or story duration). The contents were manually transcribed and the word tokens per story were counted.

Fluency and disfluency measures

Fillmore, Kempler and Wang (2014) define fluency as "the ability to talk at length with few pauses, (…) to fill time with talk. A person who is fluent (…) does not have to stop many times to think of what to say next or how to phrase it" (p.93). In addition, according to Zellner (1994, p. 48) "people are disfluent if they often hesitate, make non-functional pauses and make speech errors and self-corrections." Thus, fluency can be measured not only by measures of speech rate (that gives a general idea of the efficiency of the speech production process) but also by the absence of a set of features that characterizes disfluency. In this study, we used a measure of speech rate which was automatically obtained using a Praat script (De Jong & Wempe, 2009). Specifically, the script detects potential syllable nuclei in terms of peaks in intensity (dB) that are preceded and followed by dips in intensity. It then divides the number of syllables produced in each audio file by the file's total duration in seconds (i.e., speech rate is given as number of syllables/s). Moreover, based on previous studies (Bergmann, Sprenger, & Schmid, 2015; Götz, 2013; Kormos, 2014, among others), instances of any of the following types of disfluencies were manually annotated by a single annotator (examples from our data are reported below in Italian and translated in English for convenience of the reader):

- **Repetitions**: of sounds (e.g., stuttering; "il pesce è di nuovo dentro l-l'acquario", "the fish is inside th-the acquarium again"), repetitions of words (e.g., "c'è un gruppo di di uccelli", "there is a group of of birds"); and repetitions of longer segments (e.g., "si toglie il collare e lo butta… e lo butta per terra", "he takes off his collar and throws it… and throws it on the ground");
- **Insertions**: of words or phrases when speech needs further qualification or detail (e.g., "degli uccellini stanno mangiando delle briciole di pane in un prato… – un bel po' di uccellini.", "Some birds are eating bread crumbs on the ground… – a lot of birds."; "il gatto nascosto sopra l'armadio (…) graffia l'uomo … sulla fronte l'uomo" "The cat hidden on top of the wardrobe (…) scratches the man … the forehead of the man");
- **Interruptions**: abrupt interruptions of a word, or pronunciation of an isolated incoherent sound (e.g., "il gatto mangia tut - entrambe le porzioni di cibo" , "the cat eats al- both portions of food"; "nella scena successiva tiene sollevato il topo davanti a lei per s per darglielo","in the next scene it holds the mouse in front of her to s to give it to her");

- **Self-corrections**: syntax-based (e.g., rephrasing); lexicon-based (a word is replaced with another word); phonology-based (slip of the tongue or unclear pronunciations);
- **Filled pauses:** (sounds like "ehm", "mmm") and prolongations of vowels (e.g., "alloraaa il gatto", "theeen, the cat");
- **Silent pauses:** annotated automatically by a Praat script (De Jong & Wempe, 2009).

The absolute count of all types of disfluencies was converted into a relative measure (e.g., number of filled pauses per 100 words).

Acoustic analysis

The acoustic analysis of speech was done using the **Praat** software (Boersma & Weenink, 2018). To explore whether fundamental frequency (F0) and intensity were modulated differently across the N and R conditions, a set of pitch and intensity measures were extracted with Praat for every audio file. The F0 data distributions were plotted and examined for each speaker individually; the distribution curves suggested that overall, modal voice register was centered between 100 and 500Hz. Previous literature has shown that for female speakers vocal fry register excursions fall in a low frequency F0 range that is generally below 100Hz (Hollien & Michel, 1968; McGlone, 1967; Murry, 1971) with a mean of approximately 50Hz (as reported in the literature review provided in Blomgren, Chen, Ng, & Gilbert, 1998). Thus, we decided to set F0 floor and ceiling to 100Hz and 500Hz respectively for all participants. Setting the floor to 100Hz allowed us to avoid vocal fry effects on the F0 measures. After setting F0 floor and ceiling, the F0 metrics were extracted for every audio file (story) via a publicly available Praat script by Jonas

Lindh[4]. The script extracts a pitch value every 10 ms of speech via autocorrelation algorithm for the whole audio file (story told). It then computes automatically: F0 mean, minimum, maximum, and standard deviation (the latter as a measure of pitch variability). As a second measure of pitch variability, Pitch Variation Quotient (PVQ) was also computed (Hincks, 2005). PVQ is a metric derived from the F0 standard deviation, which is expressed as a percentage of the mean (see Hincks 2005, who proposed this metric as a measure of perceived liveliness).

In the same way, intensity listings were extracted with an adapted version of the Praat script mentioned above which works similarly to the one used for extracting F0 metrics: loudness listings were extracted for every audio file and, subsequently, mean, minimum, maximum intensity, as well as standard deviation were computed.

Statistical analysis

The data analysis focused on a total of 19 variables of interest: (1) Story duration (in seconds), (2) Number of words per story, (3) Repetition rate, (4) Insertion rate, (5) Interruption rate, (6) Self-correction rate, (7) Filled pauses rate, (8) Silent Pauses rate, (9) Total Disfluencies rate (including 3, 4, 5, 6, 7), (10) Speech rate, (11) Minimum F0, (12) Maximum F0, (13) Mean F0, (14) F0 standard deviation, (15) Pitch Variation Quotient (PVQ), (16) Minimum intensity, (17) Maximum intensity, (18) Mean intensity, (19) Intensity standard deviation. Table 2 shows the main descriptive statistics of each of the 19 variables separated by Condition (N, R).

[4] https://github.com/YoeriNijs/PraatPitch

Table 2 Main descriptive statistics

	Mean		SD	
Variable	**N**	**R**	**N**	**R**
Story duration (s)	22.62	23.89	7.64	7.81
n. of words	63.76	66.03	20.78	23.25
Repetitions	1.68	1.68	1.75	1.87
Insertions	0.56	0.47	0.95	0.87
Interruptions	0.97	1.33	1.3	1.43
Self-corrections	1.23	1.32	1.61	1.58
Filled pauses	5.23	5.55	4.37	3.59
Silent pauses	4.8	5.32	3.51	3.71
Disfluencies (tot)	9.67	10.35	5.4	5.35
Speech rate (syll/dur)	4.42	4.34	0.58	0.53
F0 min (Hz)	105.44	106.48	10.95	11.36
F0 max (Hz)	383.64	389.98	82.87	72.7
F0 mean (Hz)	190.82	189.2	16.12	15.91
F0 var. (Hz)	33.44	32.75	11.43	11.2
PVQ	0.17	0.17	0.05	0.05
Intensity min (dB)	28.11	27.92	2.76	2.9
Intensity max (dB)	72.91	72.87	4.79	4.64
Intensity mean (dB)	60.12	59.89	4.03	4.24
Intensity var. (dB)	10.2	10.27	1.54	1.64

The effect of gesture restriction (within-subjects factor) on speech was tested by running a total of 19 Linear Mixed Effects Models (henceforth LMEMs, R function *lmer* in *lme4* package; Bates, Mächler, Bolker, & Walker, 2014). Each model included one of the 19 dependent variables listed in table and had *Condition* (N, R) as a fixed effect, and both *Story* and *Participant* as random intercepts. P-values are obtained by likelihood ratio tests of the full model against the model without the fixed effect of interest (i.e., Condition). The tests were then corrected for multiple testing via False Discovery Rate (i.e., Benjamini-Hochberg procedure, Benjamini & Hochberg, 1995). The results are reported with the adjusted False Discovery Rate (FDR) critical values.

2.3 Results

Table 3 shows the results of the 19 LMEMs. No significant effect of the ability to gesture on any of the dependent variables was found. In other words, stories told when participants were free to gesture were not longer in terms of duration (s) (est. = 1.277, S.E. = 0.667, Chisq = 3.645) and did not change in terms of *number of words* (est. = 2.270, S.E. = 2.096, Chisq = 1.1752); Also, none of the *disfluency rates* or *speech rate* (est. = -0.08, S.E. = 0.044, Chisq = 3.323) significantly changed between the two conditions. As for F0 and intensity, speech was not affected by gesture restriction.

Table 3 Results of the LMEMs per dependent variable

Variable	Estimates	S. E.	C.I.		t	Chisq	p	FDR
			Lower	Higher				
Story duration (s)	1.277	0.667	-0.034	2.589	1.914	3.645	0.056	0.0053
n. of words	2.270	2.096	-1.85	6.39	1.083	1.1752	0.278	0.0237
Repetitions	0.003	0.23	-0.449	0.455	0.015	2e-04	0.988	0.0500
Insertions	-0.09	0.122	-0.33	0.15	-0.735	0.5426	0.461	0.0342
Interruptions	0.362	0.186	-0.004	0.728	1.943	3.754	0.053	0.0026
Self-corrections	0.088	0.212	-0.33	0.506	0.415	0.173	0.678	0.0447
Filled pauses	0.325	0.397	-0.456	1.106	0.817	0.67	0.413	0.0289

Silent Pauses	0.525	0.365	-0.192	1.241	1.439	2.069	0.15	0.0158
Disfluencies (tot)	0.688	0.627	-0.544	1.92	1.098	1.208	0.272	0.0211
Speech rate (syll/dur)	-0.08	0.044	-0.165	0.006	-1.827	3.323	0.068	0.0132
F0 min (Hz)	1.035	1.4	-1.715	3.786	0.74	0.55	0.459	0.0316
F0 max (Hz)	6.340	9.402	-12.135	24.814	0.674	0.457	0.499	0.0395
F0 mean (Hz)	-1.619	0.852	-3.293	0.055	-1.9	3.594	0.058	0.0079
F0 var. (Hz)	-0.693	0.764	-2.195	0.809	-0.907	0.826	0.364	0.0263
PVQ	-0.002	0.004	-0.009	0.005	-0.502	0.253	0.615	0.0421
Intensity min (dB)	-0.19	0.102	-0.39	0.009	-1.874	3.499	0.061	0.0105
Intensity max (dB)	-0.037	0.346	-0.717	0.643	-0.106	0.011	0.915	0.0474
Intensity mean (dB)	-0.23	0.16	-0.545	0.085	-1.435	2.059	0.151	0.0184
Intensity var. (dB)	0.07	0.099	-0.124	0.264	0.711	0.508	0.476	0.0368

Note: N. of obs: 200; **Groups:** Participants, 20 | Story, 10. **C.I.**, Confidence interval: Lower 2,5%; Higher 97,5% (R package confint). **FDR:** False Discovery Rate adjusted alpha levels (Benjamini-Hochberg correction for multiple testing); Levels "N" (baseline) and "E" were recoded by contrasts (i.e., 0 was in between each level, instead of being equal to N).

Thus, the results show that speech length, fluency and prosodic features did not change when the speakers did not gesture compared to when they did gesture.

2.4 Discussion and Conclusions

Previous studies have shown that gesture restriction can in some cases affect fluency and speech content. However, previous findings have not always been consistent. Also, little attention has been paid to the potential effects of restraining gestures on speech acoustics specifically. The present study was aimed to gain further insight into the direct influence of the inability to gesture on speech prosody. An experiment was set up to elicit spontaneous story telling narratives in an ecologically valid setting, in which effort was made to let speakers be comfortable with the task and naturally interact with the listener. The study takes into account a comprehensive set of measures related to temporal narrative features, fluency measures, and also focuses on acoustic measures related to pitch and intensity. However, the results show no significant effects of restraining the use of gestures on speech. These results are commented separately for fluency, discourse length, and acoustic features.

As for **fluency**, we expected a lower paced and more disfluent speech produced in the restraining condition, since previous research showed that the inability to gesture can lead to lexical access difficulties or more general planning difficulties. However, our study does not provide evidence that gesture restriction has any detrimental effects on speech fluency in terms of speech rate or disfluencies rate. As for speech rate, a possible explanation could be that the stories produced were too short to allow speech rate measures to be representative of longer speech excerpts. However, this seems unlikely. The average length of the stories told in our experiment is between 22.62 seconds (in N condition) and 23.89 seconds (in R condition). Studies on the stabilization of speech rate in Brazilian Portuguese suggest that to give a representative idea of the speech rate of longer segments, speech excerpts should be at least around 9 seconds long (Arantes & Lima, 2017) or 12 seconds long (Arantes, Eriksson, & Lima, 2018).

It may be that, in the kind of task used for this study, which did not specifically challenge memory or lexicon, participants could compensate for the inability to gesture by drawing upon other cognitive strategies and capitalizing on analytical thinking. Moreover, even though gestures are restrained, we cannot exclude that any other movements participants did with their heads, legs and shoulders had some role in the process of speaking. In other words, the imagery and actions to be described can still be embodied/simulated by other body movements e.g., legs, head, lips and this could still play some positive role in the process of speaking.

Furthermore, to shed more light on the question of how the inability to gesture can affect speech fluency, a future investigation could be implemented with an analysis of disfluent speech in relation to, e.g., spatial and motor content expressions or an analysis of speech chunking in general. In fact, the inability to gesture can cause speech planning difficulties due to the fact that gestures have a role in packaging spatio-motor information into chunks ready to be expressed in speech (Information Packaging Hypothesis, Kita, 2000), and this may impact how speakers express spatial relations and action-related content (that is, spatial content expressions might be the most affected, in terms of fluency, by the inability to gesture, as found in Rauscher et al. (1996)).

As for **speech length measures**, we did not find any difference in story length or in the number of words used. On the one hand, when gestures are restrained, speakers may rely more on the speech modality, and speech might need to integrate information that cannot be expressed visually (Emmorey & Casey, 2001; Melinger & Levelt, 2004, but not confirmed in Hostetter et al., 2007). This would possibly require more words and longer speaking time. On the other hand, if the gesture stream is inhibited, this may lead speakers, for example, to just exclude some (spatial) information from their speech (Kita, 2000). To clarify this issue, future work can complement speech length measures with an analysis of the content differences between speech produced in N vs R conditions.

As far as the **acoustic analysis**, we explored whether restraining gesture has any effects on acoustic properties such as fundamental frequency and intensity. Our results do not provide evidence for it and are consistent with Hoetjes et al. (2014). However, empirical evidence coming from investigations on speech prosody and gesture production suggests that encouraging the use of gestures has an effect on speech acoustics, with reference to F0, intensity and spectral properties (Krahmer & Swerts, 2007; Pouw, Harrison, et al., 2019). Our results, at a first sight, appear to be in partial contradiction with the abovementioned studies. However, encouraging and restraining speakers' gestures should not be considered exact polar opposites: the two kinds of instructions (i.e., encouraging and restraining gestures) can still impose additional cognitive load (Hostetter & Alibali, 2019; Marstaller & Burianová, 2013) that can interfere differently with prosodic modulation, or they can even turn on speakers' preconceptions about how speech with or without gesture should sound; these might make speakers try to speak differently to adapt to these preconceptions (i.e., speakers might interpret the instruction to gesture as a request to enact more, or speak in a clearer way, louder, or perhaps slower, etc.).

As a last note, asking speakers to sit on their hands, as we did, does not necessarily restrain them from moving other parts of the body, e.g., the forearms, shoulders, head, and legs. Rimé et al. (1984), in fact, report that prohibiting hands movements can increase movements in other parts of the body, including the eyes, lips, fingers and legs (this is also observed by Hoetjes et al. (2014) and Dobrogaev (1929)). Our impression is that this applies to our data too: unsurprisingly, speakers still moved their shoulders and head while sitting on their hands and we cannot exclude that these movements could somehow replace actual hand gestures in their role for speech production. Walkington et al. (2019) recently found that when gestures are restrained and are not overtly produced, valid geometry conjectures justifications can still be successfully provided with no detrimental effects of gesture inhibition on speech production, and math reasoning. The authors propose that gestures are a mere *byproduct* of reasoning processes and do not cause any facilitation on it. By contrast, as Hoetjes et al. (2014) propose, it can be argued that even when people do not actually visibly produce a gesture or movement, this does not necessarily mean that they did not *intend* to produce a gesture (i.e., a motor command can be there even though it is not overtly produced as a gesture) and that speech and gestures are so closely related that not even by physically restraining speakers from using their hands can actually inhibit the effects/role that gestures have on speech production.

To conclude, while the present study provides evidence that gesture restriction does not affect discourse length, fluency or F0 and intensity variations in semi-spontaneous narrative speech, we believe that further investigations on semi-spontaneous speech could allow to shed more light into the mechanisms underlying gesture and speech production.

