Gestures do semantic work

JK;HR;MC

Psychology, PPLS, University of Edinburgh

Josiah King

Philosophy, Psychology and Language Sciences

University of Edinburgh

7 George Square

Edinburgh EH8 9JZ, UK

J.P.J.King@sms.ed.ac.uk

Abstract

When a task becomes more conceptually demanding, speakers tend to produce more gestures. Previous research pins this increase either on gestures helping the speaker to package more complex information into appropriate units for speech, or on some of the communicative load being traded off from speech to gesture. The former view holds that speech and gesture co-vary, increasing in parallel with cognitive load. The latter account suggests the inverse: A negative relationship between speech and gesture indicating a transference of communicative load to whichever modality is least difficult.

In the present study we elicit semi-naturalistic dialogues, and measure the relative durations of gestures and speech directly, in order to establish whether these metrics co-vary (as would be predicted by the packaging account) or correlate inversely (as would be predicted by a trade-off). Pairs of participants took part in a shape-matching game, alternating in the roles of director and matcher. Directors saw two shapes (one easy, one difficult) for two seconds, and subsequently described them to their partner.

In line with an information packaging account, speech duration and gesture duration increased in parallel for easy-to-name objects. Importantly, when participants were referring to objects which were more difficult to verbally encode, gesture duration increased at a higher rate than speech duration. This suggests that, when naming difficult objects, gesture does more than simply help the speaker to package verbal information. Instead, gesture serves an additional communicative purpose, adding to, but not trading off against, the verbal descriptions which are uttered.

Gestures do semantic work

During conversation, speakers often move their hands and arms in meaningful ways. These movements might add emphasis to speech, indicate location, or depict properties of objects, movements, actions or space. Research suggests that when a task becomes more conceptually demanding, speakers tend to produce more gestures. This might be because gestures have benefits for speech, either in aiding lexical access (Krauss, Chen, & Gotfexnum, 2000; Rauscher, Krauss, & Chen, 1996) or in helping to wrap information into speech (Kita, 2000). Alternatively, gesture production might increase with conceptual demand because some of the communicative load is being traded off from speech to gesture (Bangerter, 2004; de Ruiter, 2006; Melinger & Levelt, 2004).

In contrast to common methods of measuring the number of discrete gestures per word, experimental trial or some other construct, the present study measures the relative durations of gestures and speech directly, in order to establish whether they co-vary (as would be predicted by a gesturing-to-benefit-speech account) or correlate inversely (as would be predicted by a trade-off). With few studies having investigated the speech-gesture relationship in a dialogical setting where two people are making explicit contributions, we elicit semi-naturalistic dialogues in which participants jointly engage in producing referring expressions.

Although gesturing has traditionally been treated as a direct intent to communicate meaning to an addressee, recent research has suggested that it may in part be motivated by the cognitive benefits it has for the speaker. Advantages of gesturing have been found in many areas: In speech production and planning processes (Kita, 2000; Krauss & Hadar, 1999; Morsella & Krauss, 2004; Rauscher et al., 1996; Rose & Douglas, 2001); spatial working memory (Morsella & Krauss, 2004; Wesp, Hesse, Keutmann, & Wheaton, 2001); conceptual planning (Melinger & Kita, 2007); and even doing mental arithmetic (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001).

Explanations of the mechanism by which gesturing may benefit production of speech have been varied. One suggestion holds that gesturing increases activation on relevant items in the mental lexicon, thus facilitating access (Krauss et al., 2000).

Another (Hadar & Butterworth, 1997) suggests that gesturing prevents visuo-spatial imagery from decaying, providing speech production processes with higher quality information. In a similar vein, the Information Packaging Hypothesis (Kita, 2000; Kita & Özyürek, 2003) claims that gesturing helps speakers to package complex information into appropriate units for speech.

Common to all these accounts is the claim that as conceptual load increases, speech and gesture increase in parallel. A 2009 study by So et al. found evidence which patterned with this claim. Participants were asked to describe scenes from videotaped vignettes (e.g., a man giving a woman a basket), and their use of both speech and gesture to indicate characters in the scene was measured. So et al. found that speakers more often used a gesture to identify a referent if it was also specified in speech. So et al. viewed their results as evidence in support of an account of speech and gesture going 'hand-in-hand' — i.e. the two modalities co-varying.

