JSLHR

Research Article

Effects of Encouraging the Use of Gestures on Speech

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

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. This 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 fundamental frequency [F0] and intensity).

Method: Twenty 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 1 of the following 2 conditions: (a) nonencouraging condition (N), that is, no instructions about gesturing were given, and (b) encouraging condition (E), that is, 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) and (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.

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, Chen, & Gottesman, 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, 2018), 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., Esteve-Gibert, Borràs-Comes, Asor, Swerts, & Prieto, 2017; Esteve-Gibert & Prieto, 2013; Loehr, 2012; Shattuck-Hufnagel, Yasinnik, Veilleux, & Renwick, 2007, among many others). 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; e.g., Krivokapić, Tiede, & Tyrone, 2017; for a general overview on the synchronization between gestures and speech, see also Wagner, Malisz, & Kopp, 2014). Finally,

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

Correspondence to Alice Cravotta: alice.cravotta@phd.unipd.it

Editor-in-Chief: Julie Liss
Editor: Stephanie Borrie

Received December 12, 2018
Revision received April 18, 2019
Accepted May 14, 2019
https://doi.org/10.1044/2019_JSLHR-S-18-0493

Disclosure: The authors have declared that no competing interests existed at the time of publication.

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 co-occurring speech. There is some evidence that hand gestures can alter the prosodic and spectral properties of the speech they co-occur with. Krahmer and 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*, on *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 co-occurring 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 and 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 three common words (CIAO, NO, and STOP) and one pseudoword (LAO) in four gesture conditions (absent, present, meaningful/related, nonmeaningful gestures). They found that when the words were coproduced 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, and Dixon (2018) 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 and 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, and fluency. Hostetter, Alibali, and 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 spatiomotor events. Also, spontaneous conversations produced by speakers that were prevented from gesturing showed a general decrease in imagery content (Rimé, Schiaratura, Hupet, & Ghysselinckx, 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, Forrest, Lickley, & Beck, 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, Krahmer, & Swerts, 2014; Hostetter et al., 2007; Jenkins, Coppola, & Coelho, 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 et al., 2003; Graham & Heywood, 1975; Morsella & Krauss, 2004) and especially when describing spatial content (Rauscher, Krauss, & Chen, 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. 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 were 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.'s idea had been previously proposed by Dobrogaev (1929),[1] an early study that is often reported in the literature (e.g., Krauss et al., 2000; McClave, 1998; Rauscher et al., 1996; Wagner et al., 2014) but does not contain any quantitative analysis, nor does it provide specific details on the methodology used.

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 (i.e., both conditions can lead to side effects that can interfere differently with fluency and prosody), 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 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, and 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, and Casasanto (2014) also found that asking speakers to perform nonreferential 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 for successfully recalled words were shorter in the beat gesture than in the iconic gesture condition and in the no-instruction condition (the longest reaction times 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. The study of Vilà-Giménez and 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 nonreferential hand gestures that associate with speech) during story telling improved children's narrative performance and fluency in the posttest phase, when no instructions about gestures were given. Also, Vilà-Giménez, Igualada, and 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, and 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 toward 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 posttest 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 interword intervals, speaking time, total sentence duration, speaking rate, and phrasing strategy by triggering a more natural speech chunking; Garcia, Cannito, & Dagenais, 2000; Garcia & Cobb, 2000; Garcia & Dagenais, 1998; Hustad & Garcia, 2005).

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 this study was 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), and F0 and intensity. Considering previous findings, we expect that encouraging the use of hand gestures benefits speech fluency and has an impact on F0 and intensity.

Method

This study used a narration task in which the participants had to watch and describe a set of comic strips in two different conditions: nonencouraging gesture condition (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.

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

Participants

Twenty female native speakers of Italian participated in the experiment. They were all from the Veneto region (M_{age} = 24.2 years, SD = 2.9). Seventeen of them were undergraduate students at the University of Padua, and three of them were alumni 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 women are more expressive and produce more gestures when speaking than men (e.g., 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 1 for an example and see the Appendix for the whole set of 16 comic strips). The comic strips were carefully selected and adapted so that they were considered equivalent in terms of complexity and length (four-scene narrations). 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 selected 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 an HD video camera (JVC GZ-HD7E Everio), and speech was recorded (16-bit .wav files, 44.1-kHz sampling rate) as a separate audio track using a MIPRO wireless head-mounted microphone with a bodypack transmitter connected to a Zoom R16 digital audio mixer. All levels were set prior to the first participant and remained consistent throughout the data collection. 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) who 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.

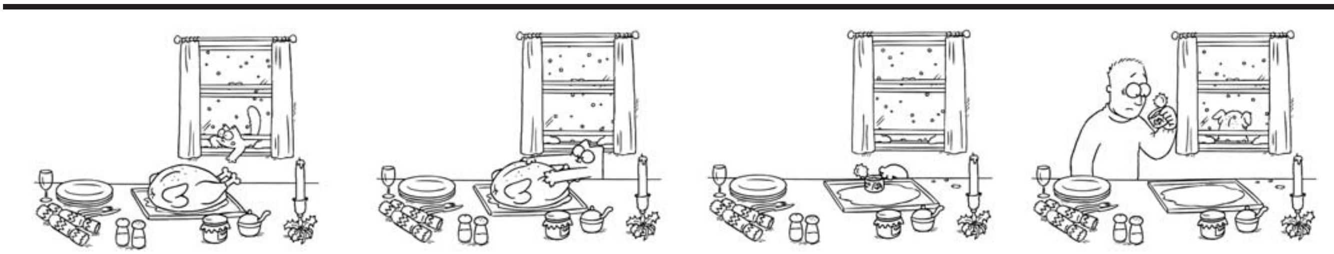
Each participant entered the room and was first given an informed consent form to sign. She was introduced to the confederate listener as if he was also a fellow participant. Both the participant and the listener were given written instructions. The participant received the following instructions (translated 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 show, 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) will be able to reconstruct it by placing the four comic cards that make up the strip in the correct order.

