Speech Communication 135 (2021) 25–36

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Exploring the effects of restraining the use of gestures on narrative speech

Alice Cravotta[a,1,*], Pilar Prieto[b,c,1], M. Grazia Busà[a,1]

[a] Dipartimento di Studi Linguistici e Letterari (DiSLL), Università degli Studi di Padova, Italy
[b] Institució Catalana de Recerca i Estudis Avançats, ICREA, Barcelona, Catalunya, Spain
[c] Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra, Barcelona, Catalunya, Spain

ARTICLE INFO

Keywords:
Gesture
Fluency
Disfluency
Narrative speech
Speech
Hand gestures

ABSTRACT

Research on gesture production has emphasized the strong relationship between speech and gesture. Studies have explored whether the inability to gesture is detrimental to speech at different levels. However, findings are still inconclusive and research that focuses on a complete set of acoustic prosodic measures, including F0 and intensity are lacking. Also, studies have used very controlled tasks but evidence is lacking about spontaneous speech. The present study investigates the effects of restraining hand gestures on semi-spontaneous 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 speech variables were examined: speech length (number of words and speech length in seconds), disfluencies (filled pauses, self-corrections, repetitions, insertions, interruptions) and prosodic properties related to speech rate, F0 and intensity. Overall, the results showed that speakers' inability to gesture does not significantly affect their narrative speech performance in terms of speech length, fluency and acoustic features. When repeating the analysis with the exclusion of the participants who gestured very little in the N condition, however, a slightly different pattern of results emerged; this leaves open the possibility that the inability to gesture may impact individuals differently, depending on the extent to which they rely on gesture when speaking. Further work is needed to shed more light on the role of gestures on the prosodic level of speech production which also takes into account individuals' communicative and cognitive inclinations.

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, and Chu, 2017; Interface Model, Kita and Özyürek, 2003), facilitate lexical access (Krauss, Chen, and 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 and Alibali, 2018), and reduce cognitive load (Cook, Yip, and Goldin-Meadow, 2012; Goldin-Meadow, Nusbaum, Kelly, and Wagner, 2001; Ping and 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., Esteve-Gibert, Borràs-Comes, Asor, Swerts, and Prieto, 2017; Esteve-Gibert and Prieto, 2013; Loehr, 2012; Shattuck-Hufnagel, Yasinnik, Veilleux, and Renwick, 2007; among many others). 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, Tiede, and Tyrone, 2017). In fact, there is evidence that language and action are closely related on a

---

* Corresponding author at: Via E. Vendramini, 13, 35137 Padova, Italy.
E-mail addresses: alice.cravotta@phd.unipd.it (A. Cravotta), pilar.prieto@upf.edu (P. Prieto), mariagrazia.busa@unipd.it (M.G. Busà).

[1] Authors are funded by: University of Padova Doctoral School of Linguistic, Philological and Literary Sciences; Erasmus Plus Inter-Institutional Mobility Agreement between the University of Padova and Universitat Pompeu Fabra; Spanish Ministry of Science and Innovation (Grant PGC2018-097007-B-I00); grant awarded by the Generalitat de Catalunya (Grant 2017 SGR-971) to the Prosodic Studies Group, Universitat Pompeu Fabra.

https://doi.org/10.1016/j.specom.2021.09.005
Received 25 July 2019; Received in revised form 30 July 2021; Accepted 17 September 2021

motoric level and co-speech gesture production engages brain areas that are functionally connected to Broca's area (Marstaller and Burianová, 2015; see Gentilucci and Dalla Volta, 2008 for a review). [For a general overview on the interaction between gestures and speech, see also Wagner, Malisz, and Kopp (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 above and to further explore the self-directed role of gesture production and its connection with 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, on the one hand, we aim to examine the impact of gesture restriction on narrative speech, with respect to fluency and speech length; on the other hand, the study also explores how the inability to gesture can impact F0 and intensity measures with the purpose to shed more light into how gesture production interacts with speech production at the prosodic level.