3 Encouraging the Use of Gestures: Effects on Speech

This chapter is based on the following publication:

Cravotta, A., Busà, M. G., & Prieto, P. (2019). Effects of Encouraging the Use of Gestures on Speech, In *Journal of Speech, Language, and Hearing Research, 62(9), 3204-3219*

Abstract

Purpose: Previous studies have investigated the effects of the inability to produce hand gestures on speakers' prosodic features of speech; however, the potential effects of encouraging speakers to gesture have received less attention, especially in naturalistic settings. The present study aims at investigating the effects of encouraging the production of hand gestures on the following speech correlates: speech discourse length (number of words and discourse length in seconds), disfluencies (filled pauses, self-corrections, repetitions, insertions, interruptions, speech rate), and prosodic properties (measures of F0, and intensity).

Method: 20 native Italian speakers took part in a narration task in which they had to describe the content of short comic strips to a confederate listener in one of the following two conditions: (1) Non-Encouraging condition (N), i.e., no instructions about gesturing were given; (2) Encouraging condition (E), i.e., the participants were instructed to gesture while telling the story.

Results: Instructing speakers to gesture led effectively to higher gesture rate and salience; Significant differences were found for (a) discourse length (e.g., the narratives had more words in E than in N); (b) acoustic measures: F0 maximum, maximum intensity, and mean intensity metrics were higher in E than in N.

Conclusion: The study shows that asking speakers to use their hands while describing a story can have an effect on narration length, and can also impact on F0 and intensity metrics. By showing that enhancing the gesture stream could affect speech prosody, this study provides further evidence that gestures and prosody interact in the process of speech production.

3.1 Introduction

In the last decades an increasing bulk of research has focused on co-speech gestures and their role in the process of speech production (see Church, Alibali, & Kelly, 2017). Scholars have proposed various theoretical models based on a set of experimental findings which predict the self-directed positive role of gestures during speech production. Gestures have been found to: contribute to utterance planning and conceptualization (Gesture-for-conceptualization-hypothesis, Kita, Alibali, & Chu, 2017; Kita & Özyürek, 2003); facilitate lexical access (Krauss et al., 2000); provide additional spatial information (Asymmetric Redundancy Sketch Model, de Ruiter, 2017); express the speaker's mental simulation of motor actions and perceptual states during speech production (Gesture as simulated action framework, Hostetter & Alibali, 2008, 2019) and reduce cognitive load (Cook, Yip, & Goldin-Meadow, 2012; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Ping & Goldin-Meadow, 2010). These effects are still under investigation from different perspectives and disciplines.

Gesture production is also very interconnected with speech production at the prosodic level. It has been shown that gestures and prosodic units are tightly synchronized from a temporal point of view. For example, gestural strokes or prominent parts of gestures (or gesture 'hits') tend to align with prosodically prominent parts of speech, or pitch accents (e.g., among others, (Esteve-Gibert et al., 2017; Esteve-Gibert & Prieto, 2013; Loehr, 2012; Shattuck-Hufnagel & Ren, 2018). Also, as shown by recordings of concurrent speech and body movements (using electromagnetic articulometry for vocal tract movements and a motion capture system for body movements), final lengthening at prosodic boundaries extends to body movements (i.e., manual gestures lengthen during speech prominence and at boundaries) (Krivokapić et al., 2017). [For a general overview on the synchronization between gestures and speech, see also Wagner et al. (2014)]. Finally both behavioral and neuroimaging studies have suggested that spoken language and arm gestures are controlled by the same motor control system (Bernardis & Gentilucci, 2006), and that both speech production and co-speech gesture production engage a neural network connected to Broca's area (Marstaller & Burianová, 2015; see Gentilucci & Dalla Volta, 2008 for a review).

Studies have shown that the production of gestures can have an impact on the prosodic features of cooccurring speech. There is some evidence that hand gestures can alter the prosodic and spectral properties of the speech they co-occur with. Krahmer & Swerts (2007) investigated whether producing a visual beat (head nod, eyebrow movement or hand gesture) on a given target word led to changes in the **acoustic realization of prominence**. They asked Dutch participants to utter a target sentence ("Amanda gaat naar Malta": *Amanda goes to Malta*) in a number of different ways by varying the distribution of the acoustic and/or visual cues for prominence. For example, a pitch accent could be produced on *Amanda*, *Malta*, or in neither of these words, and a visual beat (a manual beat gesture, a head nod, or an eyebrow movement) could be produced on *Amanda* or *Malta*. The vowels of the two target words (*Amanda* or *Malta*) were then analyzed in terms of duration, maximum fundamental frequency (F0), maximum values of higher formants (F1, F2, F3) and intensity (energy). The results showed that the production of beat gestures had an acoustic effect on the cooccurring word in terms of duration (longer durations) and spectral properties (lower F2, F3). This indicates that if a speaker produces a visual beat (either a manual beat gesture, a head nod, or an eyebrow movement) this triggers a clear direct effect on the acoustic realization of the co-occurring word, showing that visual beats have a similar emphasizing function as pitch accents. Bernardis & Gentilucci (2006) suggested that, in fact, gestures can enhance voice spectrum because the two modalities are coded as an integrated signal that is directed by a **unique communication system.** In their study, participants were asked to produce 3 common words (CIAO, NO, and STOP) and 1 pseudo-word (LAO) in four gesture conditions (absent, present, meaningful/related, non-meaningful gestures). They found that when the words were co-produced with meaningful gestures (i.e., semantically related with the co-occurring word), the F0 and spectral properties of the vowels were enhanced (specifically, F0 and F2 increased, while intensity and vowel duration did not increase). Interestingly, this effect was only present when the gestures were meaningful; that is, a random arm movement did not trigger comparable effects on the speech signal.

Experimental physiological evidence from Pouw, Harrison, & Dixon (2019) shows that the relation between gestures and the acoustic realization of co-occurring speech can also have a **biomechanical basis**. They claim that hand gesture movements can affect the actions of the muscles involved in expiration, and this could directly affect prosodic metrics of speech (e.g., contrasts in F0, as well as changes in amplitude). In their study they asked the participants to phonate a steady-state voiced vowel (e.g., 'a:') while either moving their arms (one-arm beat, two-arm beat, wrist beat) or not (passive condition) and either standing or sitting. They found that beat-like movements with high physical impetus (i.e., wrist movement excluded) affected phonation in terms of periodicity: a downbeat to upward movement phase of the beat seemed to temporally align with a peak in amplitude envelope and a peak in F0. Such peaks were observed about 50 ms before and 50 ms after the moment of maximum extension (i.e., when the hand reached its lowest point) in correspondence with the highest impetus (deceleration for stopping extension and acceleration for initiating flexion). Also, they found that performing movements when standing, as compared to sitting, increased the degree of entrainment of movement and phonation (but only for F0), due to anticipatory postural adjustments. In other words, the study showed that merely moving arms affects the acoustics of phonation at particular moments in time. This might support the idea that the gesture-speech synchrony itself emerges from biomechanical constraints and not only from neural mechanisms.

One of the methods used to investigate the potential effects of gestures on speech has been to **restrain speakers' gestures** during speech production. The effects of the inability to gesture on speech production have been mainly assessed in relation to speech content (i.e., semantic richness, spatial relations expression, imagery content), speech length, as well as fluency. Hostetter, Alibali, & Kita (2007) found that when speakers were prevented from gesturing the speech used to describe motor tasks (e.g., *how to wrap a package*) was less semantically rich in expressing spatio-motor events. Also, spontaneous conversations produced by speakers that were prevented from gesturing showed a general decrease in imagery content (Rimé et al., 1984). As for speech length, studies using different tasks (e.g., cartoon retelling, description of drawings or description of motor tasks) reported a general increase in the number of words used in the speech produced by speakers who were prevented from gesturing (Finlayson et al., 2003) or at least in the number of words used for expressing spatial content (Graham & Heywood, 1975) – though this has not been corroborated in more recent studies (e.g., Hoetjes et al., 2014; Hostetter et al., 2007; Jenkins et al., 2018). Also, speakers were found to be more likely to generate target words from definitions or pictures (i.e., picture naming task) when free to gesture than when prevented from gesturing (Frick-Horbury & Guttentag, 1998; Pine, Bird, & Kirk, 2007). With regard to speech fluency, studies have shown that speech becomes less fluent when speakers are prevented from gesturing while retelling stories or describing drawings or objects (e.g., Dobrogaev, 1929;

Finlayson, Forrest, Lickley, & Beck, 2003; Graham & Heywood, 1975; Morsella & Krauss, 2004), and especially when describing spatial content (Rauscher et al., 1996). However, not all studies have confirmed these findings: in Hoetjes et al., (2014) speakers had to give instructions on how to tie a tie, with half of the participants having to perform the task while sitting on their hands (other factors such as mutual visibility and previous experience were also tested). The study did not find effects of the inability to gesture on fluency (in terms of speech rate and filled pauses). Hoetjes et al. (2014) is also, to our knowledge, the only study that has explored the effects of restraining gestures on speech acoustics by investigating whether speech becomes more monotonous (in terms of pitch range) when speakers cannot gesture. The study found no evidence of the effects of restraining gestures on pitch range; the speech data was also tested perceptually and it showed that listeners were not able to tell, by hearing speakers' voice only, whether someone was gesturing or not while speaking. Hoetjes et al. (2014)'s idea had been previously proposed by Dobrogaev (1929), an early study which is often reported in the literature (e.g., Krauss et al., 2000; McClave, 1998; Rauscher, Krauss, & Chen, 1996; Wagner, Malisz, & Kopp, 2014) but does not contain any quantitative analysis, nor does it provide specific details on the methodology used.

Previous studies on encouraging gestures

A complementary way of exploring the role of gestures in speech production is by encouraging speakers' use of gestures. Though encouraging and restraining gesture use are not exactly polar opposites, the two types of instructions can in principle provide complementary evidence on the effects that gestures have on speech production and, more generally, can both contribute to shedding light on the cognitive and functional roles that gestures have on speaking. Nonetheless, the effects of actively eliciting gestures on speech production and its prosodic properties have been investigated to a lesser extent. As for **speech length**, Parrill, Cabot, Kent, Chen & Payneau (2016) found that participants tested in a narrative task talked slightly longer in the instructed-to-gesture phase than in the no-instruction phase (i.e., they found a significant difference in mean story duration for the narrative data in the two conditions). With respect to abilities ideally related to **fluency**, Ravizza (2003) showed that asking participants to produce meaningless hand movements (e.g., rhythmic tapping) facilitated speech production in terms of resolution of tip-of-the tongue states and **lexical retrieval**. Lucero, Zaharchuk, & Casasanto (2014) also found that asking speakers to perform non-referential gestures (i.e., beats) had a positive effect on word production. In this study, subjects were asked to recall words from definitions while they (a) had to perform either iconic gestures or (b) beat gestures; or (c) had no instructions about gesturing. Their results showed that beat gestures facilitated word production, since Reaction Times (RTs) for successfully recalled words were shorter in the beat gesture than in the iconic gesture condition and in the no-instruction condition (the longest RTs were found in the iconic gesture condition).

Evidence from applied research in **educational** and **speech rehabilitation settings** also shows that gestures can help boost speech fluency and speech articulation. Vilà-Giménez & Prieto (2018) is, to our knowledge, the only study that has explored the effects of encouraging gesture production on narrative ability and fluency. The study showed that training children by asking them to produce beat gestures (or non-referential hand gestures that associate with speech) during story-telling improved children's narrative performance and fluency in the post-test phase, when no instructions about gestures were given. Also, Vilà-Giménez, Igualada, & Prieto (2019) showed that observing storytellers who use beat gestures while telling a story improves the performance of children's narrative abilities in story retelling. Moreover, Llanes-Coromina, Prieto, & Rohrer (2018) showed that asking Catalan speakers to produce beat gestures while reading in L2 in a training phase can benefit fluency in a posttest reading task, when no instructions to gesture are provided. These studies together point towards a positive effect of encouraging the use of gestures on speech fluency, though the speech that was evaluated in terms of fluency was not produced specifically in the encouraging gesture phase, but in a post-test phase. In **speech rehabilitation** it has been shown that instructing adults with acquired dysarthria to produce hand gestures while speaking can enhance their speech intelligibility by causing an improvement of some aspects of the sentences uttered (in terms of inter-word intervals, speaking time, total sentence duration, speaking rate and phrasing strategy by triggering a more natural speech chunking) (Garcia & Cobb, 2000; Garcia & Dagenais, 1998; Garcia, Cannito, & Dagenais, 2000; Hustad & Garcia, 2005).

Summary & research question

To sum up, both encouraging and restraining the use of gesture appear to be valid ways to explore the role of gesture on speech production but findings need further investigation especially in naturalistic settings. The aim of the present study is to provide evidence, based on naturalistic data, of the potential effects of encouraging gesture use on a comprehensive set of speech cues related to prosody. The object of analysis is features that have been previously investigated in relation to hand gesture restriction and only marginally in relation to hand gesture encouragement. These are: speech discourse length, fluency (number of filled pauses, self-corrections, repetitions, insertions, interruptions, silent pauses, speech rate), as well as fundamental frequency and intensity. Considering previous findings, we expect that encouraging the use of hand gestures benefits speech fluency as well as has an impact on F0 and intensity.

3.2 Method

The present study used a narration task in which the participants had to watch and describe a set of comic strips in two different conditions: **Non-encouraging gesture (N)**, in which no instructions regarding how to gesture while narrating were provided; and **Encouraging gesture** condition **(E)**, in which participants were encouraged to use gestures while telling the story. The experiment has a within-subject design (with a within subject factor: *Condition*) in order to control for the unavoidable presence of individual differences in gestures' use in terms of types, frequency and saliency.

Participants

Twenty female native speakers of Italian participated in the experiment[5]. They were all from the Veneto region (age M = 24.2; SD = 2.9). Seventeen of them were undergraduate students at the University of Padua and 3 of them were former students from the same university. As compensation for their participation they were either given partial fulfillment of course credits or a free breakfast. Only female participants were recruited in the study for two main reasons, namely (a) to control for gender-related differences in F0 values; and (b) to control for potential gender differences in gesture production, as it might be the case that females are more expressive and produce more gestures when speaking than males (e.g., Briton & Hall, 1995; Hostetter & Hopkins, 2002).

Materials

For the narration task, we used the same kind of stories described in Section 2.2, and reported in Appendix. Similarly to Study 1, to control for

[5] Study 2 involved a different group of participants than Study 1.

potential item effects, the target comic strips were shown in two orders of presentations which were counterbalanced across conditions (see next section).

Procedure

Study 2 followed the same procedure as Study 1 (see Section 2.2) with the same research assistant participating as a listener. Similarly, after the two-story familiarization phase, the participants had to tell the first half of the comic strips set (2 extra familiarization stories + 5 target trials) in the **Non-encouraging gesture condition** (i.e., no instructions regarding how to gesture while narrating were provided; hence, **N condition**), and the second half (2 extra familiarization stories + 5 target trials) in the **Encouraging gesture condition** (i.e., the participants were encouraged to use gestures while telling the story; hence, **E condition**). The experiment has a within-subject design (with *condition* as within-subject factor). The order of the two conditions was kept the same (N, E) for all participants: this is because we believed that telling participants to "come back" to a N condition after having encouraged them to gesture was not natural and it would lead to carryover effects between E and N. On the other hand, we are aware that this experimental set-up cannot exclude possible order effects due to the fact that the E condition is always produced after the N condition. For example, participants in the E condition could be more familiar with the task, more comfortable with the setting/the listener than in the N condition, with possible effects on their productions. However, the presence of two initial general familiarization trials plus other two familiarization trials before each condition excludes the argument that the N condition was not trained enough to be comparable with the E condition.

In the E condition the participants were given the following **instructions** (translated from Italian): "Tell each story and use hand gestures to help you do so”. The instructions were kept visible during the whole E condition to remind the participants about the task. In order to avoid potential item effects half of the participants explained half of the comic strips in the N condition, while the same comic strips were explained in the E condition by the other half of the participants. By this we made sure that comic strip materials were counterbalanced across conditions.

The experiment lasted approximately 30 minutes. Audiovisual recordings of a total of 200 short narratives were obtained (20 participants × 10 target trials) lasting a total of 81.2 minutes (39.1 minutes in the N condition and 42.1 in the E condition).