The reason why we made the 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 & Wilkin, 2009) and 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, who was the same person across the whole data collection, 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 the participant and the listener was natural, he was allowed to interact with the participant after the narration task and while solving his part of the task, that is, 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

Figure 1. Example of a four-scene comic strip used for the experiment (from *Simon's Cat*, by Simon Tofield; see the Appendix).



started with a set of two initial familiarization trials to show the participants how the task should be performed and to make them confident about the camera. Specifically, each trial consisted of a three-step sequence: (a) the participant examined a four-scene comic strip to learn the story it depicted (for approximately a minimum of 5 s to a maximum of 40 s); (b) the comic strip was then concealed, and the participant told the story to the listener in a face-to-face interaction; and (c) 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 retell the first half of the comic strips set (2 extra familiarization stories + 5 target trials) in the *nonencouraging 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 an 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 setup 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 and 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 min. Audiovisual recordings of 200 short narratives were obtained (20 participants × 10 target trials), lasting 81.2 min (39.1 min in the N condition and 42.1 min in the E condition).

Assessment of the Effects of the Encouraging Gesture Prompt

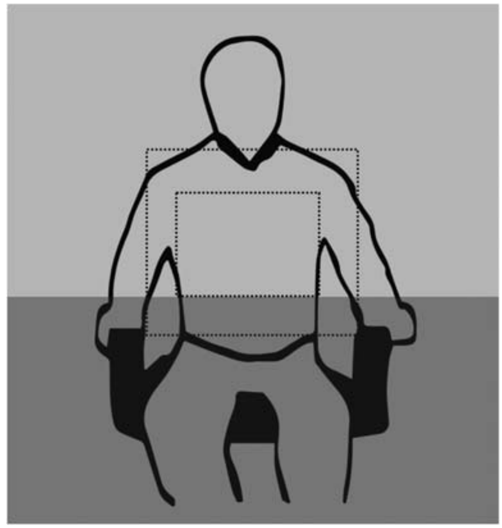
The amount of gesturing that was present in the speakers' narrations across the two conditions (N, E) was quantitatively assessed over the entire data set. This was done to ascertain that the gesture elicitation instruction in the E condition had actually caused an increase in the speakers' gestures with respect to the N condition. All instances of co-speech gestures were identified and manually annotated by the first author with the software ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). The annotation criteria consisted in the marking of all gestural *stroke* (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). Nongestural movements (self-adaptors, e.g., scratching, touching one's hair; Ekman & Friesen, 1969) were excluded. The speakers produced 2,396 gestures (1,015 in the N condition and 1,381 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's (1992) representation of the gesture space, which is divided into sectors delimited by concentric squares. For the present annotation, a two-sector version of McNeill's representation of the gesture space was used, as illustrated in Figure 2.

When the gesture stroke was produced in a more central, higher, and visually prominent area (Streeck, 1994) of the gesture space (the lighter gray area), the gesture was coded as *salient*. Whereas, when the gesture stroke was produced in a less visually prominent area (the lower darker sector), it was coded as *nonsalient* (e.g., those gestures performed while keeping the arms along the legs or on the armrests). 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 nonsalient (NS) gesture rate. The prompt to gesture worked well, as shown in Table 1 and Figure 3. The effects of encouraging gestures on gesture rate and gesture salience were tested by running three linear mixed-effects models (R function *lmer* in *lme4* package; see Bates, Mächler, Bolker, & Walker, 2014). Each model included one of the three dependent variables listed in Table 1 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 rates (est. = 4.134, $SE = 0.708$, $p < .001$) and salient gesture rates (est. = 10.723, $SE = 0.794$, $p < .001$) were significantly higher in the E condition than in the N condition, and conversely nonsalient gesture rate (est. = –.589, $SE = 0.65$, $p < .001$) was higher in the N 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,

Figure 2. Gesture space representation (based on McNeill, 1992, p. 89). Upper (light grey) area: salient gestures; Lower (dark grey) area: nonsalient gestures.



& Goldin-Meadow, 2010), leading speakers to use more gestures that also involve a higher and wider gesture space.

Data Analysis: Transcriptions, Fluency Coding and Acoustic Analysis

Speech Discourse Length

The recordings were edited so that a separate 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 story 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. A count of word tokens was also included as a measure of how talkative participants were in every story told.