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); with the exception of one study (Hoetjes, Krahmer, and Swerts, 2014), the effects on acoustic features of speech such as F0 and intensity have been overlooked (see next section). 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, and 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[2]. 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. 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 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 and Indefrey, 2014; Indefrey, 2011; Kearney and 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, while describing 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 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. 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, while sitting on their hands for half of the experiment (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. Pine, Bird and Kirk (2007) also found that, in a picture naming task, when children were free to gesture, they named more words correctly and resolved successfully more Tip-of-the-Tongue states (TOTs)[3]. However, it should be noted that, when they were restricted from gesturing, children did not experience more TOTs than when they were free to move. Conversely, Beattie and Coughlan (1999) found that restraining gesture use does not affect word recall (i.e., retrieving rare high imageability lexical items from definitions) and that participants with their arms folded had in general a more fluent retrieval process than participants that were free to gesture. 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 content and speech length, 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, and Kelly, 2017; Kita et al., 2017). As for speech content, Rimé, Schiaratura, Hupet and Ghysselinckx (1984) found that restraining speakers' gestures in 50-minutes 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, and Wang (2019) found that gesture restriction does not impact language use in math explanations. Transcriptions of the speech produced by participants assessing whether 8 geometry conjectures were true or false, and why, were analyzed through dictionary-based text analysis tools (*Coh-metrix*, McNamara, Louwerse, Cai, and Graesser, (2013); *LIWC*, Pennebaker, Booth, Boyd, and 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 the 148 different language measures, including kind and number of

---

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

[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).

Table 1
Summary of previous findings on the effects of restraining gestures on speech (from more to less recent).

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
Jenkins, Coppola & Coelho (2017)	-	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:** Within-subjects
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, 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 =worse 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
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). EXP1 and EXP 2 have similar results.		**T:** Retrieval of lexical items from definition. **L:** English. **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.*

words used and other speech patterns.

Finlayson et al. (2003) reported a higher number of words used and higher number of spatial content phrases in the speech produced in a story retelling task in the gesture-restricted condition than in the hands-free condition; an increase in the number of words used for expressing spatial content was also 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.

Focusing on the semantic level of speech, 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 gesture restriction did not impact the duration of routes descriptions on a map 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.

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 study that investigated whether speech becomes more monotonous (in terms of pitch range) when speakers cannot gesture. The study found that, in motor tasks descriptions, 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 et al. (2019) showed that gestures can actually be "heard" in the moment they occur during vocalizations, as forces reverberate to the lower vocal tract, affecting aspects of prosodic control (specifically, listeners succeeded in synchronizing their own wrist or arm movement with what they perceived in the voice acoustics to be the movement of the speaker).

This line of research, exploring the connection between gesture production and prosodic modulation, is linked with the evidence that spoken language and arm gestures are controlled by the same motor control system, as suggested by both behavioral and neuroimaging studies (see Marstaller and Burianová (2015);Gentilucci and 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 and Dixon (2018) 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. A study by Cravotta, Busà and Prieto (2019), where speakers were encouraged to use their hands while telling a set of stories, showed that the speech produced in the gesture-encouraged condition had overall higher F0 maximum metric and intensity features as compared to the non-encouraged condition. This finding from a rather naturalistic setting (semi-spontaneous storytelling) supports the idea that gesture production can have a direct influence on speech prosody.

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

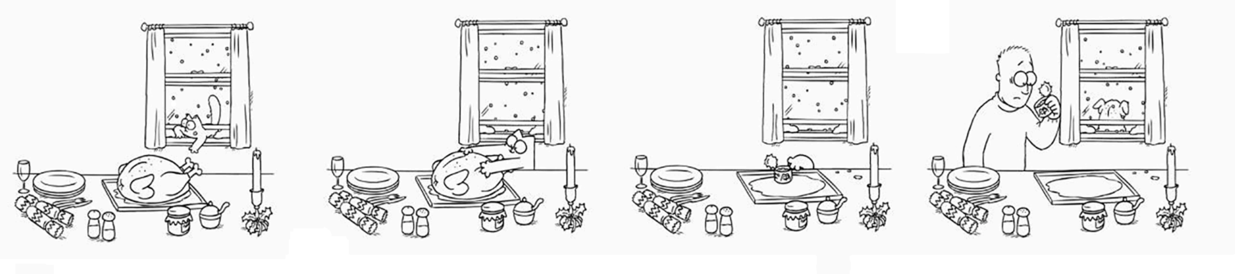

Fig. 1. Example of a 4-scene comic strip used for the experiment (from Simon's Cat by Simon Tofield, reproduced with permission).

on fluency have been found in visual objects description (Morsella and Krauss, 2004) and low codability abstract lines drawings descriptions (Graham and 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., 2017; 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 aims to investigate how the inability to use gesture impacts on speech production as it happens in its full functioning (in semi-spontaneous narrative speech). The study focuses on assessing the main prosodic 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 acoustics, which, to our knowledge, has only been explored by Hoetjes et al. (2014). Based on the main theoretical accounts described above, we expect to find that the inability to gesture negatively affects speech fluency and may also result in a reduction of the acoustic features of speech (i.e., lower F0 and intensity measures).