Assessment of the effects of the encouraging gesture prompt

The amount of gesturing that was present in the speakers’ narrations across the two conditions (N, E) was quantitatively assessed over the entire data set. This was done to ascertain that the gesture elicitation instruction in the E condition had actually caused an increase in the speakers’ gestures with respect to the N condition. All instances of co-speech gestures were identified and manually annotated by a single annotator with the software ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). The annotation criteria consisted in the marking of all gestural *strokes* (the most effortful part of the gesture that usually constitutes its semantic unit, e.g., two hands shaping together a rounded table (Kendon, 2004; McNeill, 1992). Non-gestural movements (self-adaptors e.g., scratching, touching one’s hair; Ekman & Friesen, 1969) were excluded. The speakers produced a total of 2396 gestures (out of which, 1015 in the N and 1381 in the E condition). Gesture rate was calculated per every story told as the number of gestures produced per story relative to the number of spoken words in the narration (Gestures/words*100).

In addition, to assess whether instructing speakers to gesture also changed gesture **salience**, each stroke was further classified depending on where it was performed in the gesture space. The classification was done based on McNeill (1992)'s representation of the gesture space, which is divided into sectors delimited by concentric squares. For the present annotation, a 2-sectors version of McNeill (1992)'s representation of the gesture space was used, as illustrated in Figure 4:

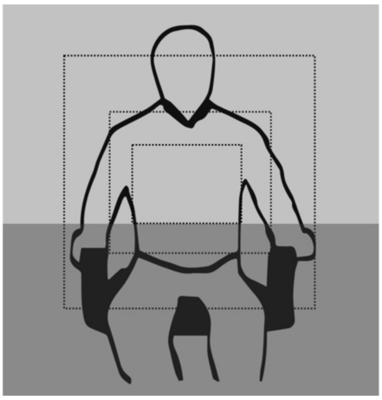


Figure 4 Gesture space representation (based on McNeill, 1992). Upper (light grey) area: salient gestures; Lower (dark grey) area: non-salient gestures.

When the gesture stroke was produced in a more central, higher and visually prominent area (Streeck, 1994) of the gesture space (the lighter grey area), the gesture was coded as *salient* (Figure 5, left frame); whereas, when the gesture stroke was produced in a less visually prominent area

Figure 5 Examples of a salient gesture (left) and non-salient gesture (right). Both gestures referred to a "tuna can" appearing in one of the stories told in the study.

(the lower darker sector), it was coded as *non-salient* (e.g, those gestures performed while keeping the arms along the legs or on the armrests (Figure 5, right frame). Salient Gesture (S) rate was computed per every story told as the number of salient gestures produced per story relative to the number of spoken words in the narrative (Salient gesture/words*100). The same was done for Non-Salient (NS) gesture rate. The prompt to gesture worked well, as shown in Figure 6 and Table 4.

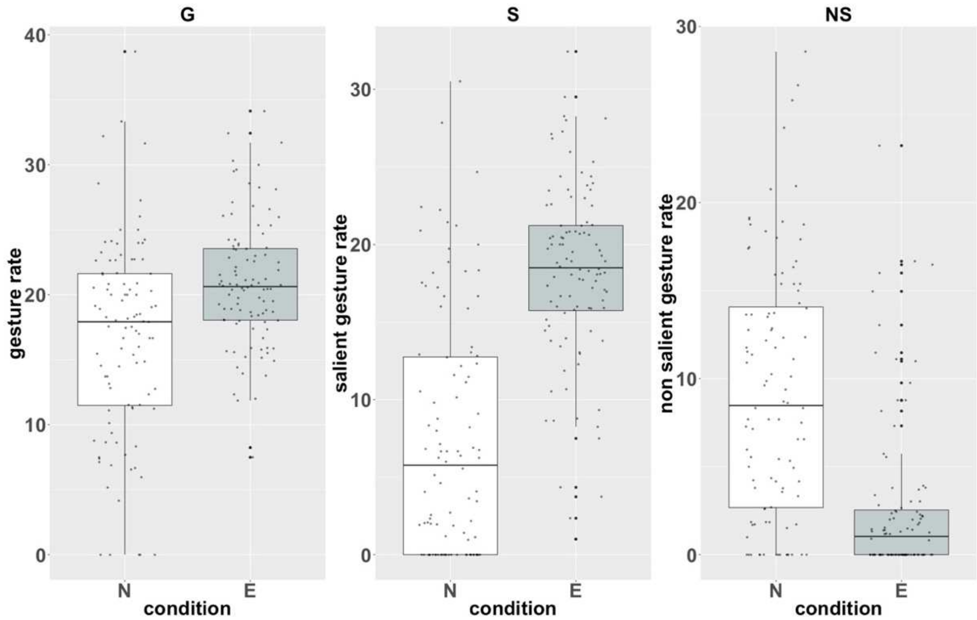


Figure 6 Gesture rate (G) and Salient gesture rate (S) and Non-Salient gesture rate (NS) per condition (nonencouraging , N; encouraging, E).

The effects of encouraging gestures on gesture rate and gesture salience was tested by running 3 Linear Mixed Effects Models (henceforth, LMEMs; R function *lmer* in *lme4* package; Bates et al., 2014). Each model included one of the 3 dependent variables listed in Table 4 and had condition (N, E) as a fixed effect and both story and participant as random intercepts. P-values are obtained by likelihood ratio tests of the full model against the model without the fixed effect of interest (i.e., condition). *Gesture rate* (est. = 4.134, SE = 0.708, $p < .001$) and *Salient gesture rate* (est. = 10.723, SE = 0.794, $p < .001$) were significantly higher in the E condition than in the N condition, and conversely *Non-Salient gesture rate* (est. = -6.589, SE = 0.65, $p < .001$) was higher in the N condition.

Table 4 Linear mixed-effects models for the effects of condition on Gesture rate, Salient gesture rate and Non-Salient gesture rate (per 100 words).

Variable	**Estimates**	**S. E.**	**C.I**		**t**	**Chisq**	**p**
			Lower	**Higher**			
Gesture rate	4.134	0.708	2.742	5.526	5.838	31.217	<.001
Salient gesture rate	10.723	0.794	9.162	12.283	13.51	125.57	<.001
Non-Salient gesture rate	-6.589	0.65	-7.868	-5.311	-10.13	80.71	<.001

Note: Models: R function lmer in lme4 package (Bates, Mächler, Bolker, & Walker, 2014). Each model included Condition (N, E) as a fixed effect and both Story and Participant as random intercepts. **N. of obs:** 200; Groups: participants, 20 | Story, 10. C.I.: Lower 2,5%; Higher 97,5% (R package confint). Levels "N" (baseline) and "E" were recoded by contrasts. P-values obtained by likelihood ratio tests of the full model against the model without the fixed effect of interest (i.e., Condition).

To sum up, the quantitative analysis of the gestures performed by the speakers suggests that encouraging speakers to use their hands while telling the stories worked well (as found in previous studies, Chu & Kita, 2011; Cook, Yip, & Goldin-Meadow, 2010), leading speakers to use more gestures that also involve a higher gesture space.

Data analysis: Transcriptions, fluency and acoustics

In this study, the data analysis was performed exactly as in Study 1 on Restraining gestures (see Section 2.2). Similarly, the following measures were taken in each story told in the two conditions (N, E): (1) speech discourse length measures; (2) fluency and disfluency measures; (3) acoustic measures.

Statistical analysis

The data analysis focused on a total of 19 variables of interest: (1) Story duration (in seconds); (2) Number of words per story; (3) Repetition rate;

(4) Insertion rate; (5) Interruption rate; (6) Self-correction rate; (7) Filled pauses rate; (8) Silent Pauses rate; (9) Total Disfluencies rate (including 3, 4, 5, 6, 7); (10) Speech rate; (11) Minimum F0; (12) Maximum F0; (13) Mean F0; (14) F0 standard deviation; (15) Pitch Variation Quotient (PVQ); (16) Minimum intensity; (17) Maximum intensity; (18) Mean intensity; (19) Intensity standard deviation. The main descriptive statistics per variable are shown in Table 5:

Table 5 Main descriptive statistics

	Mean		SD	
Variable	**N**	**E**	**N**	**E**
Story duration (s)	23.44	25.23	8.17	10.25
N. of words	62.1	67.19	21.35	24.74
Repetitions	1.47	1.95	1.85	2.18
Insertions	0.54	0.41	1.16	0.75
Interruptions	0.86	1.02	1.45	1.31
Self-corrections	1.2	1.45	1.49	1.58
Filled pauses	6.7	6.25	4,12	3.89
Silent pauses	7.61	7.76	4.69	4.27
Disfluencies (tot)	10.78	11.08	6.23	5.97
Speech rate (syll/dur)	4.24	4.21	0.53	0.62
F0 min (Hz)	105.44	105.25	9.92	11.08
F0 max (Hz)	391.86	414.43	66.36	68.86
F0 mean (Hz)	191.49	191.94	18,61	17.57
F0 var. (Hz)	33.88	35.03	9.87	10.36
PVQ	0.18	0.18	0.05	0.05
Intensity min (dB)	27.82	27.84	2.57	2.46
Intensity max (dB)	73.36	74.61	4.79	4.69
Intensity mean (dB)	60.51	61.36	3.51	3.17

Intensity var. (dB)	10.84	10.86	1.39	1.31

The effect of gesture encouragement (within-subjects factor) on speech was tested by running a total of 19 LMEMs (R function *lmer* in *lme4* package; see Bates et al., 2014). Each model included one of the 19 dependent variables listed in Table 5 and had *Condition* (N, E) as a fixed effect and both *Story* and *Participant* as random intercepts. P-values are obtained by likelihood ratio test of the full model against the model without the fixed effect of interest (i.e., *Condition*). The tests were then corrected for multiple testing via False Discovery Rate (i.e., Benjamini-Hochberg procedure, Benjamini & Hochberg, 1995). The results are reported with the adjusted False Discovery Rate (FDR) critical values.

3.3 Results

Table 6 shows the results of the 19 LMEMs. There was an effect of condition on speech discourse length, specifically on *Number of Words* that was higher in the E than in the N condition (est. = 5.09, S.E = 1.945, $p < .011$, FDR). As for the measures of fluency, there was no effect of *Condition* in any of the disfluency measures, nor there was any effect on *Speech Rate*. Finally, as for the acoustic measures related to F0 and intensity, there was an effect of condition for *Maximum F0* that was higher in E than in N (est. = 22.578, S.E=8.302, $p < .008$, FDR); and an effect was found for *Maximum Intensity* (est. = 1.254, S.E = 0.297, $p < .005$, FDR) and *Mean Intensity* (est. = 0.846, S.E = 0.15 $p < .003$, FDR), which were both higher in E than in N. While the results on the Maximum F0 are quite impactful (a mean difference of 22 Hz), a more subtle effect was found for intensity. However, we should note that a change of about 1 dB corresponds to the smallest change in loudness that can be heard in a sound booth (e.g. a change of 5 dB corresponds to doubling the loudness (Ladefoged, 2003)).

Table 6 Results of the LMEMs per dependent variable

Variable	Estimates	S. E.	C.I		t	Chisq	p	FDR
			Lower	Higher				
Story duration (s)	1.792	0.745	0.329	3.256	2.407	5.73	0.017	0.013
N. of words	5.090	1.945	1.268	8.913	2.617	6.753	0.009	0.011 *
Repetitions	0.477	0.231	0.023	0.931	2.066	4.237	0.04	0.016
Insertions	-0.132	0.133	-0.393	0.128	-0.997	0.996	0.318	0.029
Interruptions	0.16	0.182	-0.197	0.518	0.881	0.779	0.377	0.032
Self-corrections	0.242	0.197	-0.146	0.63	1.228	1.509	0.219	0.024
Filled pauses	-0.45	0.416	-1.267	0.368	-1.081	1.17	0.279	0.026
Silent pauses	0.145	0.42	-0.682	0.971	0.345	0.119	0.73	0.042
Disfluencies (tot)	0.298	0.655	-0.99	1.586	0.455	0.208	0.649	0.039
Speech rate (syll/dur)	-0.029	0.048	-0.123	0.065	-0.615	0.38	0.538	0.037
F0 min (Hz)	-0.191	1.247	-2.641	2.259	-0.153	0.024	0.878	0.050
F0 max (Hz)	22.578	8.302	6.251	38.904	2.719	7.272	0.007	0.008 *
F0 mean (Hz)	0.458	0.663	-0.846	1.761	0.69	0.478	0.489	0.034
F0 var.(Hz)	1.148	0.715	-0.257	2.554	1.606	2.574	0.109	0.018
PVQ	0.005	0.004	-0.002	0.013	1.443	2.08	0.15	0.021
Intensity min (dB)	0.0198	0.078	-0.133	0.173	0.254	0.065	0.799	0.045
Intensity max (dB)	1.254	0.297	0.67	1.837	4.221	17.043	0.001	0.005 *

Intensity mean (dB)	0.846	0.15	0.552	1.14	5.652	29.459	0.001	0.003 *
Intensity var. (dB)	0.024	0.097	-0.167	0.215	0.245	0.06	0.806	0.047

Note: N. of observations: 200; **Groups:** participants, 20 | story, 10. Confidence interval (CI): lower, 2.5%; higher, 97.5% (R package confint). **FDR:** False discovery rate adjusted alpha levels (Benjamini–Hochberg correction for multiple testing); An asterisk denotes significance after Benjamini–Hochberg correction; Levels "N" (baseline) and "E" were recoded by contrasts (i.e., 0 was in between each level, instead of being equal to N). **syll/dur** = number of syllables produced in each audio file divided by the file's duration (De Jong & Wempe, 2009); **F0 =** fundamental frequency. **PVQ =** Pitch Variation Quotient

In Figure 7 the 4 jittered boxplots refer to the 4 variables that showed a significant effect of *Condition*:

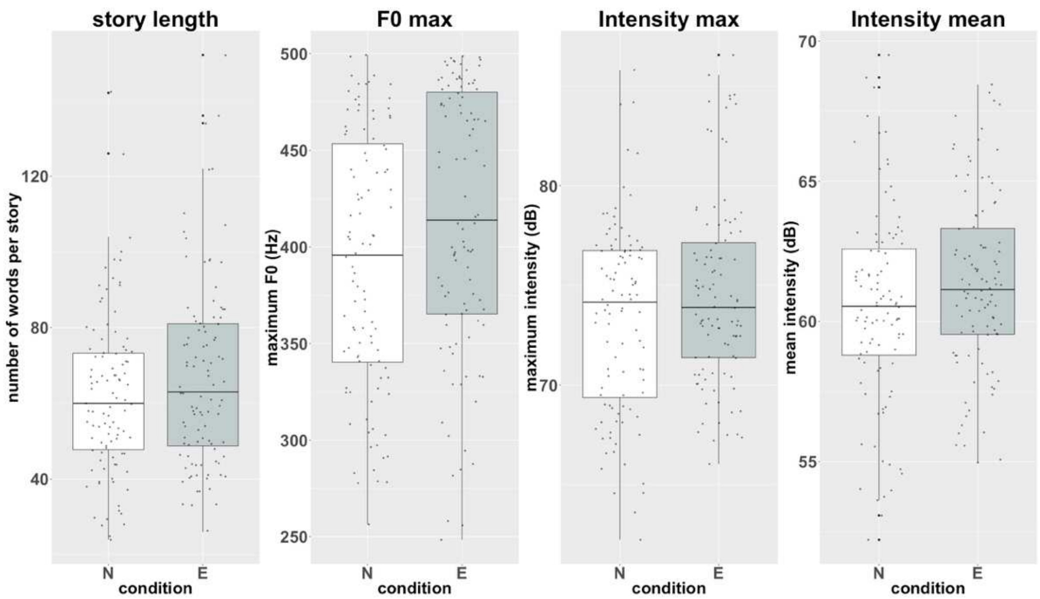


Figure 7 Box plots representing the variables that showed a significant effect of Condition (N, E). They show the full range of variation (form minimum to maximum), the likely range of variation (the interquartile range), and the typical value (the median) across condition.

3.4 Discussion and Conclusion

The aim of the present study was to explore the prosodic effects of encouraging the use of hand gestures on naturally elicited spontaneous narrative speech. The study took into account a comprehensive set of measures relating to discourse length and fluency, and acoustic measures relating to pitch and intensity.

Previous studies on the role of gestures in speech fluency and prosody have used a variety of methods and heterogeneous experimental settings and tasks (e.g., lexical retrieval via definition, objects, drawings or motor task descriptions, story retellings, vocalizations in the lab). These studies have suggested that gestures can, in some cases, have a facilitating role in the process of speech production. However, little attention has been paid to the effects that encouraging gestures can have on both fluency and prosodic features, especially in naturalistic settings.