Fluency and Disfluency Measures

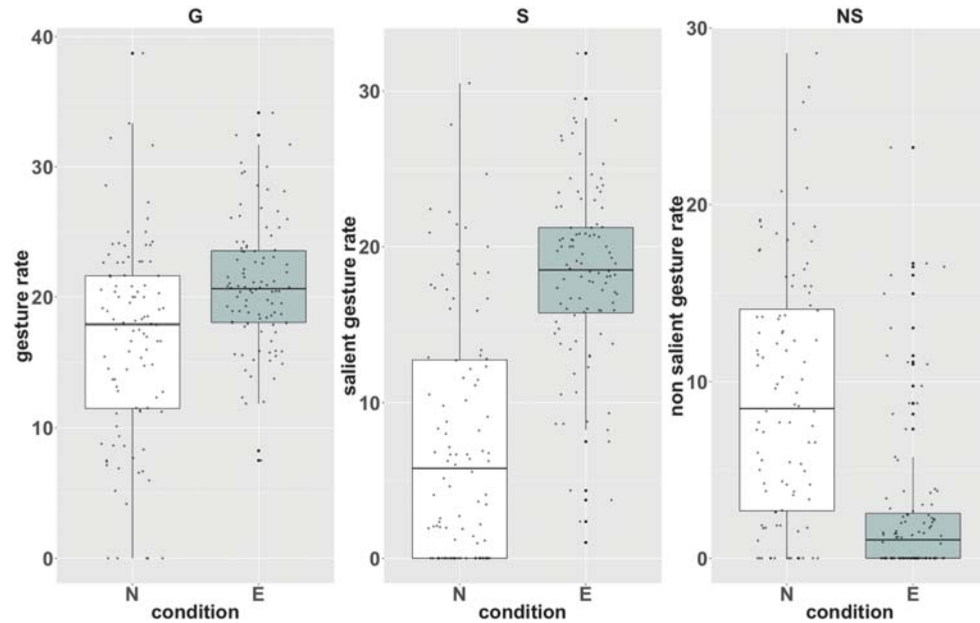
Fillmore, Kempler, and Wang (2014, p. 93) define fluency as "the ability to talk at length with few pauses, (…) to fill time with talk. A person who is fluent (…) does not have to stop many times to think of what to say next or how to phrase it." In addition, according to Zellner (1994, p. 48) "people are disfluent if they often hesitate, make nonfunctional pauses and make speech errors and self-corrections." Thus, fluency is usually measured not only by measures of speech rate but also by the absence of a set of features that characterize disfluency. In this study, we used a measure of speech rate, which was automatically obtained using a Praat script (de Jong & Wempe, 2009). 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 the first author (examples from our data are reported below in Italian and translated in English for convenience of the reader):

Table 1. Linear mixed-effects models for the effects of condition on gesture rate, salient gesture rate, and nonsalient gesture rate (per 100 words).

Variable	Estimates	*SE*	CI Lower	CI Higher	*t*	χ^2	*p*
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
Nonsalient gesture rate	−6.589	0.65	−7.868	−5.311	−10.13	80.71	< .001

Note. Models: R function lmer in lme4 package (Bates et al., 2014). Each model included condition (nonencouraging [N], encouraging [E]) as a fixed effect and both story and participant as random intercepts. Number of observations: 200; groups: participants, 20|story, 10. Confidence interval (CI): 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).

Figure 3. Gesture rate (G), salient gesture rate (S), and nonsalient gesture (NS) rate per condition (nonencouraging [N], encouraging [E]).



- *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), and phonology-based (slip of the tongue or unclear pronunciations);
- *filled pauses* (sounds like "ehm" and "mmm") and *prolongations of vowels* (e.g., "alloraaa… il gatto…," "then, *theee*…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 Praat software (Boersma & Weenink, 2018). To explore whether F0 and intensity were modulated differently across the N and E 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 modal voice register was centered between 100 and 500 Hz. Previous literature has shown that, for female speakers, vocal fry register excursions fall in a low-frequency F0 range that is generally below 100 Hz (Hollien & Michel, 1968; McGlone, 1967; Murry, 1971), with a mean of approximately 50 Hz (as reported in the literature review provided in Blomgren, Chen, Ng, & Gilbert, 1998). Thus, we decided to set F0 floor and ceiling to 100 and 500 Hz, respectively, for all participants. Setting the floor to 100 Hz 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[2]: The script extracts a pitch value every 10 ms of speech via autocorrelation algorithm for the whole audio file (story told), F0 floor and ceiling values can be specified. 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).

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

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, and maximum intensity and standard deviation were computed.

Statistical Analysis

The data analysis focused on 19 variables of interest: (a) story duration (in seconds), (b) number of words per story, (c) repetition rate, (d) insertion rate, (e) interruption rate, (f) self-correction rate, (g) filled pauses rate, (h) silent pauses rate, (i) total disfluencies rate (including 3, 4, 5, 6, 7), (j) speech rate, (k) minimum F0, (l) maximum F0, (m) mean F0, (n) F0 standard deviation, (o) PVQ, (p) minimum intensity, (q) maximum intensity, (r) mean intensity, and (s) intensity standard deviation. The main descriptive statistics per variable are shown in Table 2.

The effect of gesture encouragement (within-subject factor) on speech was tested by running 19 linear mixed-effects models (R function *lmer* in *lme4* package; see Bates et al., 2014). Each model included one of the 19 dependent variables listed in Table 2 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). The tests were then corrected for multiple testing via false discovery rate (FDR; i.e., Benjamini–Hochberg procedure; Benjamini & Hochberg, 1995). The results are reported with the adjusted FDR critical values.

Table 2. Main descriptive statistics.

	M		*SD*	
Variable	**N**	**E**	**N**	**E**
Story duration (s)	23.44	25.23	8.17	10.25
No. 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 (total)	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 variation (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 variation (dB)	10.84	10.86	1.39	1.31

Note. N = nonencouraging condition; E = encouraging condition; 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.

Results

Table 3 shows the results of the 19 linear mixed-effects models. There was an effect of condition on speech discourse length, specifically on the number of words that was higher in the E than in the N condition (est. = 5.09, *SE* = 1.945, $p < .011$, FDR). As for the measures of fluency, there was no effect of condition in any of the disfluency measures, nor was there 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, *SE* = 8.302, $p < .008$, FDR), and an effect was found for maximum intensity (est. = 1.254, *SE* = 0.297, $p < .005$, FDR) and mean intensity (est. = 0.846, *SE* = 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). In Figure 4, the four jittered box plots refer to the four variables that showed a significant effect of condition.