In contrast to So et al.'s evidence for co-variance, other studies point to the use of gesture to compensate for underspecifications in their spoken descriptions. For an example, in a communication task about spatial arrangements of connected dots of different colours where it was possible to determine the minimal content necessary for each stimulus, Melinger and Levelt (2004) examined whether omissions in speech were accompanied by compensatory gestures. Melinger and Levelt found that people who gestured produced more — and different — omissions in speech than people who did not.

Melinger and Levelt's findings pattern with other studies (Bangerter, 2004; de Ruiter, 2006; van der Sluis & Krahmer, 2007) in which the informational content of speech is found to inversely correlate with the amount or precision of gestures. These findings point to a trade-off account which holds that speakers distribute communicative load across speech and gesture such that that when speaking becomes more difficult, gesturing will increase (and vice versa). This view directly contrasts with the co-varying account, suggesting that as it becomes harder to verbally encode meaning, rates of speech will decrease while rates of gesture will increase.

To directly test the trade-off account, several studies have measured gesture production while manipulating the effort required to formulate spoken descriptions. de Ruiter (1998) found that gesture rate did not change depending on whether speakers were describing simple arrangements of shapes and vertical/horizontal lines or random arrangements of shapes and diagonal lines. However, as Morsella and Krauss (2004) point out, what de Ruiter varied in his stimuli was complexity, and not describability. In an experiment designed to tease apart these two concepts, Morsella and Krauss (2004) concluded that while visual complexity did not affect gesture rates, verbal codability did, with harder-to-name pictures (squiggles) eliciting higher rates of gesturing than easy-to-name pictures (familiar objects).

One explanation of the contrasting accounts of the speech-gesture relationship is that some types of gesture may be intended to communicate in place of speech (i.e. "A building with a roof like this [gestures a pointy roof]") while others might not. There is some evidence for the communicative intent of gestures: Alibali, Heath, and Myers (2001) found that speakers who could see their addressee used more *representational* gestures, but not *beat* gestures. Similarly, de Ruiter, Bangerter, and Dings (2012) found mutual visibility to increase both pointing and *obligatory iconic* gestures (gestures containing information not represented in speech but required for disambiguation), but not *non-obligatory iconic* gestures (gestures which are not essential to the understanding of co-occurring speech). However, de Ruiter et al.'s 2012 study found no evidence of a speech-gesture trade-off for *any* type of gesturing, with speakers' gesture rates not changing depending upon whether referring to simple, humanoid or abstract tangrams.

A further explanation for the differing findings is the variability in how speech and gesture have been measured. Previous studies have varied widely in their choice of measures for both modalities. Tending towards *rate* measures (i.e. gesture as a function of speech), the predominant strategy has involved measuring the number of discrete gestures produced when speaking (e.g. de Ruiter et al. (2012); Gerwing and Allison (2011); Hoetjes, Koolen, Goudbeek, Krahmer, and Swerts (2015); Hostetter, Alibali, and Kita (2007)). Unlike speech — where distinct phonemes and words offer comparatively clear means of measuring utterance length and duration — multiple pieces of information (and multiple types of gesture) may be produced in the time between the raising and lowering of hands. In the literature, defining individual gestures has varied with the thing being described, from “illustrating a feature of the target (for instance its shape)” when describing tangram figures (de Ruiter et al., 2012) to change in any one of “shape and placement of the hand, trajectory of the motion” when identifying referents in a narrative (So et al., 2009). On the side of speech, these measures of discrete gestures have then been calculated per trial (Morsella & Krauss, 2004), per minute (Mol, Krahmer, Maes, & Swerts, 2011), per 100 words (Gerwing & Allison, 2011; Hoetjes et al., 2015; Hostetter et al., 2007; Masson-Carro, Goudbeek, & Krahmer, 2015), per *feature description* (de Ruiter et al., 2012) or per *semantic attribute* (Hoetjes et al., 2015).

Because gestures and vocalisations can vary in durations, these metrics fail to capture the relative contributions of each. We propose a duration-based approach in which relative durations of speech and gesture are measured directly. By comparing the durations for which a speaker conveys (or attempts to convey) information via different channels when referring to either easy-to-name or hard-to-name shapes, we aim to establish whether speech and gesture trade-off against one another or co-vary. On a broad level, the two accounts would simply predict a correlation between durations of speech and gesture but in different directions: A positive correlation suggesting speech and gesture increase hand-in-hand, an inverse correlation suggesting a trade-off. Our manipulation of referent-nameability should, by either account, have no effect on the strength of the correlation: If gesture serves only as an aid to speech, then for both easy- and hard- to-name shapes we should see the same parallel increase in durations of speech and gesture. Likewise, if gesture serves only to compensate for speech, then easy- and hard-to-name shapes should both show the same negative correlation between the two modalities.