Method

The present study used a narration task in which participants had to watch and describe a set of comic strips in two different within-subject conditions: Non-restraining gesture condition (N) in which the participants were free to gesture when narrating; and Restraining gesture condition (R) in which the 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 and Hall, 1995; Chu, Meyer, Foulkes, and Kita, 2014; Goksun, Goldin-Meadow, Newcombe, and Shipley, 2013; Hostetter and Hopkins, 2002; Hostetter and Potthoff, 2012; Kita, 2009; Nicoladis, Nagpal, Marentette, and Hauer, 2018; O'Carroll, Nicoladis, and Smithson, 2015).

The prosodic analysis of the target stories will focus on speech 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 an alumnus 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 and Hall, 1995; Hostetter and Hopkins, 2002).

Materials

Sixteen 4-scene comic strips adapted from Simon's Cat by Simon Tofield were used for the narration task (see Fig. 1 for an example). The comic strips were carefully selected 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 *Procedure* 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 head-mounted 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.

As shown in Fig. 2, the camera was set in front of the participant (at 2.50 m distance) recording her upper body and face. 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 (first author) sat at the participant's side for the entire experiment.

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 listener was a fellow participant who did not know the stories in advance was to avoid potential effects of common ground (Holler and 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 back-channel 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 and Garnham, 2007)). By contrast, to ensure that the interaction between participant and listener 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 participant told the story to the listener 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

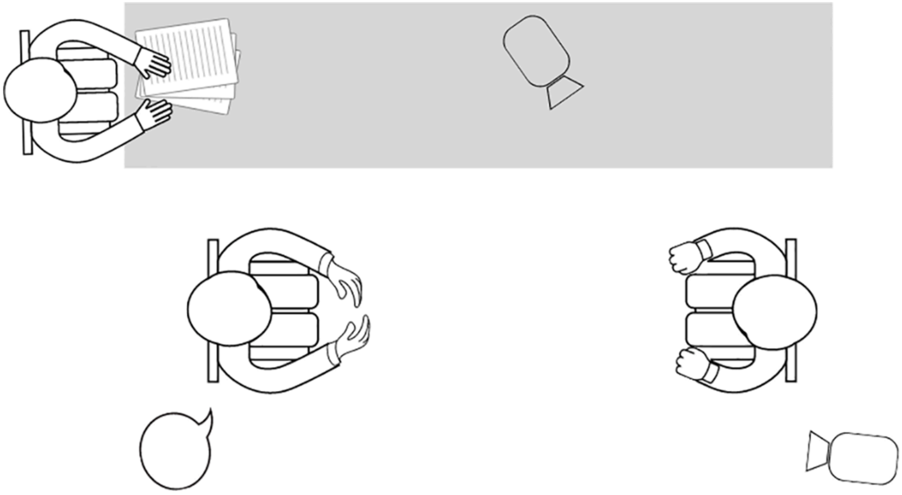

Fig. 2. Experimental set-up

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 and 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 an 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 their 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 coding and Acoustic analysis

Speech 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 final 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 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 and 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, and Schmid, 2015; Götz, 2013; Kormos, 2014, among others), instances of any of the following types of disfluencies were manually annotated by the first author (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 t-t-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 uccelli stanno mangiando delle briciole di pane per terra – un sacco di uccelli – a un certo punto, il gatto…", "Some birds are eating bread crumbs on the ground – a lot of birds – suddenly, the cat…");
- 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"). Interruptions can often precede a self-correction;

- 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…", "then, theee…cat");
- Silent pauses: annotated automatically by a Praat script ((De Jong and 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 and 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 and Michel, 1968; McGlone, 1967; Murry, 1971) with a mean of approximately 50Hz (as reported in the literature review provided in Blomgren, Chen, Ng, and 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.

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, and Walker, 2014). Each model included one of the 19 dependent variables (see Table 2) 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 and Hochberg, 1995)[5]. The results are reported with the adjusted False Discovery Rate (FDR) critical values.

In order to assess how different individuals compensate for the inability to gesture, the same analysis was performed on the full data set and on a subset of the data. The subset excludes the participants who did not gesture (or barely gestured) in the Non-restraining gesture condition (N). We hypothesize that participants who did not gesture at all (or gestured very little) in the N condition, may not rely on gestures as much as those who normally gesture more when speaking, that is, the group of gesturers may be more affected by the inability to gesture. To build the subset, we have computed the average gesture rate of the stories told in the N condition for each of the 20 participants. We have then computed the quartiles and have excluded the three participants whose average gesture rate fell in the lower quartile of the distribution (Q1=11.52 gestures per 100 words). Fig. 3 shows the gesture rate (gesture per 100 words) per participant in the N condition. The subset does not include participants 18, 17, 15.