Discussion and Conclusions

The aim of this 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 as well as 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 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 prosody, 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: (a) production of longer discourse in terms of the number of words used, (b) higher F0 maxima, and (c) louder speech. Let us now comment on each of these results separately.

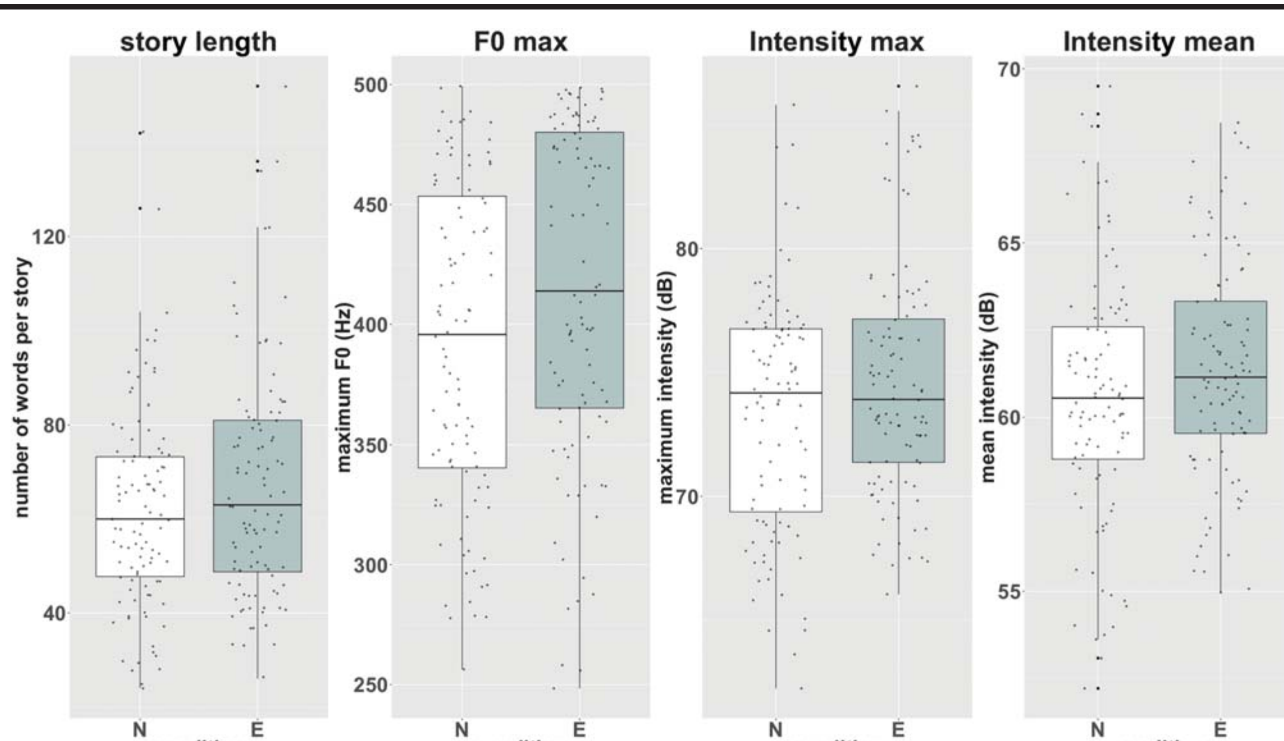
First, when instructed to gesture, speakers used more words to tell each story. The increase in the number of words in the E condition, albeit significant, is rather modest (approximately five words), and it should be interpreted with care. In spite of this, this result is coherent with Parrill et al.'s

Table 3. Results of the linear mixed-effects models per dependent variable.

Variable	Estimates	*SE*	CI Lower	CI Higher	*t*	χ^2	*p*	FDR
Story duration (s)	1.792	0.745	0.329	3.256	2.407	5.73	.017	0.013
No. words	5.090	1.945	1.268	8.913	2.617	6.753	.009	0.011*
Repetitions	0.477	0.231	0.023	0.931	2.066	4.237	.04	0.016
Insertions	−0.132	0.133	−0.393	0.128	−0.997	0.996	.318	0.029
Interruptions	0.16	0.182	−0.197	0.518	0.881	0.779	.377	0.032
Self-corrections	0.242	0.197	−0.146	0.63	1.228	1.509	.219	0.024
Filled pauses	−0.45	0.416	−1.267	0.368	−1.081	1.17	.279	0.026
Silent pauses	0.145	0.42	−0.682	0.971	0.345	0.119	.73	0.042
Disfluencies (total)	0.298	0.655	−0.99	1.586	0.455	0.208	.649	0.039
Speech rate (syll/dur)	−0.029	0.048	−0.123	0.065	−0.615	0.38	.538	0.037
F0 min (Hz)	−0.191	1.247	−2.641	2.259	−0.153	0.024	.878	0.050
F0 max (Hz)	22.578	8.302	6.251	38.904	2.719	7.272	.007	0.008*
F0 mean (Hz)	0.458	0.663	−0.846	1.761	0.69	0.478	.489	0.034
F0 variation (Hz)	1.148	0.715	−0.257	2.554	1.606	2.574	.109	0.018
PVQ	0.005	0.004	−0.002	0.013	1.443	2.08	.15	0.021
Intensity min (dB)	0.0198	0.078	−0.133	0.173	0.254	0.065	.799	0.045
Intensity max (dB)	1.254	0.297	0.67	1.837	4.221	17.043	.001	0.005*
Intensity mean (dB)	0.846	0.15	0.552	1.14	5.652	29.459	.001	0.003*
Intensity variation (dB)	0.024	0.097	−0.167	0.215	0.245	0.06	.806	0.047

Note. Number of observations: 200; Groups: participants, 20|story, 10. Confidence interval (CI): lower, 2.5%; higher, 97.5% (R package confint). False discovery rate (FDR) 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

Figure 4. Box plots representing the variables that showed a significant effect of condition (nonencouraging [N], encouraging [E]). They show the full range of variation (from minimum to maximum), the likely range of variation (the interquartile range), and the typical value (the median) across conditions.