In addition to this new metric, our study targets investigating the speech-gesture relationship in a more ecological setting. To date, much of the research into gestures has involved studies in which a single speakers’ gestures are evaluated under various conditions. Experimental paradigms have tended towards those in which participants produce speech and gesture either to an imagined future addressee (e.g. Morsella and Krauss (2004); Wesp et al. (2001)) or to an addressee who is present but in a comparatively passive role (e.g. Bangerter (2004); de Ruiter et al. (2012); Hoetjes et al. (2015); Holler and Stevens (2007)). However, both of these designs fail to capture the dynamic process of conversation. Dialogue is often considered to be a joint activity (Clark, 1996), and gesture is no exception to this. In Bavelas, Gerwing, Sutton, and Prevost (2008), results suggested independent effects of visibility and dialogue on gesturing: Speakers produced more gestures when speaking to an occluded but dialogically involved listener than when speaking to a tape recorder. With this in mind, the present study elicits semi-naturalistic dialogues in which both participants make contributions.

Experiment

Pairs of participants engaged in a collaborative matching game, in which they were tasked with matching two target shapes seen by one particpant from a set of six shapes seen by the other participant. Participants took turns in the roles of *director* and *matcher*. Target shapes varied in verbal codability — they were either easy- or hard-to-name. We also manipulated participants' familiarity with the target shapes in both roles (director and matcher) — shapes were seen 4 times acros the experiment, meaning that descriptions were elicited for when speakers had previously described the shape, when they had previously had the shape described to them, and when the shape was novel. In a subset of trials in the second half of the experiment, familiarity was directly controlled such that the hard-to-name shape was repeated in consecutive trials.

Materials

Pairs of shapes in critical trials were selected from a set of 20 critical shapes (10 easy-to-name, 10 hard-to-name). Each of the 10 easy-to-name shapes was altered (sections of the shape were rotated and/or flipped) to create the 10 hard-to-name variants. For filler trials, shapes were selected from a further set of 40 shapes (20 easy-to-name shapes and 20 hard-to-name variants).

A set of 20 *distractor* shapes (10 easy-, 10 hard-to-name) were also included, but were never described, only being presented as part of the matchers' arrays, with the aim of encouraging specificity in descriptions.

The director's array was always comprised of only two shapes. In critical trials these shapes differed in codability (one easy-to-name and one hard-to-name), and in filler trials they were both of the same codability (both easy-to-name or both hard-to-name).

The matchers' array was comprised of six shapes. Along with the two target shapes for that trial, this set contained two randomly selected filler shapes and two randomly selected distractor shapes. Codability of the shapes in the array matched that of the target shapes (i.e. for critical trials, half were easy-to-name, half were hard-to-name). Shapes were randomly positioned for both the director (Left vs. Right) and for the matcher (in a 2x3 grid).

Experimental blocks

The experiment consisted of two blocks, each containing 40 trials (20 critical and 20 fillers). In each block, every critical shape was seen twice.

In the first block, trials were sampled randomly from a set of 20 critical trials (in which each critical shape was present in two trials) and 20 filler trials (in which two shapes of the same codability were randomly selected from the set of filler shapes). Selection was constrained such that shapes were never repeated in consecutive critical trials. Whilst the majority of the shapes were described once by each participant, the probability that a given shape was described twice by the same participant was 25%.[1]

In the second block the goal was to get a measure of behaviour when familiarity with the shape was more controlled. In the second block, pairs of consecutive critical trials alternated with pairs of filler trials. For the consecutive critical trials, the difficult-to-name shape was repeated. This meant that for critical trials where participant B was the director, they were tasked with describing the difficult shape which had just been described to them in the previous trial by participant A.

[1]Whilst the intention was to make it so that each participant described each critical shape once, a typing error in the experiment script resulted in these distributions

Filler trials, like for block 1, consisted of two randomly selected shapes of the same codability from the set of 40 filler shapes.

Procedure

Participants sat facing one another with an unobstructed space between them. Each participant had a monitor and a mouse on a table to their left, positioned such that they could not see what was on their partner's monitor. The setup was designed to encourage face-to-face dialogue, and to discourage participants from leaving their hand resting on the mouse whilst speaking (the position being uncomfortable for a right-handed mouse user), thus leaving both arms free to gesture. Audio and video was recorded by two cameras positioned to the right of each participant, facing their partner.