Results of the analyses performed on the full data set and on the subset are described separately in the next section.

Results

Full data set

Table 2 shows the main descriptive statistics of each of the 19 variables separated by Condition (N, R). The table also includes the average gesture rate (number of gestures per 100 words) in the non-restraining gesture (N) condition.

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

Table 2

Main descriptive statistics (full data set). Non-restraining gesture condition (N), Restraining gesture condition (R).

	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
Gesture rate (per 100 words)	13.48	-	6.39	-

---

[4] Script written by Jonas Lindh and available at https://github.com/YoeriNijs/PraatPitch

[5] False Discovery Rate (FDR; Benjamini & Hochberg, 1995) was found to be a good compromise solution among different multiple testing correction procedures because it minimizes the type II error and at the same time it keeps the type I error under control (Pastore, Nucci, & Galfano, 2005)).

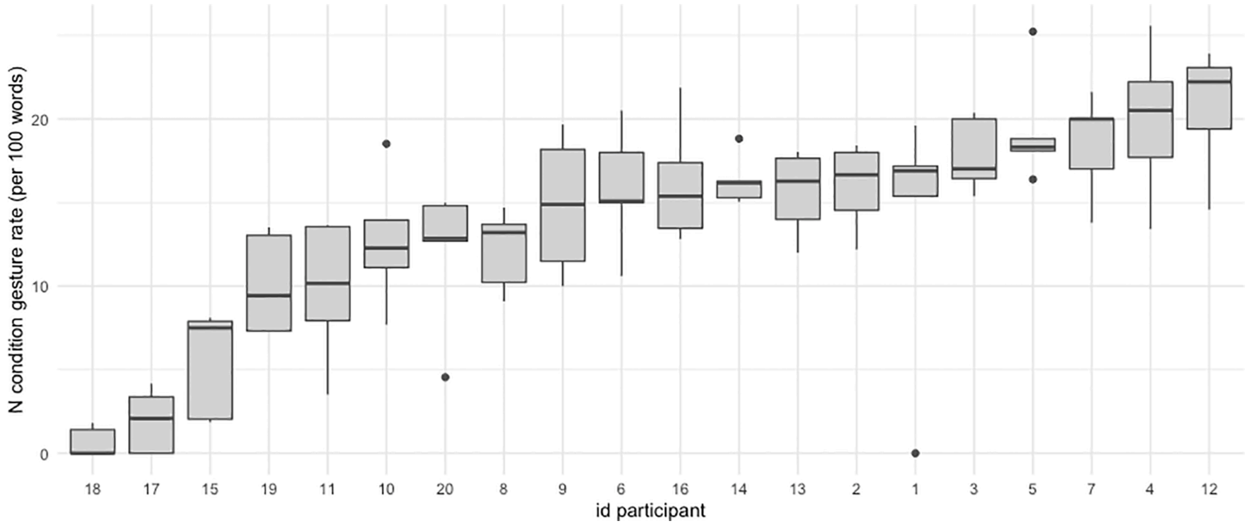

Fig. 3. Box plots reporting participants' gesture rate (per 100 words) in the stories told in the N condition (Q1=11.52, Q2=15.71, Q3=17.41, Q4=20.64)

other words, stories told when participants were free to gesture were not found to be longer in terms of duration (s) (est.= 1.277, S.E.= 0.667, χ^2=3.645) and did not change in terms of number of words (est.= 2.270, S.E.= 2.096, χ^2=1.1752); Also, none of the disfluency rates or speech rate (est.= -0.08, S.E.= 0.044, χ^2=3.323) significantly changed between the two conditions. As for F0 and intensity, speech was not affected by gesture restriction.

Data subset

Results of the analyses performed on the data subset are reported below in Table 4.

As shown in Table 4, none of the variables were significantly different across the N and R conditions for FDR corrected values. It is worth indicating that Story duration (s), F0 mean (Hz), Intensity min (dB) were significant for alpha =.05 before FDR correction. Stories told when the participants were unable to gesture were slightly longer in terms of duration (in seconds) (est.= 2.0198, S.E.= 0.6752, χ^2=8.7523, $p < .05$). However, narratives did not change in terms of the number of words used and none of the disfluency or speech rate variables changed significantly between the two conditions. As for the acoustic measures, F0 mean (Hz) (est.= -2.0742, S.E.=0.9079, χ^2= 5.153, $p < .05$) and Intensity min (dB) (est.= -0.2401, S.E.=0.1159, χ^2=4.2585, $p < .05$) resulted different between the two conditions only before FDR correction. Even though the numeric difference is modest, the direction of the effects is in line with our preliminary expectations. The FDR correction procedure was meant to minimize and keep under control both type I and type II errors. However, the fact that a significant effect for 3 out of 19 variables of interest was observed before correction means that a robust null effect is not found for this data subset. A discussion of the outcome of this analysis is proposed below.