(2016) 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 have a role on conceptualizing, phrasing, and chunking speech differently? Does it lead to more creative language use and imaginative speech? 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 (Kita et al., 2017; Kita & Özyürek, 2003; Hostetter & Alibali, 2008).

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 (Llanes-Coromina et al., 2018; Vilà-Giménez & Prieto, 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).

Third, the results showed that encouraging participants to gesture increased F0 maximum and intensity features in the speakers' narratives (i.e., 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 3), 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 et al., 2018), and future investigations are needed to further explore the existence of a shared motor control system that controls the production of both modalities, that is, 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.

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 et al. (2014). Indeed, Hoetjes et al. showed that the inability to gesture while speaking does not have any detrimental effect on pitch modulation and prosody. However, it might well be that the two experimental settings are not exact polar opposites (e.g., restraining or encouraging gestures may lead to all types of other effects that might interfere with fluency and prosody). For example, the two settings (i.e., being encouraged or restrained to gesture) could cause an increase in cognitive load that could interfere with speech planning processes, thus affecting speech prosody; also, drawing speakers' attention to gesture could switch on speakers' preconceptions about how speech produced with or without bodily expressions should be like; this can influence speakers' speaking styles (in a way that 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 et al., 2000; Garcia & Cobb, 2000; Garcia & Dagenais, 1998; 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.

Acknowledgments

We are grateful to Nicolò Marcato for helping with the recordings and also serving as the "fellow participant" during the experiment. We thank Federica Cavicchio, Simone Gastaldon, and Joan Borràs-Comes for comments and advice on the statistical analyses. We are also grateful to the anonymous 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 is supported by a grant from the University of Padova's Doctoral School of Linguistic, Philological and Literary Sciences; the third author also acknowledges Grant FFI2015-66533-P from the Spanish Ministry of Science and Innovation and Grant 2014SGR-925 from 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*. Retrieved from https://arxiv.org/abs/1406.5823

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. 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 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. https://doi.org/10.1121/1.422785

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

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

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology. General, 140*(1), 102–116. https://doi.org/10.1037/a0021790

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. (2017). *Why gesture?: How the hands function in speaking, thinking and communicating* (Vol. 7). Philadelphia, PA: John Benjamins. Retrieved from https://books.google.it/books?hl=it&lr=&id=hYerDgAAQBAJ&oi=fnd&pg=PR1&dq=Church,+R.+B.,+Alibali,+M.W.,+%26Kelly,+S.+D.+(Eds.).+(2017).+Why+gesture%3F+How+the+hands+function+in+speaking,+thinking+and+communicating.+Amsterdam,&ots=Bd9RFK-MNC&sig=MznQU4d_dBm

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.

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). Philadelphia, PA: John Benjamins.

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.

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

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. 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. *Journal of Speech, Language, and Hearing Research, 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*. New York, NY: Academic Press.

Finlayson, S., Forrest, V., Lickley, R., & Beck, J. M. (2003). Effects of the restriction of hand gestures on disfluency. *Proceedings of DISS, Gothenburg Papers in Theoretical Linguistics, Göteborg University, Sweden, 90*, 21–24.

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, 111*(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. https://doi.org/10.1044/1058-0360.0902.107

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. Retrieved from https://scholar.google.it/scholar?hl=it&as_sdt=

0%2C5&q=The+effects+of+gesturing+on+speech+intelligibility+and+rate+in+ALS+dysarthria%3A+A+case+study&btnG=#d=gs_cit&u=%252Fscholar%253Fq%253Dinfo%253AHU-kenPZwFaYJ%253Ascholar.google.com%252F%2526output%253Dcite%2526scirp%253D0%25

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. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9859884

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.

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

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* (Vol. 53). Philadelphia, PA: John Benjamins.

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.

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

Hollien, H., & Michel, J. F. (1968). Vocal fry as a phonational register. *Journal of Speech and Hearing Research, 11*(3), 600–604. https://doi.org/10.1044/jshr.1103.600

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. (2018). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*. 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 D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 1097–1102). Austin, TX: Cognitive Science Society.

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. https://doi.org/10.1075/gest.12.1.04hos

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*(5), 996–1012.

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

Jenkins, T., Coppola, M., & Coelho, C. (2018). Effects of gesture restriction on quality of narrative production. *Gesture, 16*(3), 416–431.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, United Kingdom: Cambridge University Press. Retrieved from https://books.google.it/

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. *Psychological Review, 124*(3), 245–266.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*(1), 16–32.

Kormos, J. (2014). *Speech production and second language acquisition*. New York, NY: Routledge.

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 D. McNeill (Ed.), *Language and gesture* (Vol. 2, p. 261). Cambridge, United Kingdom: 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. https://doi.org/10.5334/labphon.75

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

Llanes-Coromina, J., Prieto, P., & Rohrer, P. (2018). Brief training with rhythmic beat gestures helps L2 pronunciation in a reading aloud task. *Proceedings of the 9th International Conference on Speech Prosody, 2018,* 498–502. https://doi.org/10.21437/SpeechProsody.2018-101

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 Annual Meeting of the Cognitive Science Society, 36,* 898–903.