Taking turns in the roles of director and matcher, participants were tasked with successfully matching what was seen on the director's monitor from a set of possibilities on the matcher's monitor.

As a pair, participants were awarded points for successful matching. As an incentive, the highest scoring pair received £40, and participants were informed of this beforehand.

A high-score table was shown prior to the experiment, with participants adding their score to the table after they had played. Participants were asked to restrict their communication to within the 10 second time-window during which no images were present on either screen. The onset of this period was signalled by the images disappearing from the director's screen, and the offset was marked by the sound of a bell (see Figure 1). The aim of this was to encourage participants to look at their partners during communication, and not at their screens. Gesturing was permitted but not explicitly encouraged, as participants were told that during this period, they were "both allowed to talk, gesture, ask questions, and so on".

Feedback was given at the end of each trial (The number of correctly matched shapes was signified by an equivalent number of bell rings, with zero correct resulting in a buzzer), and the participants' cumulative score was displayed.

At the end of the experiment, participants were asked to complete a short questionnaire about their experience during the game. This questionnaire included asking whether it had occurred to them during the experiment that the study might be interested in gestures. For positive responses to this question, a follow-up question asked them to rate how much they felt that this affected their behaviour during the experiment (1="Not at all" to 7="A lot").

Coding

Audiovisual data for each pair of participants was coded using a three stage process: Audio-only and Video-only stages were used to code for speech and gesture respectively, with the third stage (both Audio and Video) used to confirm the annotations resulting from the previous stages. As each trial consisted of describing two shapes, special care was taken in the third stage to ensure that utterances and gestures were assigned to the correct referents.[2]

Speech Coding

Utterance duration, utterance length and disfluencies were coded in the Audio-only stage. Within each trial, only the first mention of each shape was used. Utterance duration (ms) was coded from the onset of the noun-phrase up until either a) speech-offset, or b) a valid interruption from the listener in either modality. Listeners' use of the collateral channel (for instance: "yep","mmhm",[nods head]) were not considered valid interruptions. Utterance length (number of words) was coded analogously, with disfluencies within the utterance period being identified and excluded from this measure.

Disfluencies were coded as falling into one of six categories: Filled pauses; Insertions; Substitutions; Articulation Errors; Deletions; Repetitions (see Shriberg (1996)). Any speech prior to the onset of the noun-phrase was coded as either fluent or disfluent.

Gesture Coding

Gestures were identified in the Video-only stage of the coding process. Any movement from the fingers up to the shoulder were considered. Only gestures which

[2]There was potential for descriptions in both modalities to be interleaved

partially overlapped an identified utterance period were included, and were assigned to the utterance which they primarily overlapped. This pairing was then confirmed in the third stage of the coding process. In any cases of a gesture being ambiguous as to which utterance it accompanied (i.e. non-representational gestures which overlap both utterance periods), the gesture was assigned to both referents.

Gestures were categorised into five types: Iconics; Beats; Points; Adaptors; and Others. Any gesture which was considered to be an attempt to represent any feature of the target shape was coded as an Iconic gesture. Beat gestures were identified as any movements which rhythmically matched prosody in speech but which *did not* represent any feature of the target shape. Point gestures were extensions of the index finger used to refer deictically to either present objects or people, or to previous parts of the discourse. Other movements were categorised as either adaptor gestures (scratching, stroking, manipulating clothing, etc.), or other miscellaneous gesticulations.

Individual gestures were identified by onset of movement, and continued either until the start of the retraction phase, or until transformation into a) a different category of gesture or b) iconic gesturing referring to a different shape (i.e. the other shape in the trial).

The third stage of the coding process (audio and video) was used to confirm the coding of the first and second stages, specifically gesture categorisation and pairing of gestures with referents. Several gestures remained ambiguous between iconic and beat even after this third stage (n referencing easy shapes, and n referencing difficult shapes).These were considered to be imprecise/lax attempts at representing the shapes in space, and thus coded as iconic gestures. Additionally, this stage was used to code whether or not the utterance referred explicitly to the gesture being made (e.g. "like this","like that","a bit here", etc.)