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 by using a novel

Table 3
Results of the LMEMs per dependent variable (full data set).

Variable	Estimates	SE	Lower CI	Upper CI	t	χ^2	p	FDR
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. CI, Confidence interval: Lower 2.5%; Upper 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).

Table 4
Results of the LMEMs per dependent variable (Subset of the data)

Variable	Estimates	SE	Lower CI	Upper CI	t	χ^2	p	FDR
Story duration (s)	2.02	0.675	0.693	3.349	2.991	8.752	0.003	0.0026
n. of words	3.29	2.281	-1.194	7.78	1.442	2.08	0.149	0.0158
Repetitions	0.105	0.246	-0.379	0.588	0.427	0.183	0.669	0.0421
Insertions	-0.136	0.14	-0.411	0.138	-0.975	0.954	0.329	0.0289
Interruptions	0.322	0.208	-0.087	0.731	1.550	2.395	0.122	0.0132
Self-corrections	0.075	0.243	-0.404	0.552	0.309	0.094	0.759	0.0447
Filled pauses	0.557	0.45	-0.328	1.442	1.238	1.534	0.216	0.0211
Silent Pauses	0.451	0.408	-0.351	1.253	1.106	1.225	0.268	0.0237
Disfluencies (tot)	0.918	0.692	-0.442	2.277	1.326	1.761	0.185	0.0184
Speech rate (syll/dur)	0.077	0.048	-0.171	0.017	-1.604	2.563	0.109	0.0105
F0 min (Hz)	0.741	1.497	-2.202	3.685	0.495	0.246	0.62	0.0368
F0 max (Hz)	9.00	10.58	-11.79	29.79	0.851	0.727	0.394	0.0342
F0 mean (Hz)	-2.074	0.908	-3.859	-0.288	-2.285	5.153	0.023	0.0053
F0 var. (Hz)	-0.791	0.836	-2.436	0.851	-0.947	0.90	0.343	0.0316
PVQ	-0.001	0.004	-0.010	0.006	-0.427	0.183	0.669	0.0395
Intensity min (dB)	-0.240	0.116	-0.468	-0.012	-2.071	4.259	0.039	0.0079
Intensity max (dB)	0.055	0.377	-0.685	0.796	0.145	0.021	0.884	0.0474
Intensity mean (dB)	-0.166	0.169	-0.498	0.167	-0.978	0.959	0.327	0.0263
Intensity var. (dB)	0.009	0.108	-0.204	0.221	0.082	0.007	0.934	0.0500

Note: N. of obs: 200; Groups: participants, 20 | Story, 10. CI, Confidence interval: Lower 2.5%; Upper 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).

experimental setting and design. An experiment was set up to elicit spontaneous storytelling 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 considers a comprehensive set of measures related to temporal narrative features, fluency measures, and also focuses on acoustic measures related to pitch and intensity. The results show no significant effects of restraining the use of gestures on speech features. It is difficult to draw firm conclusions from null-results, as 20 participants took part in the study and each participant only completed 5 trials under each condition. We cannot exclude that gesture restriction has any effects on speech, but neither can we confirm that it does. However, our results give us the opportunity to discuss and elaborate on a few aspects that we believe are of key importance and touch upon the bigger question of the self-directed role of gesture for speech production. These results are commented separately for fluency, speech 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 (see Table 1). However, our study does not provide evidence that gesture restriction has any detrimental effects on fluency. There are possible explanations for these results. First, it should be acknowledged that speech fluency can be determined by individual characteristics (e.g., short-term memory) and personality traits (e.g., extraversion, resistance to stress) that were not directly considered. For example, it has been shown that open, extraverted, and emotionally stable participants score better in verbal fluency tests (Sutin et al., 2011), and, in the case of bilinguals, high extraverted individuals have naturally more fluent L2 speech than low extraverted individuals (Dewaele and Furnham, 2000). Moreover, gesture production itself and its role in speech production largely depend on individuals’ cognitive abilities, verbal skills, personality traits, cultural, linguistic and gender differences and even experience in learning other languages (Briton and Hall, 1995; Chu et al., 2014; Gillespie, James, Federmeier, and Watson, 2014; Göksun et al., 2013; Hostetter and Hopkins, 2002; Hostetter and Potthoff, 2012; Kita, 2009; Nicoladis et al., 2018; O’Carroll et al., 2015; So, 2010). Thus, for instance, it would not be surprising if different individuals relied on gestures for linguistic fluency to different extents, and that some speakers might be naturally more affected than others when being unable to gesture. In the context of the GSA framework (Hostetter and Alibali, 2019), which predicts that every speaker has a variable threshold that reflects her own *resistance* to actually producing a gesture, speakers with a low gesture threshold may be particularly affected by the inability to gesture. In fact, the individual gesture threshold can depend on a number of factors (e.g., individual cultural taboos, social conventions, and beliefs about the appropriateness of making gestures in a certain context), but also on contextual situations (e.g., cognitive load, communicative functions, task difficulty). We have explored this aspect by analyzing a subset of the data that excludes participants (n=3) who gestured to a little extent in the N condition. The results leave open the possibility that the inability to gesture negatively affects speech in speakers who rely more on gesture in the process of speaking.