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. https://doi.org/10.1023/A:1023274823974

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

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

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.

Murry, T. (1971). Subglottal pressure and airflow measures during vocal fry phonation. *Journal of Speech and Hearing Research, 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 Communication, 69,* 1–8. https://doi.org/10.1016/J.SPECOM.2015.01.005

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

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. https://doi.org/10.1111/j.1551-6709.2010.01102.x

Pouw, W., Harrison, S. J., & Dixon, J. A. (2018). Gesture-speech physics: The biomechanical basis for the emergence of gesture-speech synchrony. *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/tgua4

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 & Review, 10*(3), 610–615. https://doi.org/10.3758/BF03196522

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.

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). Amsterdam, the Netherlands: IOS Press.

Streeck, J. (1994). Gesture as communication II: The audience as co-author. *Research on Language and Social Interaction, 27*(3), 239–267. https://doi.org/10.1207/s15327973rlsi2703_5

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. https://doi.org/10.1037/dev0000604

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. International Speech Communication Association. https://doi.org/10.21437/Speech-Prosody.2018-143

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

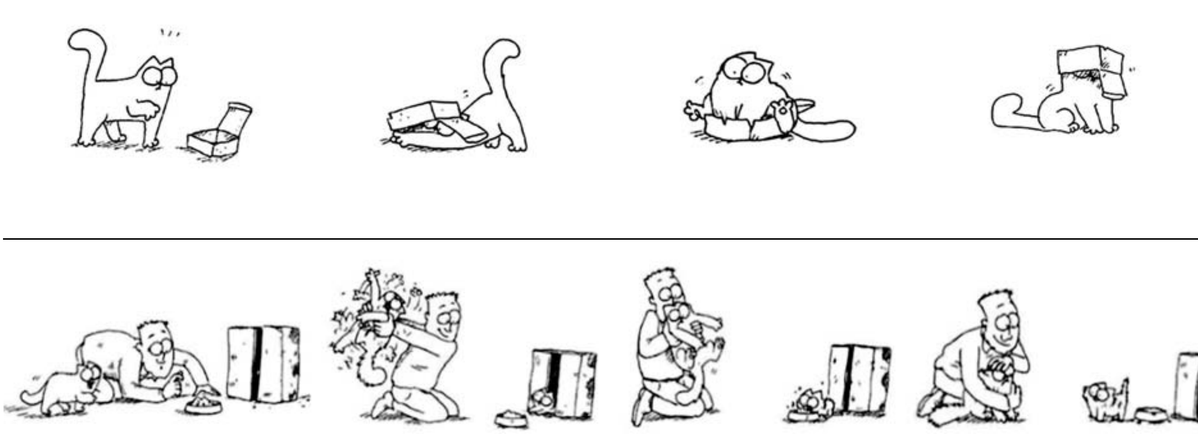
Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC '06), 2006,* 1556–1559.

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, United Kingdom: Wiley.

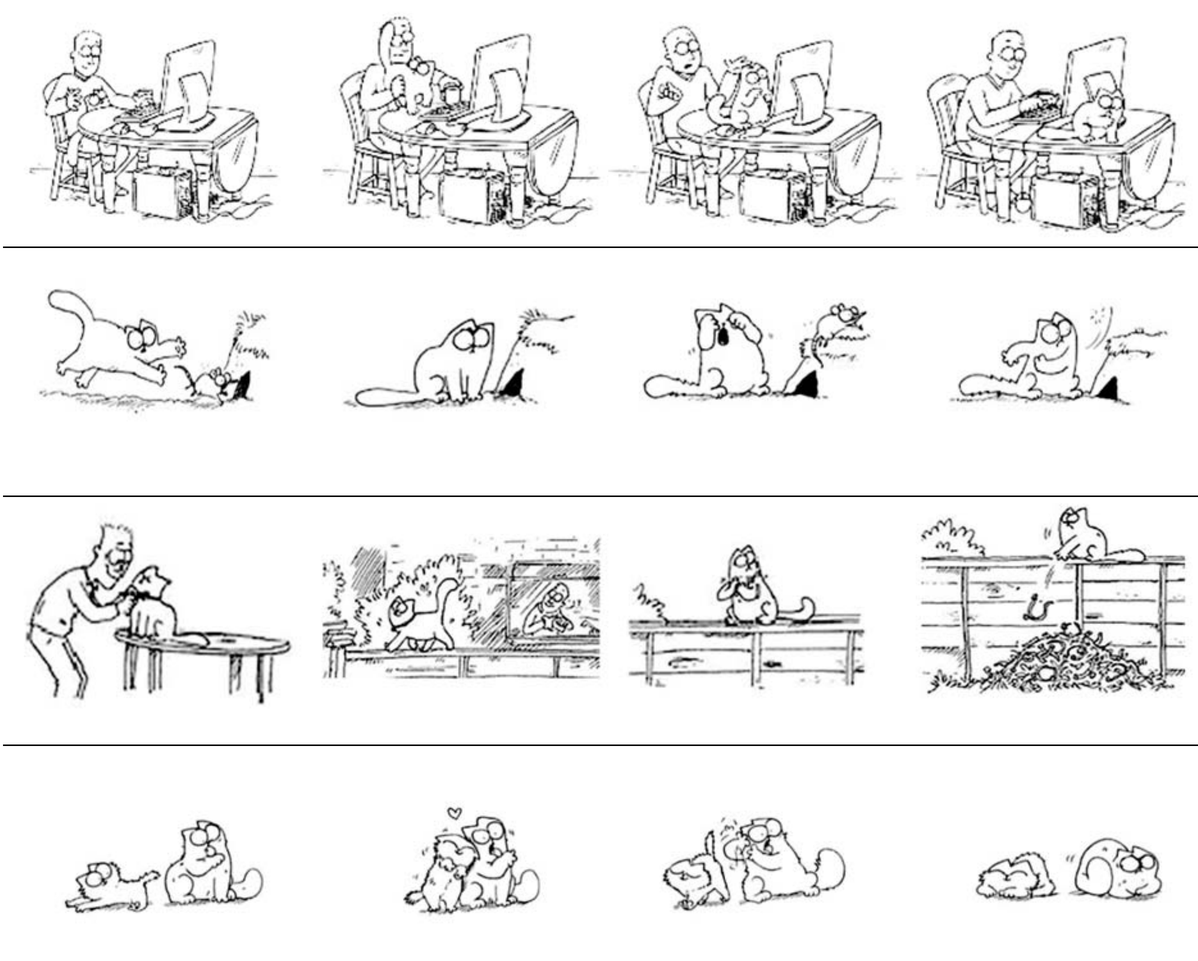
Appendix (p. 1 of 3)