Once identified, iconic gestures were coded for gesture duration analogously to the measure of utterance duration: Gesture duration for iconic gestures was measured from the onset of the first stroke or hold phase (and so excludes initial preparation phase) up until the retraction phase, or until interruption. End-of-gesture hangs (uninformative hangs immediately prior to a retraction phase) were not included.[3]

This measure of gesture duration included any hangs, false starts, or preparation which occurred within-gesture, just as utterance duration included within-utterance pauses and disfluencies. Any suspected false starts and repetitions were counted (a finger trace which is subsequently reversed counts as repetition, as does a static hold with a distinct beat gesture incorporated).

Additionally, all gestures were coded for the hands used (Left, Right, or Both), and whether the representational part of the gesture was conveyed dynamically, statically, or as a combination of both.

Second rater

For 20% of critical trials, interrater agreement was measured for utterance duration, gesture duration and gesture type, with the second rater blind to the study aims (e.g. blind to the focus on the speech duration:gesture duration ratio).

Results

Forty-four participants in 22 pairs took part in the experiment in return for £7. When questioned, 33 (75%) participants indicated that it had occurred to them that the experiment might be interested in gestures. The follow-up question showed that these participants in general felt that this did not affect their behaviour during the experiment that much: 26 gave a rating of ≤ 3, and only 1 participant gave a rating ≥ 6.

Analysis was carried out in R version 3.4.2 (**?**), using the lme4 package (**?**). Only the critical trials were included in the analysis (1760 of the 3520 referring expressions elicited). For trials in which participants made no attempt to describe one or both of the shapes — either due to running out of time, forgetting to describe a shape, or appearing to forget what shape they had seen — those descriptions were excluded from all analysis (0.7% of referring expressions).

[3]We discerned here between end-of-gesture *hangs* and end-of-gesture *holds* which continued to convey some representational content, and were thus included as part of gesture duration

Of the remaining 1748 descriptions, 1494 (86%) were accompanied by some form of gesturing. The experiment elicited far more iconic gesturing than was hoped, with 1301 descriptions (74%) being accompanied by an attempt to gesturally represent the target shape. Pairs correctly matched almost all the shapes, with listeners making an incorrect response for only 4.5% of descriptions of hard shapes and 2.8% of descriptions of easy shapes.

Use of speech

Trivially, harder-to-name shapes should require longer descriptions to disambiguate them from competitors. To confirm this, utterance duration (Z-scored) and length were analysed using mixed effects linear regression. Verbal codability (easy-to-name vs. hard-to-name), trial number (1-80, Z-scored) and their interaction were included as fixed effects, thus controlling for any change across the experiment. Random intercepts both by-shapes and by-participant were also included, with by-participant random effect of verbal codability, trial number and their interaction, and by-shape random effects of trial number.

As expected, the experimental manipulation of a shape's verbal codability influenced participants utterances, with hard-to-name shapes resulting in longer utterances, both in terms of duration ($\beta = 1.27$, SE $= 0.12$, $t > 2$) and number of words ($\beta =?$, SE $=?$, $t?$). Trial number was found to predict utterance duration, but only for hard-to-name shapes, with speakers producing shorter utterances as the experiment progressed ($\beta = -0.21$, SE $= 0.06$, $t > 2$).

Occurrence and duration of iconic gesturing

Figure 2 shows the occurrence of different types of gestures accompanying descriptions of easy- and hard-to-name shapes.

To test whether our results pattern with previous research in showing an increase of gesturing with an increase in cognitive load, both the occurrence (gesture vs. no gesture) and the duration (Z-scored) of iconic gesturing were modelled using mixed effects logistic and linear regression respectively, with the same fixed and effects structure.[4]

Patterning with the literature, for harder-to-name shapes, speakers were more likely to produce iconic gesturing ($\beta = 3.30$, SE $= 0.41$, $p < 0.05$), and spent longer gesturing ($\beta = 1.27$, SE $= 0.10$, $t > 2$) Unlike utterance duration, the reduction in length of gestures as the experiment progressed applied for descriptions of easy-to-name shapes ($\beta = -0.06$, SE $= 0.02$, $t > 2$) as well as of hard-to-name shapes, for which this reduction was greater ($\beta = -0.23$, SE $= 0.06$, $t > 2$).