The present experiment was set up to elicit spontaneous story telling narratives in an ecologically valid setting, and effort was made to let speakers be comfortable with the task and naturally interact with the listener. The results showed that encouraging speakers to gesture, while effectively increasing **gesture rate** and **gesture salience**, related to some modest changes in acoustic-prosodic features of speech: (1) production of **longer discourse** in terms of number of words used; (2) **higher fundamental frequency maxima**; and (3) **louder speech**. Let us now comment on each of these results separately.

First, when instructed to gesture, speakers used more words to tell each story. The increase in the number of words in the E condition, albeit significant, is rather modest (approximately 5 words) and it should be interpreted with care. In spite of this, this result is coherent with Parrill et al. (2016)'s findings, as well as with the positive effects that encouraging gestures have on lexical access (e.g., Lucero et al., 2014; Ravizza, 2003). We believe that a more thorough content analysis of the narratives could contribute to shedding light on the following issues: Does encouraging the use of gestures enhance the semantic density and richness of speech? Does it affect prosodic planning and speech chunking? Also, a more thorough investigation would allow a comparison of our results with those from studies investigating the effects of restraining gestures on speech production, which have yielded mixed results (e.g., Finlayson et al., 2003; Graham & Heywood, 1975; Hoetjes et al., 2014; Hostetter et al., 2007; Jenkins et al., 2018). This would be needed to reconcile the overall findings with the predictions made by the major speech-gesture production frameworks (Hostetter & Alibali, 2019; Kita et al., 2017; Kita & Özyürek, 2003).

Second, we did not find evidence of an effect of encouraging gestures on **fluency**. There was no difference between the two conditions in terms of disfluency rates and speech rate. We initially expected that encouraging gestures would have a beneficial effect on speech fluency, as found in studies that encouraged speakers to gesture in a training phase (Vilà-Giménez & Prieto, 2018; Llanes-Coromina et al., 2018). However, the direct influence of encouraging gestures on speech fluency had never been addressed specifically. In our case, it is possible that encouraging speakers to gesture might have interfered with speech production: The speakers might have been consciously thinking about producing more gestures while talking, and this might have increased the demand on their working memory, not favoring fluency (but see Cook et al., 2012, for evidence showing that drawing attention to gesture may not be detrimental to the working memory). This aspect will be further discussed in Chapter 4.

Third, the results showed that encouraging participants to gesture increased **F0 maximum** and **intensity** features in the speakers' narratives (that is, F0 maximum, Intensity mean and Intensity maximum were significantly higher in the E than in the N condition). As for F0 maximum, the results showed a difference between the two conditions of approximately 22 Hz (see Table 6) which is a highly perceivable difference. In addition, the F0 maximum variable is directly related to pitch range or F0 variation and amplitude measures, which are very relevant in the expression and detection of emphatic speech. Indeed, the acoustic-prosodic changes evident in the current study may not be apparent to an everyday listener in a naturalistic setting and more research is needed before any firm conclusion can be drawn. Nonetheless, our results are in line with proposals that defend a biomechanical interdependence between gestures and the acoustic realization of the co-occurring speech (e.g., Pouw, Harrison, et al., 2019) and future investigations are needed to further explore the existence of a shared motor control system that controls the production of both modalities, i.e., gestures and speech, which can mutually enhance/reduce each other (see Bernardis & Gentilucci, 2006). In this direction, a future analysis of our data could explore if gestures of different strengths affect speech to different degrees over time. Also, future studies might explore whether different types of gestures e.g., representational vs beat-like/non-representational gestures interact with speech prosodic features differently, as discussed in Krahmer & Swerts (2007).

We believe that further work on the effects of both restraining and encouraging gestures on speech is needed: a joint analysis of the results from the two paradigms (i.e., restraining and encouraging gestures) can help to clarify the interactions between gesture and speech prosody. With reference to this, our results might seem in disagreement with those obtained in the restraining gesture experimental setting by Hoetjes at al. (2014). Indeed Hoetjes et al. (2014) showed that the inability to gesture while speaking does not have any detrimental effect on pitch modulation and prosody. However, as previously mentioned, since the two experimental settings are not exact polar opposites, one should not expect polar opposite results. Also, encouraging speakers' to gesture could switch on speakers' preconceptions about how speech produced with bodily expressions should be like, and thus influence speakers' speaking styles.

For example, speakers might interpret the instructions as a request to enact more, or speak in a clearer way, or perhaps slower, etc.. These need to be considered as potential side effects playing a role in our study.

A possible limitation of our study that should be acknowledged is that it does not take into account individual differences that might play a role in speech and gesture production. For example, it has been shown that high extraverted individuals have naturally more fluent speech than low extraverted individuals (e.g., in the case of bilinguals, Dewaele & Furnham, 2000). Also, high and low extraverted individuals may rely on gestures for linguistic fluency to different extents. More generally, individual differences in gesture production in terms of rate, type and physical properties largely depend on individuals' cognitive abilities, personality traits, cultural, linguistic and gender differences (Briton & Hall, 1995; Chu et al., 2014; Gillespie, James, Federmeier, & Watson, 2014; Göksun et al., 2013; Hostetter & Hopkins, 2002; Hostetter & Potthoff, 2012; Kita, 2009; Nicoladis et al., 2018; O'Carroll et al., 2015). Although we controlled as much as possible for linguistic and gender differences (only female speakers from the same regional area in Italy participated in the experiment) we think that, in a future study, a data collection that takes into account other individual differences, e.g., personality traits and cognitive and linguistic abilities, might help to evaluate the extent to which the results can be generalizable (also to different cultures). Also, it could be argued that encouraging speakers to gesture while doing a different task (e.g., explaining a motor task) might yield different results. This too may be worth further investigation.

To conclude, despite some limitations, our study provides evidence that a relationship between gesture production and speech prosodic modulation is in place, in a naturalistic setting. This line of research, building on previous successful clinical attempts that used gesture encouragement as a way to improve the speech of individuals with dysarthria (Garcia & Cobb, 2000; Garcia & Dagenais, 1998; Garcia et al., 2000; Hustad & Garcia, 2005), can provide a fertile ground for therapies in communication disorders. More broadly, and from a more theoretical perspective, this study opens new questions on the direct influence of gestures on speech production that deserve further investigation.

4 Restraning and Encouraging Gestures: Methodological Remarks

Asking speakers to sit on their hands or encouraging them to gesture while speaking can be considered a valid way to explore the role of gestures in speech production. At a first glance, the two paradigms might seem opposites since they are expected to lead, respectively, to no gesture produced vs increased gesture. Though, in general, this seems to be the case, a closer inspection of the data can give a more detailed, less clear-cut picture. For example, it could be the case that even when sitting on their hands, people are still moving to some extent. Likewise, when encouraged to gesture speakers might exhibit a rich and heterogeneous pattern of multimodal cues besides hand gestures. More generally, the same instructions could affect individual speakers' behavior in slightly different ways. This could depend, for instance, on how the instructions are interpreted by each speaker, their personality traits, their baseline gesture production and their individual cognitive resources (Hostetter & Alibali, 2019; Marstaller & Burianová, 2013). In this chapter, we focus on these phenomena by providing a brief overview of some qualitative and quantitative aspects that we observed in the data analyzed in Chapters 2 and 3. Therefore, in this chapter we do not aim to bring exhaustive experimental evidence on these issues; rather, we explore them to highlight their potential and inform future work and methodological approaches.

4.1 Restraining gestures

In previous studies, restraining speakers' gesture has been carried out in different ways. For example, by asking participants (1) not to move their hands while wearing dummy electrodes on hands/arms (Morsella & Krauss, 2004; Rauscher et al., 1996); (2) to wear oven mitts or Velcro gloves attached to a table/board (Hostetter et al., 2007; Pine et al., 2007; Walkington et al., 2019); (3) to grip the bottom of the seat (Jenkins et al., 2017) or sit on their hands (Emmorey & Casey, 2001; Hoetjes et al., 2014; Özer et al., 2017); (4) to fold their arms (Beattie & Coughlan, 1999; Graham & Heywood, 1975); (5) to hold a rod or wear an apron with Velcro (Frick-Horbury & Guttentag, 1998); (6) to sit on purposefully devised armchairs (Finlayson et al., 2003; Rimé et al., 1984). In some cases, however, it was reported that these methods do not necessarily restrain participants from moving other parts of the body, e.g., the forearms, shoulders, head, and legs (as reported in Dobrogaev, 1929; Hoetjes et al., 2014). Indeed, partial immobilization of the body was shown to lead to an increased motor activity in body parts left free to move (Rimé et al., 1984).

In Study 1 (Chapter 2), a visual inspection of our data suggested that asking speakers to sit on their hands had the expected effect of inhibiting gesture use. Nonetheless, we noticed that participants still moved their shoulders, head, and legs while sitting on their hands. As a consequence, we cannot exclude that these movements somehow replaced the actual hand gestures in their role for speech production. In addition, as proposed by Hoetjes et al. (2014), people may still *intend* to produce a gesture even when not realizing an actual movement (i.e., a motor command can be there even though not overtly produced as a gesture). This is in line with McNeill's observation that "even if, for some reason, the hands are restrained and a gesture is not externalized, the imagery it embodies can still be present, hidden but integrated with speech" (McNeill, 2017, p. 78).

A related issue concerns the possible **additional cognitive load** imposed to speakers by the inability to gesture. It cannot be excluded, in fact, that this kind of setting is cognitively demanding and interacts with individual working memory capacity (Hostetter & Alibali, 2019; Marstaller & Burianová, 2013). This would again impact on speech production in different ways across participants. However, the risk of increasing the load resulting from having to remember not to gesture can be reduced by, e.g., asking speakers to sit on their hands rather than simply instructing them not to gesture.

In our study, we aimed to avoid as much as possible unnatural postures and annoying restrictions of any type. When debriefed offline, our participants commented, on the one hand, that they were fairly comfortable with sitting on their hands while speaking. On the other hand, after receiving the instructions, some of them commented spontaneously that speaking without moving their hands would be hard. For example: (EN) *So I can't move them? No, I just cannot / It's going to be hard for me* (IT: *Quindi non posso muoverle? No per me è un dramma*). In one case, a participant explicitly complained about her poor vocabulary and explanatory skills right after having told the story in the restraining gesture condition: (EN) *My poor language skills... It's clear that I've just* *woken up…* (IT*: la mia proprietà di linguaggio… si vede che mi sono svegliata da poco…*). In yet another case, a speaker experienced a long tip-of-the-tongue state (about 15 seconds of speech) while searching for the right word to describe a cat in the act of pawing inside the suitcase (i.e., walking while kneading on clothes): (EN) *It jumps in the suitcase and starts… starts… how do you say it? Starts to… it's so hard without hands!... starts touching the clothes.* (IT: *Entra nella valigia e comincia a…. a…. come si dice... comincia aa... ahm… è difficile senza mani!... toccare i vestiti*). In this case, the action to be described would involve combined alternate movements of both hands (paws). A possible explanation for this could be that the motor simulation of the action was challenged by the inability to gesture, which led to the TOT state. Furthermore, the speaker eventually selected a word that was not semantically rich enough to properly describe the action (touching vs., e.g., kneading; see Hostetter et al., 2007). Finally, it is interesting to note that it was the speaker herself to blame her inability to gesture for this problematic lexical access, which she overtly commented on. Though no conclusions can be drawn from these few examples, they at least provide some insight on how participants cope with the instructions, possibly informing the choice of experimental settings. Moreover, these are interesting cases that, framed in the major theoretical accounts, can inspire future questions.

4.2 Encouraging gestures

Researchers in previous studies have explicitly instructed participants to gesture in order to explore the effects of encouraging the use of gesture on activities such as problem-solving (Beilock & Goldin-Meadow, 2010; Chu & Kita, 2011), learning math (Broaders et al., 2007), second language pronunciation (Baills et al., 2019; Llanes-Coromina et al., 2018), speech fluency and narrative abilities (Vilà-Giménez & Prieto, 2018). These studies have shown that gestures have a beneficial role in thinking, learning, remembering, and speaking. As well, they have shown that instructing participants to gesture generally causes an increase in the participants' gesture rate. Nonetheless, to our knowledge, the only study that has directly focused on the impact of encouraging speakers to gesture on the way they gesture across genres is Parrill, Cabot, Kent, Chen, & Payneau, (2016). The study compared the differences in gesture rate and gesture type of participants that had been and had not been explicitly instructed to gesture while performing three different discourse tasks (i.e., quasi-conversation, spatial problem solving, and narration). In the study, the instruction to gesture did not change gesture rate or gesture type across the different discourse tasks, suggesting that instructing speakers to gesture will not always work (in the sense that it might not lead speakers to produce more gestures); at the same time, the instruction does not seem to impact on the type of gestures produced. In sum, the study appears to be in contrast with previous findings, mentioned above, that show that the instruction to gesture should at least contribute to increasing gesture rate.

However, gesture production may be influenced by a combination of other factors. For instance, it has been shown that gesture rate, together with gesture type and gesture physical forms (size, salience), can change and be adapted depending on (1) the shared knowledge between interlocutors (Gerwing & Bavelas, 2004; Holler & Wilkin, 2009); (2) the interlocutors' (mutual) visibility (Bavelas, Gerwing, Sutton, & Prevost, 2008; Bavelas, Kenwood, Johnson, & Phillips, 2002); (3) the addressee's feedback (e.g., gesture rate lowers when addressees are less attentive (Jacobs & Garnham, 2007)). Moreover, individual differences in gesture production in terms of rate, type, and physical properties largely depend on the individuals' cognitive abilities, personality traits, cultural and gender differences (Briton & Hall, 1995; Chu et al., 2014; Göksun et al., 2013; Hostetter & Hopkins, 2002; Hostetter & Potthoff, 2012; Kita, 2009;

Nicoladis et al., 2018; O'Carroll et al., 2015). These studies suggest that all of these features together, namely, gesture rate, type and salience, are key aspects of how gestures are produced, intended and interpreted in the wild. Therefore, it may well be that instructing participants to gesture can impact all of these aspects in combination, e.g., by increasing gesture rate, and also affecting **gesture types** and **salience**. Also, it may well be that the prompt to gesture switches on some more general speakers' preconceptions about how enhanced bodily expressions should be like. For example, the participants may interpret the instruction as a request to enact more or produce clearer gestures, which should be intentionally designed for the benefit of the addressee. This may result in the use of different multimodal strategies, including forms of **enactment**. Enactments are well documented in sign languages (e.g., *role shifts*, *constructed action*, Ferrara & Johnston, 2014) and can be found in spoken languages as well. They are bodily mimetic demonstrations that give interlocutors a direct visual experience of the original event/action/character. In face-to-face interactions, enactments allow the addressee to feel as they are witnessing the event in real-time (for a review see Stec, 2012).

When setting up an experiment or data collection that requires explicitly asking participants to gesture while speaking, it might be important to assess the possible **general impact** of the instruction to gesture on factors such as gesture type (i.e., representational vs. non-representational gestures). Moreover, the potential impact of instructing speakers to gesture on more general **multimodal cues** can be worth assessing and controlling for in experimental settings.

Effects of encouraging gestures on gesture production

In this section we explore the potential effects of explicitly asking speakers to gesture on the **types** of gesture produced and on more general multimodal strategies, with a focus on **enactment**. We do so by carrying out an additional analysis of the data collected for Study 2 (Chapter 3). In Section 3.2, we ascertained that the gesture elicitation instruction had actually caused an increase in the speakers' gestures with respect to the Non-encouraging condition (N). To do so, the data was manually annotated with the software ELAN (Wittenburg et al., 2006) and all instances of co-speech gestures were identified. The annotation criteria consisted in the marking of all gestural *strokes* and their further distinction into salient vs non-salient depending on where they were performed (see Section 3.2, Figure 4 and Figure 5). The analysis showed that encouraging gestures effectively increased speakers' **gesture rate and salience** (see Section 3.2, Figure 6).