In general, while we controlled as much as possible for task difficulty, communicative situation, and linguistic and gender differences, we think that future comparable studies should also explore how individual speakers are able to use gestures as a compensation tool during communication and to what extent this depends on individuals’ cognitive dispositions (for a review see Özer and Göksun, 2020). Also, to further explore the impact of being unable to gesture on speech fluency, a more detailed analysis could be carried out on the data presented in this study (or similarly collected data) to explore whether spatial and motor content expressions are the most affected, in terms of fluency, by the inability to gesture, as found in Rauscher et al. (1996)). 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.

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 and Casey, 2001; Melinger and 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 on gesture restriction in semi-spontaneous speech could analyze content differences between speech produced in N vs R conditions, with a focus on the semantic richness of spatial content phrases (see Hostetter et al., 2007) or, more generally, on the amount of spatio-motor information and imagery content expressed in speech. Also, a quantitative analysis of the narratives' vocabulary richness together with a detailed analysis of the speech chunking strategies could be a viable way to further explore the issue.

As far as the acoustic analysis, we explored whether restraining gesture has any effects on acoustic properties such as fundamental frequency and intensity. We expected that the inability to gesture would have reduced F0 and intensity measures. Our results do not provide evidence for it and are consistent with Hoetjes et al. (2014). This seems to be in partial contradiction with evidence coming from previous investigations on speech prosody and gesture production. These suggest that encouraging the use of gestures has an effect on speech acoustics, with reference to F0, intensity and spectral properties (Cravotta, Busà and Prieto, 2019; Krahmer and Swerts, 2007; Pouw, Harrison, and Dixon, 2018). 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 and Alibali, 2019; Marstaller and 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 instructions to gesture as a request to enact more, or speak in a clearer way, louder, or perhaps slower, etc.). With that said, the results of our study do not allow us to exclude the possibility that gesture restriction affects speech acoustics features of speech. It may be that our design, based on semi-spontaneous speech, was not able to capture subtle acoustic differences when compared to, e.g., Krahmer and Swerts (2007) and Pouw et al. (2018)'s very controlled settings. However, we believe that our study can inform future studies on the intricacies of investigating speech acoustic correlates in a naturalistic design, and it can also be a good starting point for evaluating the power of future studies based on pre-existing parameter estimates.

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)). A motion capture study by Serré, Dohen, Susanne, and Rochet-Capellan, (2020) in which participants had to recall stories in different movement conditions with hands being either free, blocked or constrained, has shown that head motion may actually be inhibited when the hands are prevented from gesturing leading to an overall stiffening of the upper body. Motion capture may be a good way to control for this aspect, as in our data 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 physically restraining speakers from using their hands can actually inhibit the effects/role that gestures have on speech production.

To conclude, the results of our study are not in contradiction with the idea that gesture has a self-directed role for speech production. While we show that restraining speakers from gesturing may not affect their narrative speech performance in terms of fluency and acoustic features, we cannot exclude that this may be the case for individuals that naturally gesture more. As our analysis on the subset of the data suggests, the impact of the inability to gesture on speech may depend on the extent to which single individuals rely on gesture in the process of speaking and how they are able to compensate for the inability to gesture with their individual cognitive dispositions (see Özer and Göksun, 2020). We believe that the issue merits further investigation in order to shed more light into the role of gesture at the prosodic level of speech production, and also, to further explore the self-directed role of gesture in speech planning and production for individuals with different communicative and cognitive inclinations.

Author statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they haveparticipated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript.