The speech/gesture relationship

To investigate whether gesturing co-varies or trades off against speech, the duration of iconic gesturing (Z-scored) was modelled by including utterance duration as a predictor of gesture duration: Utterance duration (Z-scored), verbal codability, trial number, and all interactions were included as fixed effects. Random intercepts and random slopes for utterance duration, verbal codability and their interaction were included by-subject, along with random intercepts and effects for utterance duration by-shape.[5]

Figure 3 shows the relationship between durations of speech and gesture, split by verbal-codability. Durations of speech and gesture co-varied: Across all shapes, as the duration of speech increased, so did the duration of iconic gesturing ($\beta = 0.70$, SE $= 0.04$, $t > 2$). Importantly, however, the duration of iconic gesturing increased at a higher rate when describing hard-to-name shapes over easy-to-name shapes ($\beta = 0.34$, SE $= 0.06$, $t > 2$).

The speech/gesture relationship and familiarity with the referent

Whilst the above analyses control for participants' familiarity with shapes by including trial number as a predictor, we were also interested in how repeated production and exposure to descriptions of a given shape influenced speakers' use of gesture. We

[4]With some speakers producing iconic gestures for all hard-to-name shapes, verbal codability could not be included as a random effect by-participants when modelling the occurrence.

[5]To allow the model to converge, random effects of trial number were not included on the basis of there being less variance for trial number both by-participant and by-shapes than other random effects mentioned

coded a shape's familiarity to a speaker as falling into one of five categories: New to the experiment; Described (speaker has described this shape before but not had it described to them); Heard (speaker has had this shape described to them before but has not described it themselves); Heard and Described (speaker has had to both describe this shape before, and has also previously had it described to them); and Heard in the previous trial (speaker has had this shape described to them in the trial immediately preceding this one). Table 1 shows a breakdown of these categories by both verbal codability and experimental block.

This coding scheme for familiarity was included as a predictor in the model of the speech-gesture relationship in place of trial number. Only having both heard *and* described a shape previously was found to affect the speech-gesture relationship: For easy-to-name shapes — but not hard-to-name shapes — the increase of gesture with speech was reduced ($\beta = -0.28$, SE $= 0.11$, $t > 2$) (See Figure 4).

Discussion

In line with previous research, our results found that both speech and gesture were used differently depending on whether they were describing easy- or hard-to-name shapes. Descriptions of hard-to-name shapes were more likely to be accompanied by an iconic gesture, and were accompanied by longer durations of gesturing. The duration of gesturing decreased over the course of the experiment as participants became more familiar with the shapes, patterning with previous research of the use of gesture for repeated referents (Hoetjes et al., 2015).

At a broad level, participants' use of gesture patterned with their use of speech, with harder-to-name shapes resulting in longer verbal descriptions. The durations of speech and gesture clearly show the two modalities going hand-in-hand: The longer participants spoke for, the longer they spent representing that shape in gesturing. However, when describing hard-to-name shapes, participants' use of gesture increased at a higher rate with speech than it did for easy-to-name shapes. Although these results support the idea that the use of iconic gesture parallels the use of speech in generating descriptions, they also show a tendency to favour gesture over speech when describing harder-to-name shapes. This suggests that gestures have a communicative function: When referring to difficult objects, gesture does more than simply help the speaker to package verbal information. Importantly, gesture adds to, but does not trade off against, the verbal descriptions which it accompanies.

Interestingly, the exclusion of any description in which there was spoken reference to the gestural component ("like this", "a bit here") did not change the pattern of results (20% of descriptions), nor did the exlusion of descriptions in which there was no iconic gesturing (26% of descriptions). This suggests that the favouring of gesture for harder shapes is not simply a result of participants explicitly drawing listeners' attention to the gestural component of a description (effectively explicitly stating a trade-off), nor is it due to a tendency to attempt descriptions of easier shapes without recourse to gesturing. Instead, the use of gesture is more nuanced: Communication occurs within a vector space of speech and gesture, which varies according to the relative efforts placed in either modality.

We propose that measuring the relative durations of speech and gesture offers a clearer window into how speakers distribute these communicative effort. It is in the distribution of effort (and not necessarily of information) that we see the trade-off between speech and gesture — the harder something is to convey in speech, the more effort is placed in gesturing.