Complete Set of Four-Scene Comic Strips (*Simon's Cat* by Simon Tofield)

Familiarization

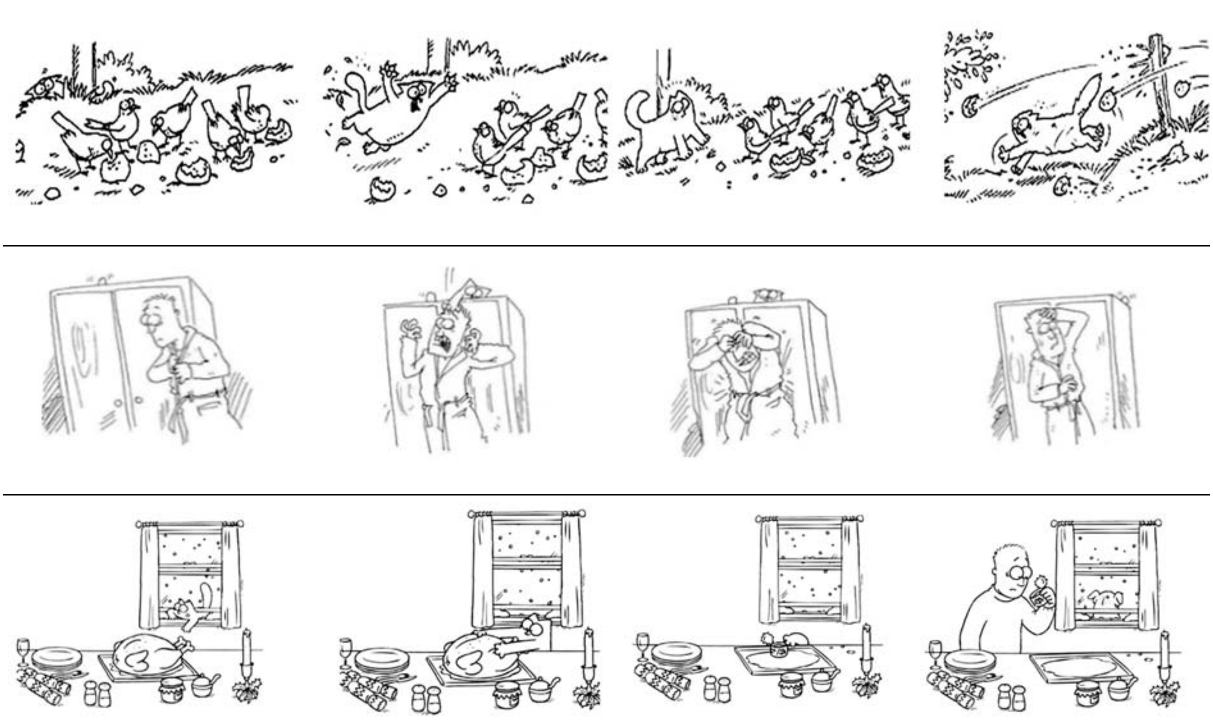


N/E

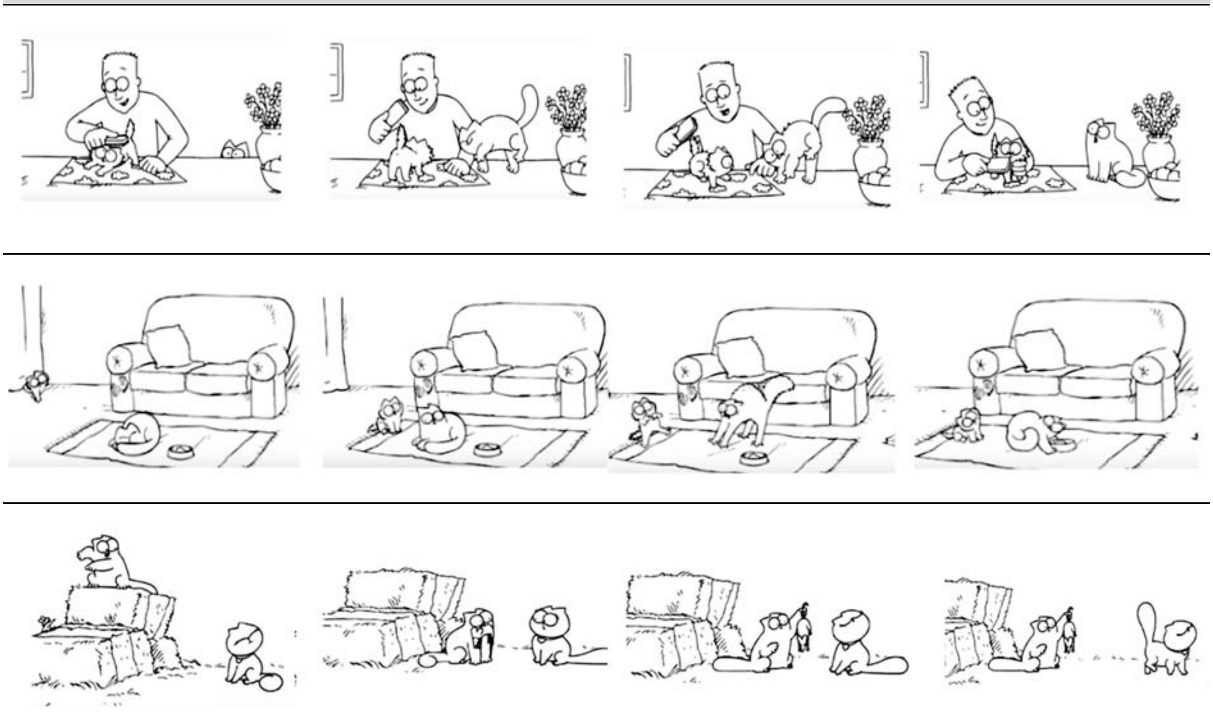


Appendix (p. 2 of 3)

Complete Set of Four-Scene Comic Strips (*Simon's Cat* by Simon Tofield)