In the additional analysis we present here, we have further annotated the data by distinguishing the gestures performed in both conditions (N, E) between **Representational** (i.e., gestures that represent semantic

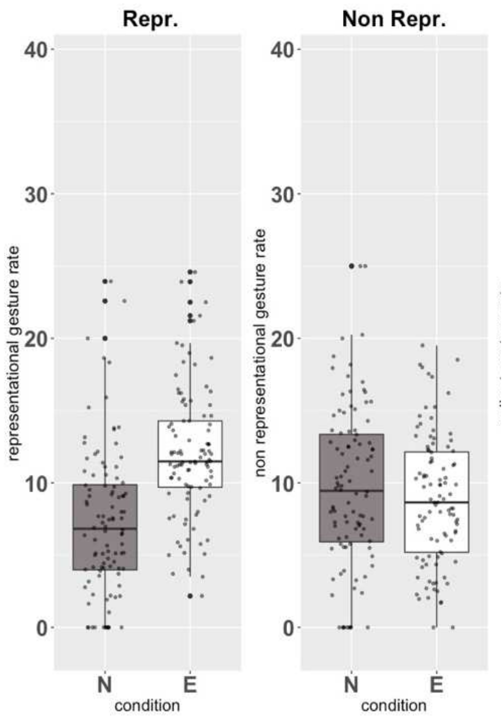


Figure 8 Box plots reporting Representational and Non-representational gesture rate (per 100 words) in the two conditions (Non-encouraging and Encouraging gestures, N, E).

information via handshape, trajectory, or location) and **Non-Representational gestures** (i.e., gestures whose handshape, trajectory and location do not visually recall the referent expressed in the co-occurring speech, see Section 1.2. For example, a *palm up* gesture). The results of the analysis, as shown in Figure 8, show that the rate of representational gestures (per 100 words) is higher in the E condition stories than in the N condition (est. = 4.776, S.E =0.586, $p < .001$)[6]. For Non-Representational gestures rate, there was no significant difference between the two conditions (N, E). This suggests that drawing speakers' attention on gestures impacts not only on gesture rate and their use of space (as found in Chapter 3) but also on the types of gestures speakers produce. That is, when encouraged to gesture, participants make use of **more representational gesture** than they would naturally do without any instruction to gesture. It is worth noting that a narrative task itself may be more likely to elicit more representational gestures (e.g., iconics) compared with other speech tasks and this is worth further investigation.

To evaluate whether the instruction to gesture also led speakers to *enact* more, we have further annotated the data from Study 2 (Chapter 3) by marking as **enactment** any moment in which the speaker made a bodily depiction/demonstration of some actions or characters of the story by means of facial expressions, and combined movements of the hands, head and shoulders, as illustrated in Figure 9. We found a significantly higher

---

[6] The effect of gesture encouragement (within-subjects factor) on gesture was tested via two Linear Mixed Effects Models (R function *lmer* in *lme4* package; see Bates et al., 2014). Each model included: Representational (R) gesture rate or Non-Representational (NR) gesture rate and had Condition (N, E) as a fixed effect and both Story and Participant as random intercepts. P-values are obtained by likelihood ratio tests of the full model against the model without the fixed effect of interest (i.e., Condition).

occurrence of enactments in the E condition (N=7; E=44; t= 3.944, p< .001), suggesting that the gesture encouragement instruction can activate and affect other body articulators and lead to more general involvement of the body in the storytelling.

To conclude, these effects on gesture production can give a clearer idea of how speakers interpret the prompt to gesture. For example, participants might interpret the instruction as a request to produce clearer gestures that are intentionally designed for the **benefit of the addressee**. This might result in the use of more representational gestures or instances of enactment.

Figure 9: Examples of enactments. From left to right: **(A)** E lei (la gattina) si volta così: sprezzante*;* (EN) *And the cat turns away like this: haughty* | **(B)** Arriva il gatto all'improvviso con gli artigli di fuori a far paura (agli uccelli); (EN) *And the cat suddenly arrives with its claws out to scare (the birds)* | **(C)** la gatta sgrana gli occhi; (EN) *the cat with its eyes wide open* | **(D)** E il gatto grande si… si schifa; (EN) *and the bigger cat is… is disgusted* | **(E)** Si allontana e lo guarda male. (EN) *it walks away and glares at him.*

Furthermore, being aware of the impact of explicitly asking speakers to gesture on the way participants gesture may be worth consideration. For example, Krahmer & Swerts (2007) speculated that gesture of different types may have different effects on the speech produced, due to the fact that they may have different "sources" in the speech production process (e.g., beat like movements arising later, at the formulation phase, and representational gesture at an earlier stage of the conceptualization phase). Therefore, asking speakers to gesture would not be a good elicitation method if the main interest of a study is, for example, assessing how beat gestures specifically impact on prosodic features. We believe that this question should be addressed in future research, perhaps finding a way to specifically elicit beat-like movements in semi-spontaneous speech.

Lastly, in the data collected for Study 2, some participants commented on the encouraging condition prompt by saying that they would already gesture a lot without the need to be asked to do it. This suggests that the same instruction can potentially be interpreted in different ways by different speakers, e.g., those who do not gesture much in their baseline/natural speaking style vs spontaneous high-gesturers. Moreover, it is worth discussing a couple of notable cases that can be of interest to explore how speakers coped with the instruction to gesture. In one case, a participant described a scene where the cat is holding a mouse in its paw and used the word "hand" instead of "paw". Something similar happened also to a different participant. She described a scene where the cat is on top of an aquarium and puts a paw inside it to fetch a fish; she says "arm" while enacting the action, and then she makes a self-correction and says "a paw": EN: *therefore it puts an arm…* ***a paw, not an arm****, inside the aquarium* (IT: *quindi mette un brac…* ***una zampa non un braccio*** *dentro l'acquario*). This question may be worth further consideration, as it suggests that encouraging speakers to gesture may enhance the speakers' mental simulation of motor actions while speaking but this can end up, in some cases, to **interfere** with speech production and lexical search.

To conclude, we believe that both restraining and encouraging the use of gestures can lead to some side effects that can be worth considering depending on the different research questions to be addressed. Also, the two kinds of instructions can lead to differences in the individual speaker's interpretations of the prompts that can, in fact, interact differently with speech production.

5 General Discussion and Conclusions

This dissertation reports on two empirical studies aimed to explore the effects of restraining and encouraging the use of gesture on speech prosodic features. In the two studies, participants were asked to describe short stories to a listener and, in doing so, they were either restrained or encouraged to gesture. The resulting stories were analyzed by taking into account a comprehensive set of measures related to temporal narrative features, fluency measures, and acoustic measures related to pitch and intensity. The results provide evidence that encouraging the use of gestures can have effects on the speech produced in the narration task in terms of speech length and prosodic features (i.e., increased F0 maxima values and intensity); by contrast, restraining gestures does not seem to affect speech. These results will be commented below in light of previous research and hypotheses.

5.1 Restraining the use of gestures

The first study (Chapter 2) provided evidence that restraining gestures does not significantly impact on discourse length, fluency or F0 and intensity measures in semi-spontaneous narrative speech.

Speech length

As for speech length measures, we expected that when gestures are restrained, speech length would change for different reasons. These can be related to the selection of the information to be included in speech. For example, when relying on the speech modality only, speakers cannot count on the use of gestures to express some information visually, e.g., showing with the hands the shape of an object, or spatial relations between objects. As a consequence, speakers would need to integrate the information that cannot be expressed visually by using more words (Emmorey & Casey, 2001; Melinger & Levelt, 2004, but not confirmed in Hostetter et al., 2007). On the other hand, the speech "packaging" process can be affected when the gesture stream is inhibited. This may lead speakers, for example, to just exclude some (spatial) information from their speech (Kita, 2000). As a consequence, speech might result to be shorter. However, Finlayson et al., (2003) and Graham & Heywood (1975) respectively found longer speech when speakers were unable to gesture, or the use of more spatial content phrases and words. Indeed, other studies did not find any effect of the inability to gesture on speech length (Hoetjes et al., 2014; Hostetter et al., 2007). Based on these mixed results, we initially speculated that the inability to gesture could have an impact on the selection of information to be verbally expressed, on the speech chunking strategy and utterance length. We expected that these variations might become detectable as differences in the number of words or speech length. However, we did not find any significant difference in speech length between the stories told in the two experimental conditions (i.e., the restraining gesture condition vs. the non-restraining condition).

In the future, to assess the possible differences between the two conditions, the *amount* and *type* of *information* that is expressed verbally could be analyzed together with the prosodic structure of the speech. This could provide evidence on whether inhibiting the gesture stream leads speakers to lower the amount of imagery and activity content in their speech (as in Rimé et al., 1984) or to reduce reference to spatio-motor information or its semantic richness (as tested but only partly confirmed in Hostetter et al., 2007). Also, through the comparison of the potentially different prosodic strategies between the two conditions, (e.g., regarding utterance length, information structure) it might be possible to better assess the impact of the inability to gesture on prosodic planning.

Fluency

We expected that restraining speakers from gesturing would lead to lower paced and more disfluent speech, as an effect of either lexical access difficulties (Rauscher et al., 1996) or more general planning difficulties (Kita et al., 2017; Kita, 2000). In the former case, it may be that the inability to gesture makes speakers struggle more in finding the right lexical items when needed. On the other hand, according to the latter proposals, gestures can offer a powerful way to organize imagistic components to be expressed in the linear stream of speech. This means that inhibiting gestures may challenge the speakers who are trying to describe any kind of spatio-motor event. However, the results showed no effects of restraining the use of gestures on fluency.

The comic strips that were used in the study feature actions by animals and people, for example, a cat stealing a cooked chicken from a table, a character trying to put the cat in a cat carrier (see the Appendix). The elements featured in the comic strips were every-day objects/actions that could be easy to recall by the participants telling the stories. Also, stories consisted of short sequences which were rather easy to memorize. A possible explanation of our results on fluency is that speakers accommodated for the inability to gesture because they were free to tell simple stories at their own length and pace, with no particular challenge imposed on memory or lexicon. As a consequence, even though spatio-motoric thinking was challenged by the inability to gesture, speakers might have compensated by drawing upon other cognitive strategies and capitalizing on analytical thinking. However, along these lines, even though hand gestures were restrained, we cannot exclude that any other movements participants did with their heads, legs and shoulders had some role in the process of speaking. In other words, the imagery and actions to be described can still be embodied/simulated by other body movements e.g., legs, head, lips and this could still play some positive role in the process of speaking.

Prosodic features

As for the acoustic analysis, we explored whether restraining gestures has any effects on acoustic properties such as fundamental frequency and intensity. Our results, consistently with a study that investigated the same issue directly (Hoetjes et al., 2014), show that the inability to gesture does not impact F0 and intensity metrics. However, studies investigating how gesture production interacts with speech acoustics and spectral properties have suggested that the two modalities are related both at a motor control level (Bernardis & Gentilucci, 2006) and biomechanically (Pouw, Harrison et al., 2019), and are both controlled by a common neural network that is connected to Broca's area (Marstaller & Burianová, 2015). Also, enhancing the gesture stream has been found to enhance the spectral properties of the co-occurring speech (Bernardis & Gentilucci, 2006; Krahmer & Swerts, 2007). Therefore, by contrast, we conjectured that restraining gesture use would inhibit prosodic features by lowering the speech acoustics measures that are typically associated with emphatic speech, prosodic contrasts, and expressivity (F0 excursions and intensity). However, our results did not meet our expectations. A possible explanation for the lack of effect on **prosodic features** is that, when unable to gesture, speakers may try to compensate for the absence of gestures with prosodic strategies; however, since the two modalities are connected at the motor and biomechanical level, the speakers' attempt would eventually not be effective. A second explanation might be that sitting on hands does not restrain movements completely (as mentioned above and in Chapter 4). One can wonder whether, in the process of speaking, moving other parts of the body can have a comparable impact on speech as producing hand gestures.

To conclude, the fact that speakers, in general, well accommodated for the inability to gesture might be due to the fact that we chose a non-invasive restraining method that was reported to be quite comfortable (Chapter 4), and possibly even being a spontaneous posture in some real-life circumstances. Along these lines, another explanation could be that in some every-day circumstances (e.g., formal situations), people are used to reduce and control for their gestures while speaking.

5.2 Encouraging the use of gestures

The second study (Chapter 3) showed that encouraging speakers to gesture effectively increased speakers' gesture rate and gesture salience, and led to (1) the production of longer discourse in terms of number of words used; (2) some modest changes in acoustic-prosodic features of speech, namely higher fundamental frequency maxima and louder speech.

Speech length

Similarly to what has been discussed above (Section 2.4), we expected that drawing speakers' attention to producing gestures might influence speech production on at least two levels, both involving different processes: the selection of the information to be included in speech, and information packaging processes required for speech production. Both of these levels may result in different speech chunking strategies and different amount of information included in speech. Also, it may well be that both the selection of information and its packaging process potentially involve prosodic structuring at some point. Our results suggest that encouraging the use of gesture can impact on these levels. However, in the future, further textual analysis, with a focus on speech content and structure, will be needed (1) to disentangle whether these effects can be explained in light of the major theoretical models for speech-gesture production and (2) to explore the idea that gestures and prosodic structures are potentially co-planned in the process of speech production (Shattuck-Hufnagel, 2019).

Fluency

As for fluency, we did not find evidence that having speakers produce more gestures leads them to speak more fluently. Based on the predictions of the main theoretical models (Section 1.3), we expected that gesturing more would lead speakers to speak more fluently. However, the direct influence of encouraging gestures on fluent (connected) speech had not been previously explored. A possible explanation of our results is that, when encouraged to gesture, speakers might have been consciously thinking about producing more gestures while telling the stories. Therefore, the prompt to gesture might have increased the demand on the speakers' working memory (though, see Cook et al., 2012, for evidence showing that drawing attention to gesture may not be detrimental to the working memory). Along these lines, it is also possible that speakers encouraged to gesture had to *plan* the gestures to produce while telling the stories and this may have generally interfered with the speech planning and production processes not favoring fluency. However, we did not find that instructing speakers to gesture is detrimental to speech fluency in any way. Finally, as for **speech rate**, it may well be that asking speakers to gesture more while speaking can have led them to rather aim for *clarity* of speech, and thus not cause them to speak faster at all.

Prosodic features

As for fundamental frequency (F0) and intensity measures, the results are in line with the predictions of the studies reviewed in Section 1.3. If speech and gesture share a motor control system, it may well be that enhancing the gesture stream would enhance speech and result in changes in acoustic patterns (Bernardis & Gentilucci, 2006; Krahmer & Swerts, 2007). Probably the most important finding reported in this investigation was that, while restraining the use of gesture does not seem to impact speech F0 and intensity (as also found in Hoetjes et al., 2014) encouraging gestures' use has a more direct "enhancing effect" on the acoustics of speech. This idea is also supported by brain-imaging studies claiming that speech and gesture production are both controlled by a common motor control system engaged with a neural network connected to Broca's area (e.g., Gentilucci & Dalla Volta, 2008; Marstaller & Burianová, 2015).

However, in the future, further analysis of our audio-visual data (i.e., story-telling set) can explore whether gestures of different strength and kinematics can directly impact on speech acoustics over time. This would serve to explore the potential biomechanic interdependence between speech and gestures that Pouw, Harrison, et al. (2019) have found in steady vocalizations produced in the lab, and have claimed to play a role in the emergence of gesture-speech synchrony. We believe that this would shed more light on the complex interconnectedness between speech articulation and gesture production: together with a biomechanical interdependence between the two modalities, prosodic contrasts and expressivity can be enhanced and "helped" by the gesture modality. Also, a future contribution would be to explore the potentially different effects of different types of gestures on prosodic features (as suggested by Krahmer & Swerts, 2007). Further investigations in this direction would contribute to clarifying the direct mechanisms and functions that gestures have for speech production and the potential role of gestures on the planning and generation of prosodic structures in the process of speech production (Shattuck-Hufnagel, 2019).

To conclude, the main contribution of this dissertation is that it explored the potential functions of gesture production in the speech planning and articulation phase, suggesting that gesture production could enhance prosodic and acoustic features of speech. At the same time, it showed that the inability to gesture can be well accommodated in terms of fluent speech production and spoken prosody in semi-spontaneous narrative speech. However, future work is needed to further explore how (and if) gesture and prosodic structure are jointly planned and produced.

References

Abner, N., Cooperrider, K., & Goldin-Meadow, S. (2015). Gesture for Linguists: A Handy Primer. *Linguistics and Language Compass*, *9*(11), 437–451.

Alibali, M. W., Yeo, A., Hostetter, A. B., & Kita, S. (2017). Representational gestures help speakers package information for speaking. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why Gesture?: How the hands function in speaking, thinking and communicating* (pp. 15–38). Amsterdam/Philadelphia: John Benjamins Publishing Company

Arantes, P., Eriksson, A., & Lima, V. G. (2018). Minimum sample length for the estimation of long-term speaking rate. *Proceedings of the 9th International Conference on Speech Prosody*, 661–665.

Arantes, P., & Lima, V. G. (2017). Towards a Methodology To Estimate Minimum Sample Length for Speaking Rate. *Revista Do GEL*, *14*(2), 183–197.