References

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of Visibility between Speaker and Listener on Gesture Production: Some Gestures Are Meant to Be Seen,. *Journal of Memory and Language*, *44*, 169–188. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S0749596X00927529` doi: 10.1006/jmla.2000.2752

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*(6), 415–419. doi: 10.1111/j.0956-7976.2004.00694.x

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*(2), 495–520. doi: 10.1016/j.jml.2007.02.004

Clark, H. H. (1996). *Using language*. Cambridge university press.

de Ruiter, J. P. (1998). *Gesture and speech production*. [Sl: sn].

de Ruiter, J. P. (2006). Can gesticulation help aphasic people speak, or rather, communicate? *Advances in Speech Language Pathology*, *8*(2), 124–127. doi: 10.1080/14417040600667285

de Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The Interplay Between Gesture and Speech in the Production of Referring Expressions: Investigating the Tradeoff Hypothesis. *Topics in Cognitive Science*, *4*(2), 232–248. doi: 10.1111/j.1756-8765.2012.01183.x

Gerwing, J., & Allison, M. (2011). The flexible semantic integration of gestures and words: Comparing face-to-face and telephone dialogues. *Gesture*, *11*(3), 308–329. Retrieved from `http://www.jbe-platform.com/content/journals/10.1075/gest.11.3.03ger` doi: 10.1075/gest.11.3.03ger

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychological science : a journal of the American Psychological Society / APS*, *12*(6), 516–522. doi: 10.1111/1467-9280.00395

Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, *115*(1-2), 147–172. Retrieved from `https://www.degruyter.com/view/j/semi.1997.115.issue-1-2/semi.1997.115.1-2.147/semi.1997.115.1-2.147.xml` doi: 10.1515/semi.1997.115.1-2.147

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., & Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language*, *79-80*, 1–17. Retrieved from `http://dx.doi.org/10.1016/j.jml.2014.10.004` doi: 10.1016/j.jml.2014.10.004

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology*, *26*(1), 4–27. Retrieved from `http://search.ebscohost.com/login.aspx?direct=true{&}db=ufh{&}AN=24195444{&}site=ehost-live{%}5Cnhttp://jls.sagepub.com/cgi/doi/10.1177/0261927X06296428` doi: 10.1177/0261927X06296428

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, *22*(3), 313–336. doi: 10.1080/01690960600632812

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (Vol. 1, pp. 162–185). Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32. doi: 10.1016/S0749-596X(02)00505-3

Krauss, R. M., Chen, Y., & Gotfexnum, R. F. (2000). Lexical gestures and lexical access: a process model. *Language and gesture*, *2*, 261.

Krauss, R. M., & Hadar, U. (1999). The Role of Speech-Related Arm/Hand Gestures in Word Retrieval. *Gesture, speech, and sign*, 93–116. Retrieved from http://www.google.com/url?sa=t{&}rct=j{&}q={&}esrc=s{&}source=web{&}cd=1{&}ved=0CFkQFjAA{&}url=http://www.columbia.edu/{~}rmk7/PDF/K{%}26H.pdf{&}ei=157HT-fMM4TPhAfU5qC0Cw{&}usg=AFQjCNHZzh3m5L3{_}4EQ1Na5ce2FZ9gWkwQ{&}sig2=5-DHpjgqUxXJvfY5Bo2F2w doi: 10.1080/14417040600667293

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2015). Can you handle this? The impact of object affordances on how co-speech gestures are produced. *Language, Cognition and Neuroscience*, *3798*(March), 1–11. Retrieved from http://www.tandfonline.com/doi/full/10.1080/23273798.2015.1108448 doi: 10.1080/23273798.2015.1108448

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, *22*(4), 473–500. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/01690960600696916 doi: 10.1080/01690960600696916

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture*, *4(2)*(2004), 119–141. doi: 10.1075/gest.4.2.02mel

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2011). Seeing and being seen: The effects on gesture production. *Journal of Computer-Mediated Communication*, *17*(1), 77–100. doi: 10.1111/j.1083-6101.2011.01558.x

Morsella, E., & Krauss, R. M. (2004). The Role of Gestures in Spatial Working Memrory and Speech. *American Journal of Psychology*, *117*(3), 411–424. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15457809 doi: 10.2307/4149008

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. Retrieved from http://pss.sagepub.com/content/7/4/226.short doi: 10.1111/j.1467-9280.1996.tb00364.x

Rose, M., & Douglas, J. (2001). The differential facilitatory effects of gesture and visualisation processes on object naming in aphasia. *Aphasiology*, *15*(10), 977–990. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/

02687040143000339{%}5Cnhttp://www.informaworld.com/10.1080/02687040143000339 doi: 10.1080/02687040143000339

Shriberg, E. E. (1996). Disfluencies in SWITCHBOARD. *Proceedings International Conference on Spoken Language Processing*, 11–14.