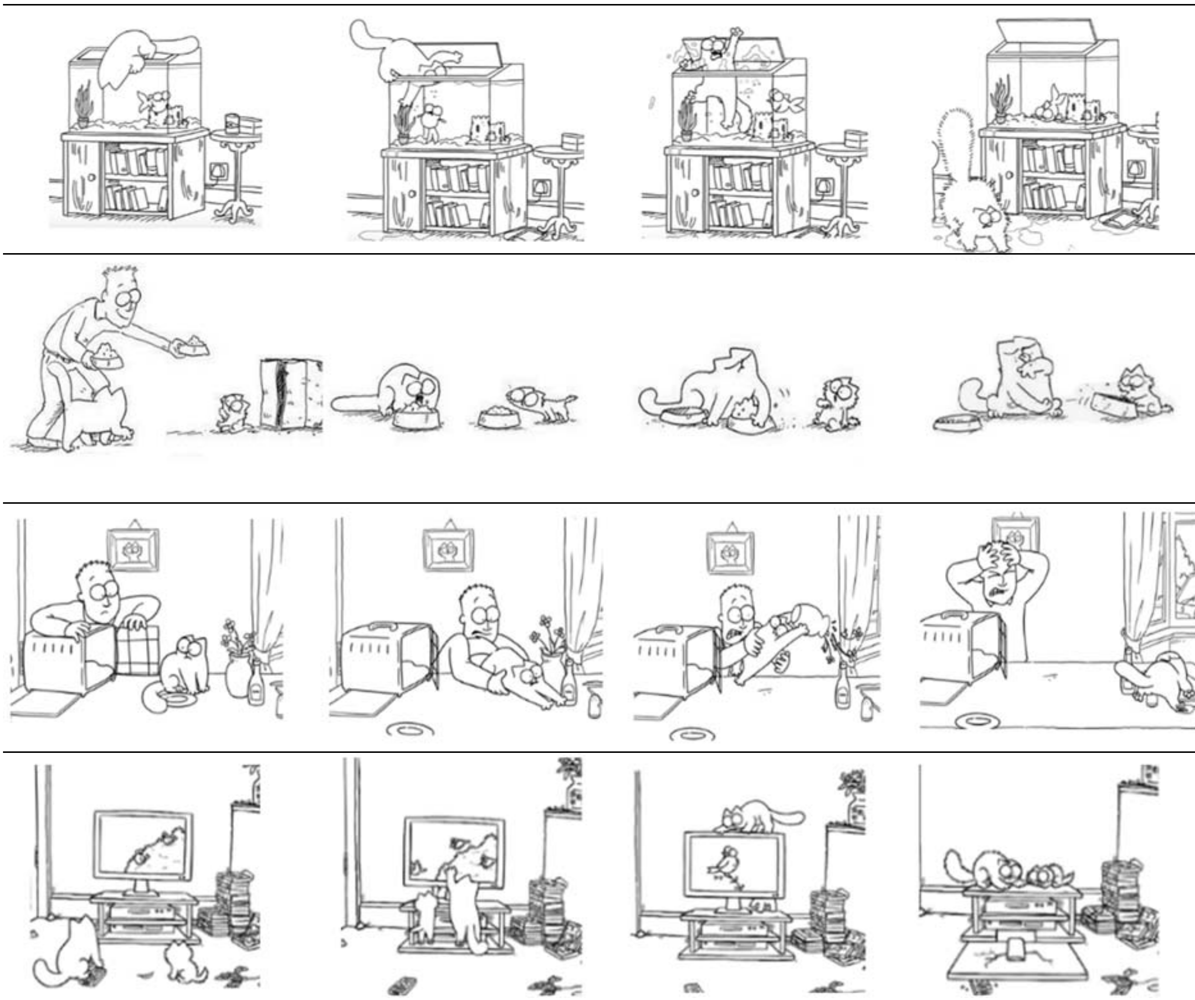


E/N



Appendix (p. 3 of 3)

Complete Set of Four-Scene Comic Strips



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.

Note. Images from *Simon's Cat* by Simon Tofield. Reprinted with permission. © Simon's Cat Ltd.