Arbib, M. A. (2012). *How the brain got language: The mirror system hypothesis* (Vol. 16). Oxford: Oxford University Press.

Armstrong, D. F., & Wilcox, S. E. (2007). *The gestural origin of language*. New York: Oxford University Press.

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.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv Preprint ArXiv:1406.5823*.

Bavelas, J., Chovil, N., Coates, L., & Roe, L. (1995). Gestures Specialized for Dialogue. *Personality and Social Psychology Bulletin, 21*(4), 394–405.

Bavelas, J., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, *15*(4), 469–489.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*(2), 495–520.

Bavelas, J., Kenwood, C., Johnson, T., & Phillips, B. (2002). An experimental study of when and how speakers use gestures to communicate. *Gesture*, *2*(1), 1–17.

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

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture Changes Thought by Grounding It in Action. *Psychological Science*, *21*(11), 1605–1610.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Bergmann, C., Sprenger, S. A., & Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychologica*, *161*, 25–35.

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

Blomgren, M., Chen, Y., Ng, M. L., & Gilbert, H. R. (1998). Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *The Journal of the Acoustical Society of America*, *103*(5), 2649–2658.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.43. Retrieved from http://www.praat.org/

Briton, N. J., & Hall, J. A. (1995). Beliefs about female and male nonverbal communication. *Sex Roles*, *32*(1–2), 79–90.

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making Children Gesture Brings Out Implicit Knowledge and Leads to Learning. *Journal of Experimental Psychology: General*, *136*(4), 539–550.

Brown, A. S. (1991). A Review of the Tip-of-the-Tongue Experience. *Psychological Bulletin*, *109*(2), 204–223.

Campisi, E., & Mazzone, M. (2016). Do people intend to gesture? A review of the role of intentionality in gesture production and comprehension. *Reti, Saperi, Linguaggi*, *3*(2), 285–300.

G. R. Hammond (Ed.). Cerebral Control of Speech and Limb Movements. (1990). In, *Advances in Psychology* (Vol. 70) Amsterdam, The Netherlands: Elsevier Science.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102–116.

Chu, M., & Kita, S. (2016). Co-thought and co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 257–

270.

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: the role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, *143*(2), 694–709.

Church, R. B., Alibali, M. W., & Kelly, S. D. (Eds.). (2017). *Why gesture? : how the hands function in speaking, thinking and communicating* (Vol. 7). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Cienki, A., & Müller, C. (2008). Metaphor, gesture, and thought. In R. W. Gibbs JR (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 483–501). Cambridge: Cambridge University Press.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of Memory and Language*, *63*(4), 465–475.

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.

Cooperrider, K., Abner, N., & Goldin-Meadow, S. (2018). The Palm-Up Puzzle: Meanings and Origins of a Widespread Form in Gesture and Sign. *Frontiers in Communication*, *3*(23), 1–16.

Corballis, M. C. (1999). The Gestural Origins of Language: Human language may have evolved from manual gestures, which survive today as a “behavioral fossil” coupled to speech. *American Scientist*, *87*, 138–145.

Corballis, M. C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences*, *26*(02), 199–208.

Cravotta, A., Prieto, P., & Busà, M. G. (2019). Encouraging Gesture Use in a Narration Task Increases Speakers’ Gesture Rate, Gesture Salience and the Production of Representational Gestures. *Proceedings of the*

6th Conference on Gesture and Speech in Interaction (GeSpIn). Paderborn, Germany, 21-26.

De Iorio, A. (1832). *La mimica degli antichi investigata nel gestire napoletano*. Napoli: stamperia e cartiera del Fibreno.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390.

de Ruiter, J. P. (2017). The asymmetric redundancy of gesture and speech. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why Gesture?: How the hands function in speaking, thinking and communicating* (Vol. 7, pp. 59–75). Amsterdam/Philadelphia: John Benjamins Publishing Company.

De Ruiter, J. P. (2000). The production of gesture and speech. In McNeill, D. (Ed.), *Language and Gesture* (Vol. 2, pp. 284–311). Cambridge: Cambridge University Press.

Dewaele, J.-M., & Furnham, A. (2000). Personality and speech production: A pilot study of second language learners. *Personality and Individual Differences*, *28*(2), 355–365.

Dobrogaev, S. M. (1929). Ucnenie o reflekse v problemakh iazykovedeniia [Observations on reflexes and issues in language study]. *Iazykovedenie i Materializm*, 105–173.

Efron, D. (1941). *Gesture and environment.* Oxford, UK: King's Crown Press.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, *1*(1), 49–98.

Emmorey, K., & Casey, S. (2001). Gesture, thought and spatial language. *Gesture*, *1*(1), 35–50.

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., & 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.

Ferrara, L., & Johnston, T. (2014). Elaborating who's what: A study of constructed action and clause structure in Auslan (Australian sign language). *Australian Journal of Linguistics*, *34*(2), 193–215.

Fillmore, C. J., Kempler, D., & Wang, W. S. Y. (Eds.). (2014). *Individual differences in language ability and language behavior*. New York: Academic Press.

Finlayson, S., Forrest, V., Lickley, R., & Beck, J. M. (2003). Effects of the restriction of hand gestures on disfluency. In R. Eklund (Ed.), *Proceedings of Disfluency in Spontaneous Speech Workshop (DiSS '03), Gothenburg Papers in Theoretical Linguistics* (Vol. 90, pp. 21–24). Göteborg, Sweden.

Flanagan, J. R., Ostry, D. J., & Feldman, A. G. (1990). Control of Human Jaw and Multi-Joint Arm Movements. In G. E. Stelmach & P. A. Vroon (Eds.), *Advances in Psychology. Cerebral Control of Speech and Limb Movements* (Vol. 70, pp. 29–58). Amsterdam, The Netherlands: Elsevier Science.

Frick-Horbury, D., & Guttentag, R. E. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. *The American Journal of Psychology*, *3*(1), 43–62.

Garcia, J. M., Cannito, M. P., & Dagenais, P. A. (2000). Hand Gestures: Perspectives and Preliminary Implications for Adults with Acquired Dysarthria. *American Journal of Speech-Language Pathology*, *9*(2), 107–115.

Garcia, J. M., & Cobb, D. (2000). The effects of gesturing on speech intelligibility and rate in ALS dysarthria: A case study. *Journal of Medical Speech-Language Pathology*, *8*(4), 353–357.

Garcia, J. M., & Dagenais, P. A. (1998). Dysarthric sentence intelligibility: contribution of iconic gestures and message predictiveness. *Journal of Speech, Language and Hearing Research, 41*(6), 1282–1293.

Gentilucci, M., & Dalla Volta, R. (2008). Spoken language and arm gestures are controlled by the same motor control system. *The Quarterly Journal of Experimental Psychology, 61*(6), 944–957.

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture, 4*(2), 157–195.

Gillespie, M., James, A. N., Federmeier, K. D., & Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition, 132*(2), 174–180.

Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the Revolution to Embodiment. *Perspectives on Psychological Science, 8*(5), 573–585.

Göksun, T., Goldin-Meadow, S., Newcombe, N. S., & Shipley, T. (2013). Individual differences in mental rotation: What does gesture tell us? *Cognitive Processing, 14*(2), 153–162.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's Role in Speaking, Learning, and Creating Language. *Annual Review of Psychology, 64*(1), 257–283.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychological Science, 12*(6), 516–522.

Götz, S. (2013). Fluency in native and nonnative English speech. In *Studies in Corpus Linguistics* (Vol. 53). Amsterdam, The Netherlands: John Benjamins Publishing.

Graham, J. A., & Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology, 5*(2), 189–195.

Graziano, M. (2014). The development of two pragmatic gestures of the so-called Open Hand Supine family in Italian children. In M. Seyfeddinipur & M. Gullberg (Eds.), *From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon.* (pp. 311–330). Amsterdam: John Benjamins Publishing Company.

Graziano, M., & Gullberg, M. (2018). When Speech Stops, Gesture Stops: Evidence From Developmental and Crosslinguistic Comparisons. *Frontiers in Psychology*, *9*(879), 1–17.

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

Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, *115*(1–2), 147–172.

Hagoort, P., & Indefrey, P. (2014). The Neurobiology of Language Beyond Single Words. *Annual Review of Neuroscience*, *37*, 347–362.

Hewes, G. W. (1992). Primate Communication and the Gestural Origin of Language. *Current Anthropology*, *33*(S1), 65–84.

Hincks, R. (2005). Measures and perceptions of liveliness in student oral presentation speech: A proposal for an automatic feedback mechanism. *System*, *33*(4), 575–591.

Hoetjes, M., Krahmer, E., & Swerts, M. (2014). Does our speech change when we cannot gesture? *Speech Communication*, *57*, 257–267.

Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes*, *24*(2), 267–289.

Hollien, H., & Michel, J. F. (1968). Vocal Fry as a Phonational Register. *Journal of Speech and Hearing Research*, *11*(3), 600–604.

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 & Review*, *26*(3), 721–752.

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). Does sitting on your hands make you bite your tongue? The effects of gesture prohibition on speech during motor descriptions. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, *29*, 1097-1102. Nashville, USA.

Hostetter, A. B., & Hopkins, W. D. (2002). The effect of thought structure on the production of lexical movements. *Brain and Language*, *82*(1), 22–29.

Hostetter, A. B., & Potthoff, A. L. (2012). Effects of personality and social situation on representational gesture production. *Gesture*, *12*(1), 62–83.

Hübscher, I. (2018). *Preschoolers ' pragmatic development : How prosody and gesture lend a helping hand*. PhD Diss: Universitat Pompeu Fabra, Barcelona.

Hübscher, I., Garufi, M., & Prieto, P. (2019). The development of polite stance in preschoolers: how prosody, gesture, and body cues pave the way. *Journal of Child Language*, *46*(05), 825–862.

Hustad, K. C., & Garcia, J. M. (2005). Aided and Unaided Speech Supplementation Strategies: Effect of Alphabet Cues and Iconic Hand Gestures on Dysarthric Speech. *Journal of Speech, Language and Hearing Research*, *48*, 996–1012.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*(255), 1–16.

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

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, *6*(11–12), 19–40.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language*, *56*(2), 291–303.

Jenkins, T., Coppola, M., & Coelho, C. (2017). Effects of gesture restriction on quality of narrative production. *Gesture*, *16*(3), 416–431.

Kearney, E., & Guenther, F. H. (2019). Articulating: the neural mechanisms of speech production. *Language, Cognition and Neuroscience*, 34(9), 1214–1229.

Kendon, A. (1972). Some relationships between body motion and speech. In Siegman, A. & Pope, B. (Eds), *Studies in Dyadic Communication, 7*, 177-210, New York: Pergamon Press

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*, 207–227. The Hague: Mouton

Kendon, A. (1992). Some Recent Work from Italy on Quotable Gestures (Emblems). *Journal of Linguistic Anthropology*, *2*(1), 92–108.

Kendon, A. (1997). Gesture. *Annual Review of Anthropology*, *26*(1), 109–128.

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

Kendon, A. (2008). Some reflections on the relationship between 'gesture' and 'sign.' *Gesture, 8*(3), 348–366.

Kendon, A. (2017). Pragmatic functions of gestures. *Gesture*, *16*(2), 157–175.

Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes, 24*(2), 145–167.

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review, 3*(124), 245.

Kita, S. van Gijn, I., & van der Hulst, H. (1997). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth & M. Fröhlich (Eds.), *Gesture and Sign Language in Human-Computer Interaction. Proceedings of the International Gesture Workshop* (Vol. 1371, pp. 23–35).

Kita, S. (2000). How representational gestures help speaking. In McNeill, D. (Ed.), *Language and Gesture* (pp. 162–185). Cambridge: Cambridge University Press.

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.

Kormos, J. (2014). *Speech production and second language acquisition*. New York: Routledge. Taylor & Francis Group.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*(3), 396–414.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: a process model. In McNeill, D. (Ed.), *Language and Gesture* (pp. 261–283). Cambridge: Cambridge University Press.

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)*(3), 1–26.

Ladefoged, P. (2003). *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Malden, Massachusetts: Wiley- Blackwell.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, Massachusetts: MIT press.

Levelt, W. J. M. (1999). Producing spoken language: a blueprint of the speaker. In P. Hagoort & C. . M. Brown (Eds.), *The neurocognition of language* (pp. 83–122). Oxford, UK: Oxford University Press.

Llanes-Coromina, J., Prieto, P., & Rohrer, P. (2018). Brief training with rhythmic beat gestures helps L2 pronunciation in a reading aloud task. *9th International Conference on Speech Prosody 2018*, 498–502.

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

Lucero, C., Zaharchuk, H., & Casasanto, D. (2014). Beat gestures facilitate speech production. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, *36*. Quebec City, Canada.

Marstaller, L., & Burianová, H. (2013). Individual differences in the gesture effect on working memory. *Psychonomic Bulletin & Review*, *20*(3), 496–500.

Marstaller, L., & Burianová, H. (2015). A common functional neural network for overt production of speech and gesture. *Neuroscience*, *284*, 29–41.

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

McGlone, R. E. (1967). Air flow during vocal fry phonation. *Journal of Speech and Hearing Research*, *10*(2), 299–304.

McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. (2013). *Coh-Metrix version 3.0 [Computer software]*. Retrieved from http://cohmetrix.com/

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

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

McNeill, D. (2017). Gesture-speech unity. What it is, where it came from. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why Gesture?: How the hands function in speaking, thinking and communicating* (Vol. 7, pp. 77–102). Amsterdam/Philadelphia: John Benjamins Publishing Company

McNeill, D., & Duncan, S. (2000). Growth points in thinking-for-speaking. In McNeill, D. (Ed.), *Language and Gesture* (pp. 141–161). Cambridge: Cambridge University Press.

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture*, *4*(2), 119–141.

Mol, L., Krahmer, E., & van de Sandt-Koenderman, M. (2013). Gesturing by Speakers With Aphasia: How Does It Compare? *Journal of Speech, Language, and Hearing Research*, *56*(4), 1224–1236.

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology*, *117*(3), 411–424.

Munari, B. (1963). *Supplemento al dizionario italiano*. Milano: Muggiani.

Murry, T. (1971). Subglottal Pressure and Airflow Measures During Vocal Fry Phonation. *Journal of Speech and Hearing Research*, *14*(3), 544–551.

Nicoladis, E., Nagpal, J., Marentette, P., & Hauer, B. (2018). Gesture frequency is linked to story-telling style: evidence from bilinguals. *Language and Cognition*, *10*(4), 641–664.

Novack, M. A., & Goldin-Meadow, S. (2017). Gesture as representational action: A paper about function. *Psychonomic Bulletin & Review*, *24*(3), 652–665.

O'Carroll, S., Nicoladis, E., & Smithson, L. (2015). The effect of extroversion on communication: Evidence from an interlocutor visibility manipulation. *Speech Communication, 69,* 1–8.

Özçalişkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, *96*(3), B101–B113.

Özer, D., Tansan, M., Özer, E. E., Malykhina, K., Chatterjee, A., & Göksun, T. (2017). The effects of gesture restriction on spatial language in young and elderly adults. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 1471-1476. London, UK.

Parrill, F., Cabot, J., Kent, H., Chen, K., & Payneau, A. (2016). Do people gesture more when instructed to? *Gesture*, *15*(3), 357–371.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic Inquiry and Word Count: LIWC 2015 [Computer software]*. Retrieved from http://www.liwc.net/

Pine, K. J., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science*, *10*(6), 747–754.

Ping, R., & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science*, *34*(4), 602–619.

Poggi, I. (1983). La mano a borsa: analisi semantica di un gesto emblematico olofrastico. *Comunicare Senza Parole. La Comunicazione Non-Verbale Nel Bambino e Nell'interazione Sociale Tra Adulti. Roma: Bulzoni*, 219–238.

Poggi, I. (2002). Symbolic gestures: The case of the Italian gestionary. *Gesture*, *2*(1), 71–98.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, *77*(2), 97–132.

Pouw, W., de Nooijer, J. A., van Gog, T., Zwaan, R. A., & Paas, F. (2014). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology*, *5*, 1–14.

Pouw, W., Harrison, S. J., & Dixon, J. A. (2019). Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony. *Journal of Experimental Psychology: General.* http://dx.doi.org/10.1037/xge0000646

Pouw, W., Mavilidi, M.F., van Gog, T., & Paas, F. (2016). Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity. *Cognitive Processing*, *17*(3), 269–277.