Declaration of Competing Interest

In reference to the manuscript "Exploring the Effects of Restraining the Use of Gestures on Narrative Speech" submitted for consideration as a research article in the journal *Speech Communication,* the authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

We are grateful to Nicolò Marcato for helping with the recordings and also serving as the "fellow participant" during the experiment. We are also grateful to the reviewers for their insightful comments, and to all the colleagues who gave valuable feedback at different international meetings. We also thank the Simon's Cat team for allowing us use of their comic strips. This collaborative research was the result of an Erasmus Plus Inter- Institutional Mobility Agreement between the University of Padova and the Pompeu Fabra University. The first author has received support by a grant from the University of Padova's Doctoral School of Linguistic, Philological and Literary Sciences; the third author also acknowledges Grant PGC2018-097007-B-I00 from the Spanish Ministry of Science and Innovation and Grant 2017 SGR _ 971 from the Generalitat de Catalunya to the Prosodic Studies Group.

References

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

Beattie, G., Coughlan, J., 1999. An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. Br. J. Psychol. 90, 35–56.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. Royal Statistical Society: Series B (Methodological) 57 (1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bergmann, C., Sprenger, S.A., Schmid, M.S., 2015. The impact of language co-activation on L1 and L2 speech fluency. Acta Psychol. (Amst) 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. J. Acoust. Soc. Am. 103 (5), 2649–2658. https://doi.org/10.1121/1.422785.

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.

Brown, A.S., 1991. A review of the tip-of-the-tongue experience. Psychol. Bull. 109 (2), 204–223.

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. J. Experiment. Psychol.: General 143 (2), 694–709.

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

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.

Cravotta, A., Busà, M.G., Prieto, P., 2019. Effects of encouraging the use of gestures on speech. J. Speech Language and Hearing Res. 62 (9), 3204–3219. https://doi.org/10.1044/2019_JSLHR-S-18-0493.

De Jong, N.H., Wempe, T., 2009. Praat script to detect syllable nuclei and measure speech rate automatically. Behav. Res. 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). John Benjamins Publishing Company.

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, pp. 105–173.

Emmorey, K., Casey, S., 2001. Gesture, thought and spatial language. Gesture 1 (1), 35–50. https://doi.org/10.1075/gest.1.1.04emm.

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. J. Acoust. Soc. Am. 141 (6), 4727–4739. https://doi.org/10.1121/1.4986649.

Esteve-Gibert, N., Prieto, P., 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. J. Speech Lang. Hear. Res. 56 (3), 850–864. https://doi.org/10.1044/1092-4388(2012/12-0049.

Fillmore, C.J., Kempler, D., Wang, W.S.Y., 2014. Individual differences in language ability and language behavior. Academic Press, New York.

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

Frick-Horbury, D., Guttentag, R.E., 1998. The effects of restricting hand gesture production on lexical retrieval and free recall. Am. J. Psychol. 3 (1), 43–62.

Gentilucci, M., Dalla Volta, R., 2008. Spoken language and arm gestures are controlled by the same motor control system. Q. J. Exp. Psychol. 61 (6), 944–957.

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. https://doi.org/10.1016/J.COGNITION.2014.03.012.

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. https://doi.org/10.1007/s10339-013-0549-1.Individual.

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

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

Graham, J.A., Heywood, S., 1975. The effects of elimination of hand gestures and of verbal codability on speech performance. Eur. J. Soc. Psychol. 5 (2), 189–195.

Hagoort, P., Indefrey, P., 2014. The neurobiology of language beyond single words. Annu. Rev. Neurosci. 37, 347–362. https://doi.org/10.1146/annurev-neuro-071013-013847.

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 Commun. 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. https://doi.org/10.1080/01690960802095545.

Hollien, H., Michel, J.F., 1968. Vocal fry as a phonational register. J. Speech Hear. Res. 11 (3), 600–604. https://doi.org/10.1044/jshr.1103.600.

Hostetter, A.B., Alibali, M.W., 2019. Gesture as simulated action: Revisiting the framework. Psychon. Bull. Rev. 26 (3), 721–752. https://doi.org/10.3758/s13423-018-1548-0.

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. In: Proceedings of the 29th Annual Meeting of the Cognitive Science Society, 29. Nashville, USA, pp. 1097–1102.

Hostetter, A.B., Hopkins, W.D., 2002. The effect of thought structure on the production of lexical movements. Brain Lang. 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. https://doi.org/10.1075/gest.12.1.04hos.

Indefrey, P., 2011. The spatial and temporal signatures of word production components: A critical update. Front. Psychol. 2 (255), 1–16. https://doi.org/10.3389/fpsyg.2011.00255.

Jacobs, N., Garnham, A., 2007. The role of conversational hand gestures in a narrative task. J. Memory and Language 56 (2), 291–303. https://doi.org/10.1016/j.jml.2006.07.011.