Pouw, W., Paxton, A., Harrison, S., Dixon, J. A., Grossma, S., Hart, O., & Vick, W. (2019). *Social Resonance: Acoustic Information about Upper Limb Movement in Voicing.* PsyArxXiv, May 2019, https//doi.org/10.31234/osf.io/ny39e

Prieto, P., Cravotta, A., Kushch, O., Rohrer, P. L., & Vilà-Giménez, I. (2018). Deconstructing beat gestures: A labelling proposal. *Proceedings of the 9th International Conference on Speech Prosody*, 201–205.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, *7*(4), 226–231.

Ravizza, S. (2003). Movement and lexical access: Do noniconic gestures aid in retrieval? *Psychonomic Bulletin and Review*, *10*(3), 610–615.

Rimé, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, *8*(4), 311–325.

Seyfeddinipur, M., & Kita, S. (2001). Gesture as an Indicator of Early Error Detection in Self-Monitoring of Speech. *Proceedings of DiSS'01:Disfluency in Spontaneous Speech*, 29–33. Edinburgh, Scotland, UK.

Seyfeddinipur, M., Kita, S., & Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition*, *108*(3), 837–842.

Shattuck-Hufnagel, S. (2019). Toward an (even) more comprehensive model of speech production planning. *Language, Cognition and Neuroscience*, 1–12.

Shattuck-Hufnagel, S., & Prieto, P. (2019). Dimensionalizing co-speech gestures. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1490–1494). Melbourne, Australia: Canberra, Australia: Australasian Speech Science and Technology Association Inc.

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), 1–13.

Stec, K. (2012). Meaningful shifts. *Gesture*, *12*(3), 327–360.

Stokoe, W. C. (2001). *Language in hand: Why sign came before speech.* Washington: Gallaudet University Press.

Streeck, J. (1994). Gesture as Communication II: The Audience as Co-Author. *Research on Language and Social Interaction*, *27*(3), 239–267.

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: a comparative analysis. *Journal of Phonetics*, *30*(4), 629–654.

Tomasello, M. (2008). *Origins of human communication.* Cambridge, Massachusetts: MIT press.

Tuite, K. (1993). The production of gesture. *Semiotica*, 83–105.

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. (2018). Encouraging children to produce rhythmic beat gestures leads to better narrative discourse performances. *Proceedings of the 9th International Conference on Speech Prosody 2018*, 704–708.

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

Walkington, C., Woods, D., Nathan, M. J., Chelule, G., & Wang, M. (2019). Does restricting hand gestures impair mathematical reasoning? *Learning and Instruction*, *64*, 101225.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*. Genova, Italy.

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (pp. 41–62). Chichester: John Wiley.

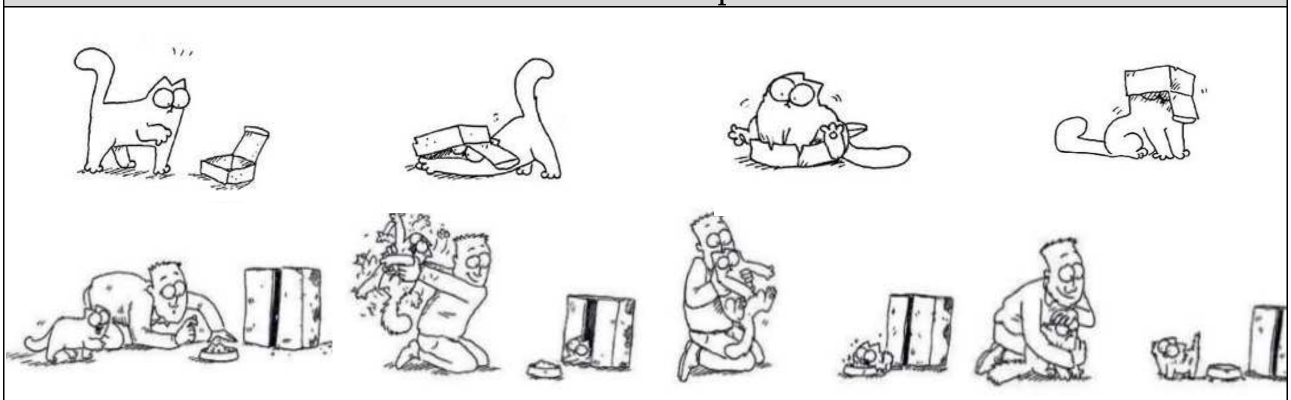
Appendix

Materials for Study 1 and Study 2

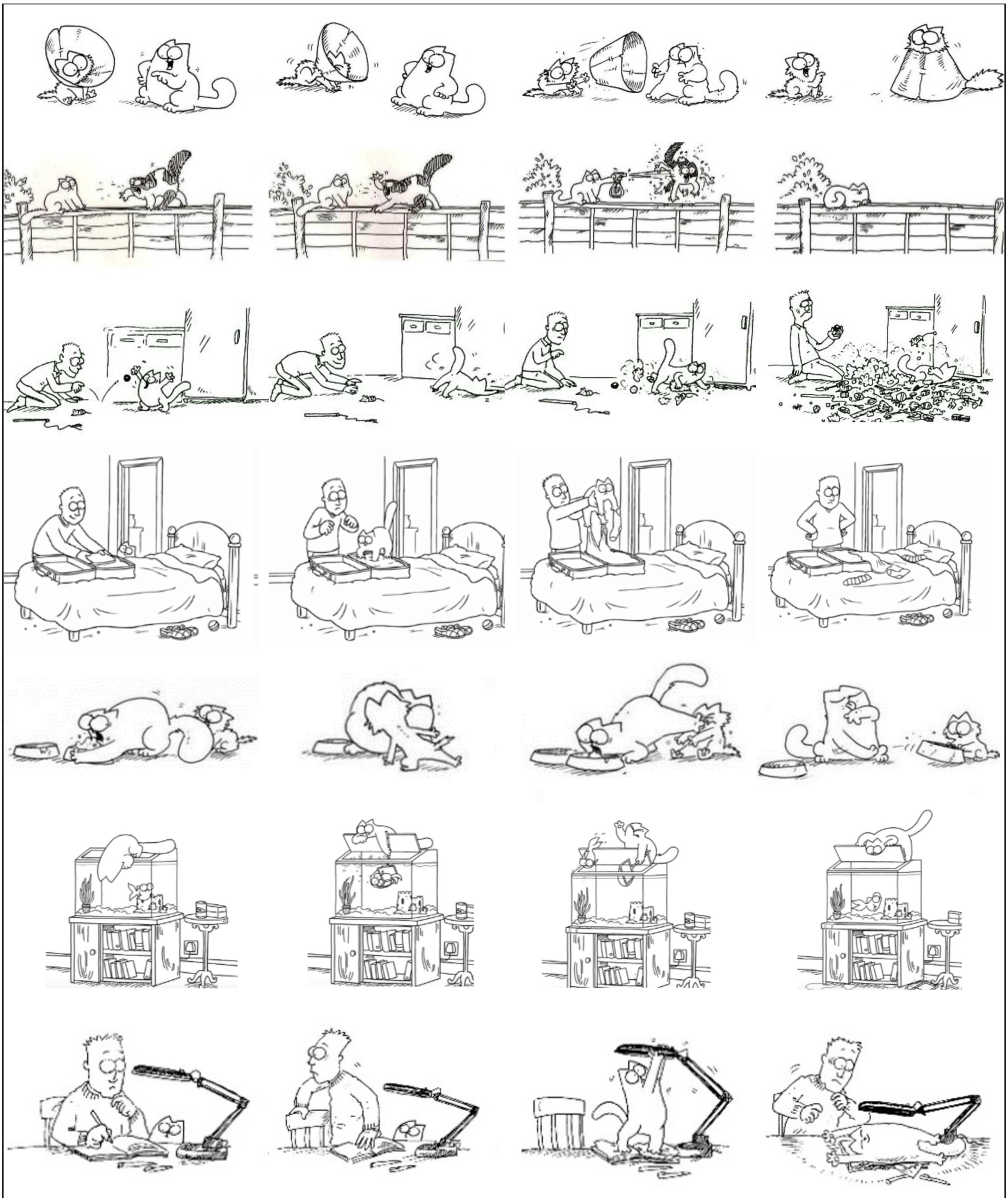
The comic strips used for both studies were adapted from *Simon's Cat* by Simon Tofield and are reproduced below with permission.

Study 1: Restraining Gestures

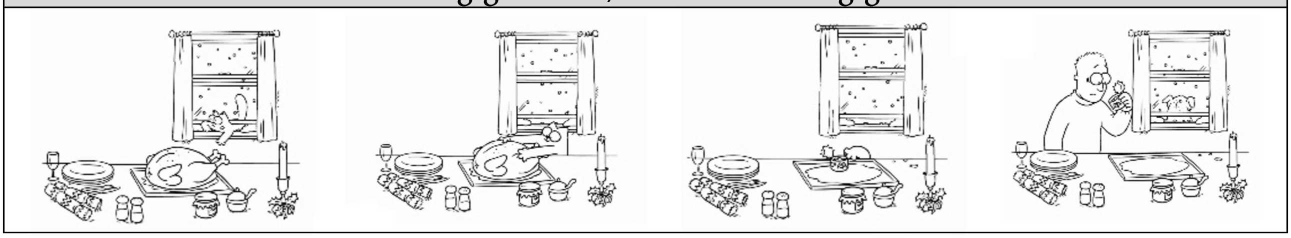
Familiarization phase

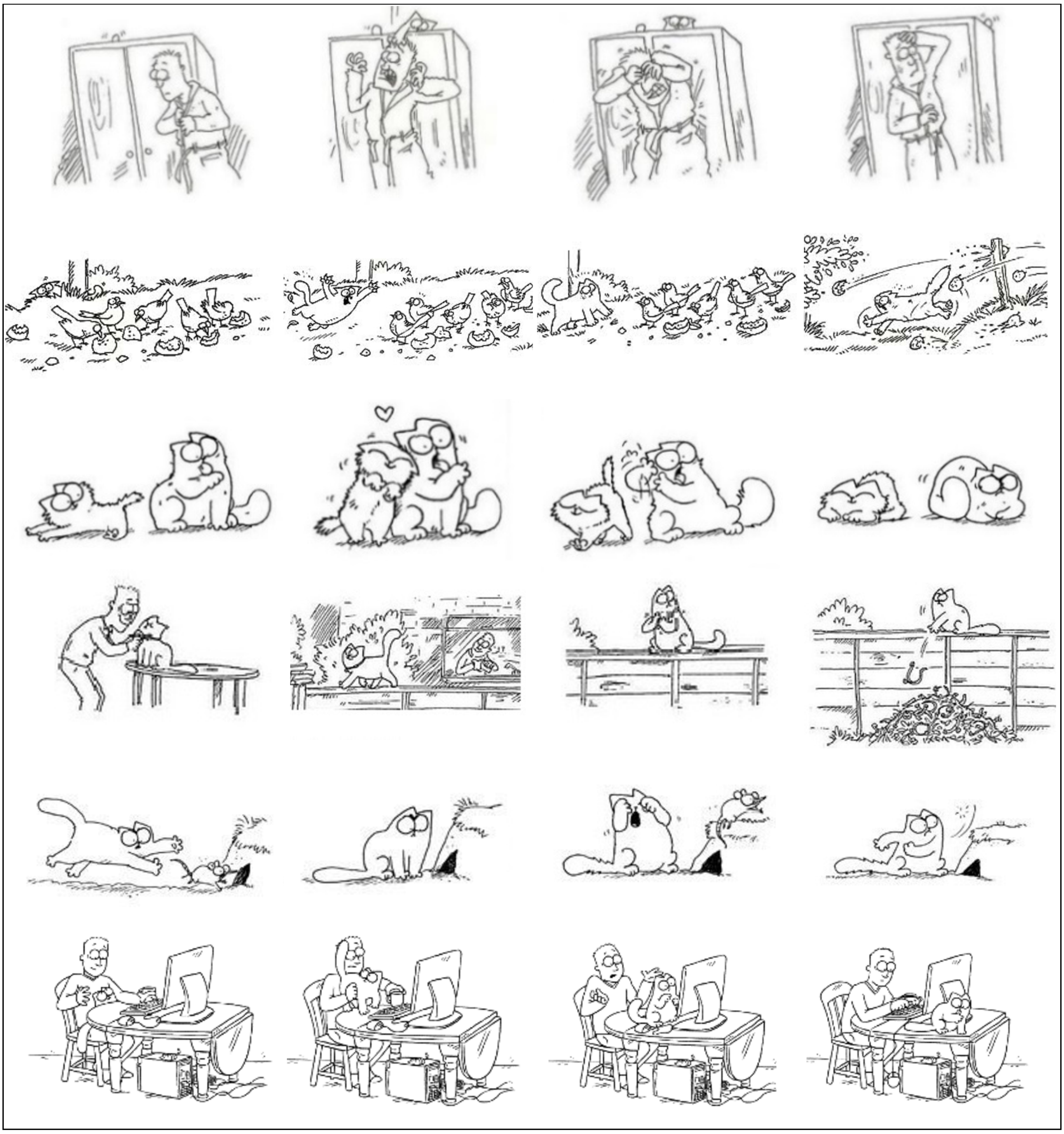


Non-restraining gestures / Restraining gestures



Restraining gestures / Non-restraining gestures

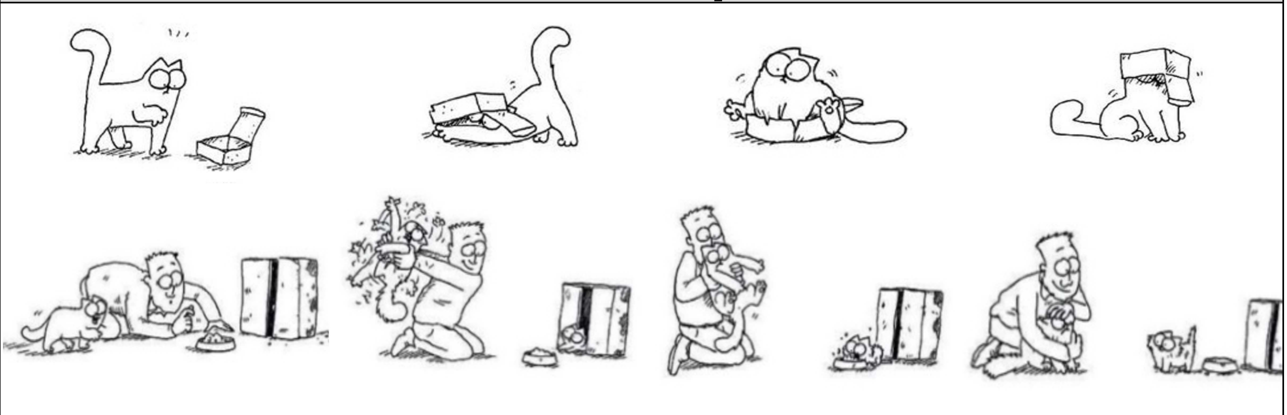




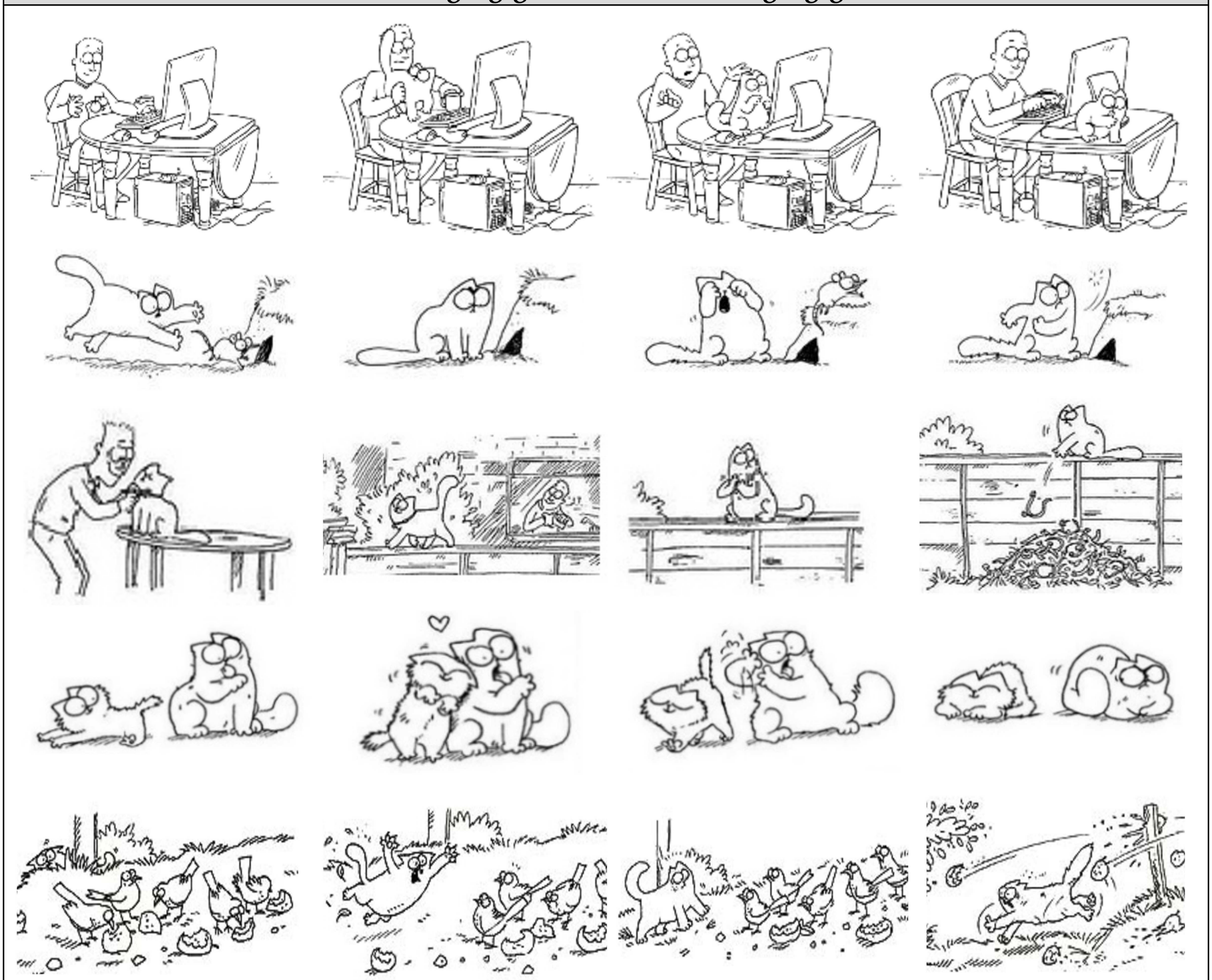
Note: **N** (Non-Restraining condition), **R** (Restraining condition); Half of the participants explained half of the comic strips in the N condition, while the same comic strips were explained in the R condition by the other half of the participants. Condition order was kept the same (N, R) between the two groups of participants.

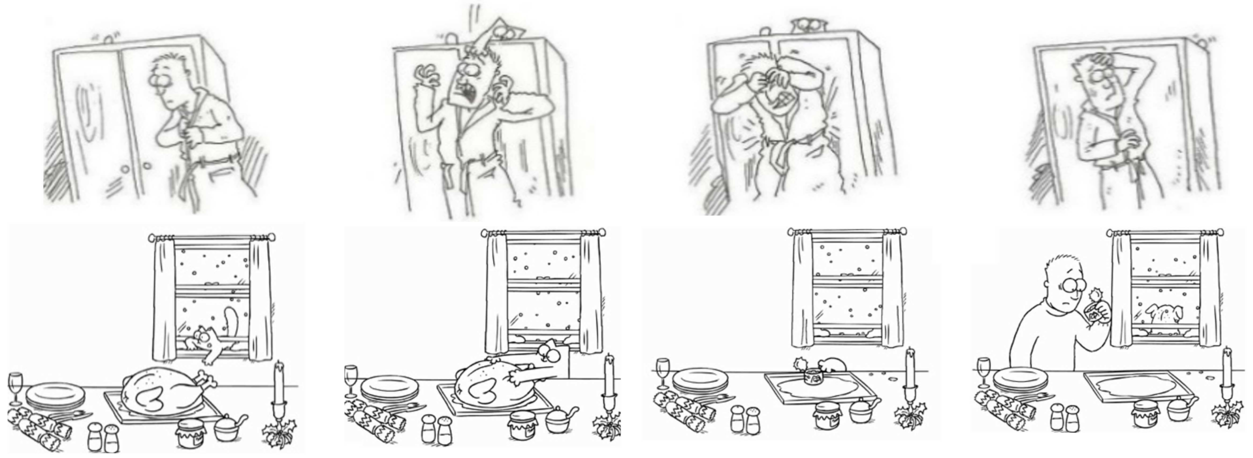
Study 2: Encouraging Gestures

Familiarization phase

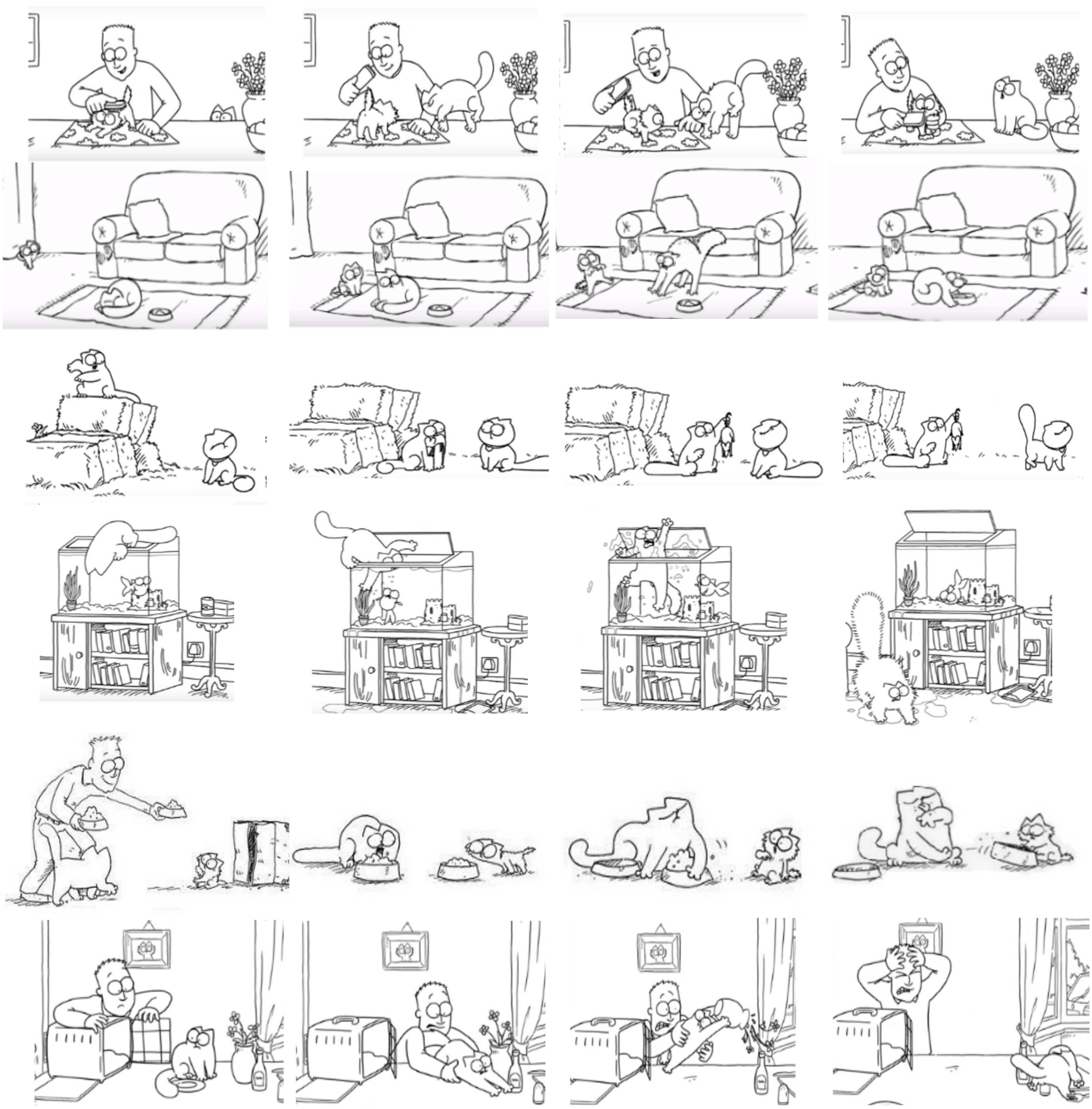


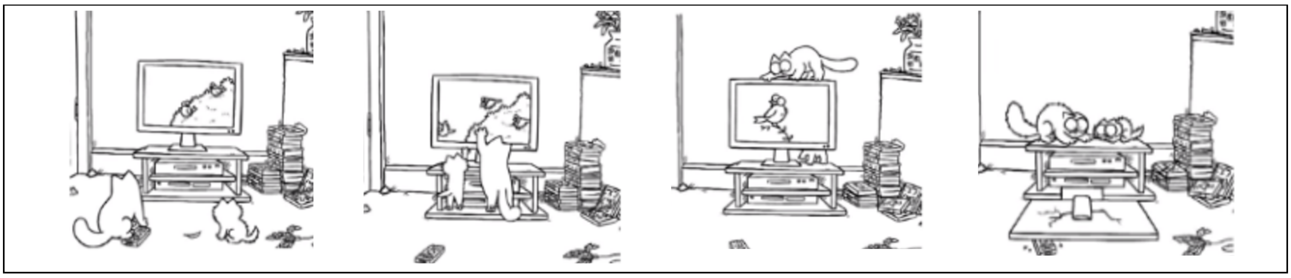
Non-encouraging gestures / Encouraging gestures





Encouraging gestures / Non-encouraging gestures





Note: **N** (Non-Encouraging condition), **E** (Encouraging condition); Half of the participants explained half of the comic strips in the N condition, while the same comic strips were explained in the E condition by the other half of the participants. Condition order was kept the same (N, E) between the two groups of participants.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor Maria Grazia Busà for her presence, support and guidance through the good and bad moments of my PhD. Thank you, Grazia, for these years together at the Language and Communication Lab (LCL). I've learned so much along the way, and this wouldn't have been possible without you.

I am also grateful to my co-supervisor, Pilar Prieto, whose support, knowledge, and enduring optimism (together with her cheering-up emails) were sometimes all that kept me going. Thank you, Pilar, for having believed in the project (and in me) from the beginning and for your prompt and detailed feedback on my thousands of emails and pieces of writings. Last but not least, thank you for your warm welcome in Barcelona that made me feel part of the GrEP research group from the very first day.

I would also like to thank the dissertation reviewers, Núria Esteve-Gibert and Maria Graziano, for their dedication to reading this work and for their precious comments and observations.

I am grateful to the Department of Linguistic and Literary Sciences (DiSLL) of the University of Padova; the DiSLL's director Anna Bettoni and the PhD program's directors Rocco Coronato and Annalisa Oboe. I would also like to thank the DiSLL's technicians, with special mentions to Paolo Scanferlato.

I am particularly grateful to Maria Teresa Musacchio for her help in finding participants for the experiments among her students. And I would like to extend my gratitude to all other people that participated in the experiments by getting just some fun comics experience and a breakfast (thanks to the *Bar Boh* or Boh Café for the partnership). My heartfelt thanks go to Nicolò "The Listener" Marcato for the great assistance with the piloting of the experiments and with the final data collection and for having shown the greatest endurance in listening to the same stories all over again pretending not to know what they were about. Thank you so much! I wouldn't have done this without you.

I wish to extend my thanks to my lovely PhD colleagues in Padova: Diana Antonello, Pietro Bertocchini, Teresa Cancro, Camilla Covazzi, Chiara De Paoli, Giulia Nalesso, Olga Retziou, Francesco Roncen, Valeria Russo, Giovanna Todaro, Raffaele Tondini, Angelica Vedelago, Susanna Zellini. Our WhatsApp group itself has been a wonderful cooperating experience providing support and fun. I wish we went out for drinks more often, though! I'm also happy to have shared this experience with Virginia Zorzi and Dario del Fante (I wish you the best of luck!). Thanks to Simone Gastaldon for meeting here and there and for sharing ideas, help and papers. And also for the statistics meetings! My thanks also go to Page Piccinini. I never met her in person but her online Statistics & R course was very helpful along the way.

I wish to express my gratitude now to various people that I was lucky enough to meet over the years and that inspired me so much. Emanuela Campisi and Asli Özyürek for the best first glance of what studying multimodal language is all about; Paolo Mairano that I need to thank also for the interesting and relevant references he had always ready for me and for spot advice on audio recording and acoustic analysis; Antonio Romano, for his inspiring seminars and advice on voice acoustics.

Human connections have been very important to me and they helped me keep going over the years with enthusiasm and motivation. So I want to list here some great colleagues that I have met at conferences that made them so special and exciting (pretty much in order of appearance): Bjørn Wessel-Tolvig, Manon Lelandais, Vito Evola, Marieke Hoetjes, Valentina De Iacovo, Mike Stevens, Matt Stave, Morgana Proietti, Kim Ouwehand, Wim Pouw, James Trujillo, Lieke Van Maastricht. Also, I feel blessed to have met Anita Slonimska and Luigi Palumbo: it was great to share with you the old times in Nijmegen (back in 2014) and the first approach to gesture studies.

I wish to send my *grazie* to Naoto Ienaga for the nice moments we shared during his research visit in Padova. And to Andrea Liotto: it was nice to have you with us at LCL. I also want to express my gratitude to Federica Cavicchio: your advice on experiment design, statistics, R and all sorts of things really made me "rolling more smoothly" towards the end of all this. Our chats eating *tramezzini al Nazionale* or *biscottini vegani dalle ragazze* are among the best memories of the past years at LCL.

Also, I wish to thank the amazing people of the GrEP group (Grup d'Estudis de Prosòdia) in Barcelona: Pilar, Florence Baills, Joan Borràs-Comes, Iris Hübscher, Evi Kiagia, Olga Kushch, Judith Llanes-Coromina, Mariia (Marusia) Pronina, Patrick Rohrer, Ingrid Vilà-Giménez, Laura Vincze, Yuan Zhang. I'm very grateful to all of you for always letting me feeling part of the GrEP, even from far away. It meant a lot to me and it has been a great source of motivation and enthusiasm (together with all the fun and lovely moments we shared). Ps. Judith & Ingrid you have made your best to make me missing Catalunya - and its food - so badly, but I'm grateful anyway. Also, I wish to thank Ionut-Teodor Sorodoc and Laura Aina from the GLiF for the great coffee/cigarette breaks at UPF (and our tapas nights too).

To the dearest friends and LCL colleagues in Padova, Tommaso Balsemin, Sara Brugnerotto, Isabella Matticchio goes my deepest gratitude. It took a while (Sara?) but, after that little while, I realized I couldn't have wished for better colleagues and friends than the three of you! You have been a mix of friendship and support service that was mostly achieved via drinks, food and endless laughs (or audio messages). In short, thank you Tb, Sara, Isa (and Paola!) for being always there in the last three years.

And I'll switch to Italian now, to thank all the people who have known me for the better part of my life and have made it somewhat easier to get to this point today.

Un grazie speciale ad Alice Spadaro, Salvo Sapienza/Arkee e Marco Costa per esservi fatti trovare a Barcelona sul più bello durante il mio dottorato (e per mille altre cose che già sapete). Fab, Cioffee, Tosa e Chiara perché sapete tutto dal principio e però rimanete *amichetti*, nonostante tutto. Cioffee merita anche dei ringraziamenti per alcune cose più specifiche tra le quali (1) le consulenze sull'Inglese ottenute *for free* sia per articoli in fase di stesura che per la tesi stessa, (2) la condivisione di dissidi sul piano vita.

Sono immensamente grata alla mia famiglia che è abbastanza grande da farmi sentire sommersa di affetto. A Mamma e Papà (Egle e Rino) per essere la mia personale fonte inesauribile di amore, sostegno e comprensione (chi sa quante cose mi sarei *preclusa* senza il vostro supporto e incoraggiamento). A Puque che fa sempre il tifo per me, come una mamma (e a Luca che fa bene a fare sempre il tifo per lei). Allo zio Marco che compare, quando può, nei momenti chiave. Grazie alla zia Cume per essere presente da sempre con dolcezza e indubbia comicità. Sono grata a Sasy e Piciu per il loro essere nonni sempre a tempo pieno anche nella distanza. Grazie a Rosanna, Renzo e Annica per esserci da otto anni come una famiglia da cui tornare. E poi volevo dire grazie a Sandro. Ti ho trovato 8 anni fa poggiato sul pavimento di un'aula come tu avresti trovato monete, banconote e quadrifogli. Sei la mia fortuna più grande.