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 Neurosci. 34 (9), 1214–1229. https://doi.org/10.1080/23273798.2019.1589541.

Kita, S., 2009. Cross-cultural variation of speech-accompanying gesture: A review. Language and Cognitive Processes 24 (2), 145–167. https://doi.org/10.1080/01690960802586188.

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

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

Kita, Sotaro, 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. J. Memory and Language 48 (1), 16–32.

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

Krahmer, E., Swerts, M., 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. J. 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, David (Ed.), Language and Gesture. Cambridge University Press, Cambdridge, pp. 261–283.

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 Phonol. 8 (1), 1–26, 3.

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

Marstaller, L., Burianová, H., 2013. Individual differences in the gesture effect on working memory. Psychon. Bull. Rev. 20 (3), 496–500. https://doi.org/10.3758/s13423-012-0365-0.

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. J. Psycholinguist. Res. 27 (1), 69–89. https://doi.org/10.1023/A:1023274823974.

McGlone, R.E., 1967. Air flow during vocal fry phonation. J. Speech Hear. Res. 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/.

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

Morsella, E., Krauss, R.M., 2004. The role of gestures in spatial working memory and speech. Am. J. Psychol. 117 (3), 411–424.

Murry, T., 1971. Subglottal pressure and airflow measures during vocal fry phonation. J. Speech Hear. Res. 14 (3), 544–551. https://doi.org/10.1044/jshr.1403.544.

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. https://doi.org/10.1017/langcog.2018.25.

O'Carroll, S., Nicoladis, E., Smithson, L., 2015. The effect of extroversion on communication: Evidence from an interlocutor visibility manipulation. Speech Commun. 69, 1–8. https://doi.org/10.1016/J.SPECOM.2015.01.005.

Özer, D., Göksun, T., 2020. Gesture Use and Processing: A Review on Individual Differences in Cognitive Resources. Front. Psychol. 11, 1–15. https://doi.org/10.3389/fpsyg.2020.573555. November.

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. In: Proceedings of the 39th Annual Meeting of the Cognitive Science Society. London, UK, pp. 1471–1476.

Pastore, M., Nucci, M., Galfano, G., 2005. False Discovery Rate: applicazione di un metodo alternativo per i confronti multipli con misure ripetute. Giornale Italiano Di Psicologia, pp. 1–12. Retrieved from. http://www.mulino.it/rivisteweb/scheda_articolo.php?id_articolo=20642.

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. Dev. Sci. 10 (6), 747–754.

Ping, R., Goldin-Meadow, S., 2010. Gesturing saves cognitive resources when talking about nonpresent objects. Cogn. Sci. 34 (4), 602–619. https://doi.org/10.1016/j.pmrj.2014.02.014.Lumbar.

Pouw, W., Harrison, S.J., Dixon, J.A., 2018. Gesture-speech Physics: The biomechanical Basis of Gesture-Speech Synchrony. 1–46. https://doi.org/10.31234/osf.io/tgua4.

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.* https://doi.org/https://doi.org/10.31234/osf.io/ny39e.

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

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. Motiv. Emot. 8 (4), 311–325.

Serré, H., Dohen, M., Susanne, F., Rochet-Capellan, A., 2020. Speaking while moving: Does the head compensate for the hands not being able to move? In: Tiede, M., Whalen, D.H., Gracco, V. (Eds.), 12th International Seminar on Speech Production. IOS PRESS, pp. 226–229.

Shattuck-Hufnagel, S., Yasinnik, Y., Veilleux, N., & Renwick, M. (2007). A Method for Studying the Time Alignment of Gestures and Prosody in American English:Hits' and Pitch Accents in Academic-Lecture-Style Speech. In A. Esposito, M. Bratanic, E. Keller, & M. Marinaro (Eds.), *Fundamentals of verbal and nonverbal communication and the biometric issue* (Vol. 18, pp. 34–44). IOS PRESS.

So, C.W., 2010. Cross-cultural transfer in gesture frequency in Chinese-English bilinguals. Language and Cognitive Processes 10 (25), 1335–1353. https://doi.org/10.1080/01690961003694268.

Sutin, A.R., Terracciano, A., Kitner-Triolo, M.H., Uda, M., Schlessinger, D., Zonderman, A.B., 2011. Personality traits prospectively predict verbal fluency in a lifespan sample. Psychol. Aging 26 (4), 994–999. https://doi.org/10.1037/a0024276.

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

Walkington, C., Woods, D., Nathan, M.J., Chelule, G., Wang, M., 2019. Does restricting hand gestures impair mathematical reasoning? Learn. Instr. 64, 101225 https://doi.org/10.1016/j.learninstruc.2019.101225.

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