So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science*, *33*(1), 115–125. doi: 10.1111/j.1551-6709.2008.01006.x

van der Sluis, I., & Krahmer, E. (2007). *Generating Multimodal References* (Vol. 44) (No. April 2014). doi: 10.1080/01638530701600755

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *American Journal of Psychology*, *114*(4), 591–600. doi: 10.2307/1423612

Table 1

Familiarity across experimental blocks.

	New	Heard	Described	Heard & Described	Heard in previous trial
Block 1					
Easy-to-name	220	99	118	0	0
Hard-to-name	220	106	110	0	0
Block 2					
Easy-to-name	0	117	62	244	11
Hard-to-name	0	53	64	118	206

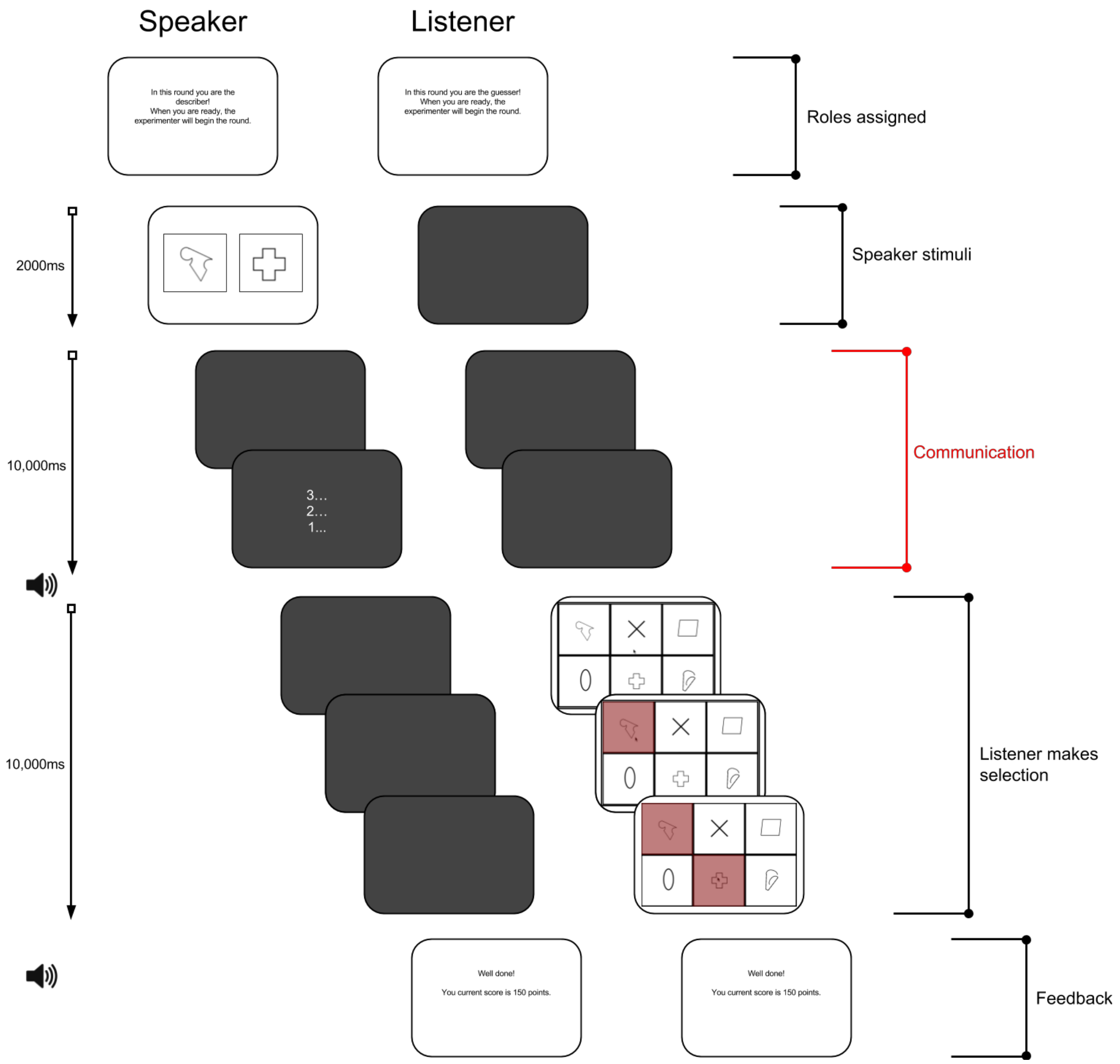

Figure 1. Procedure of a given trial.

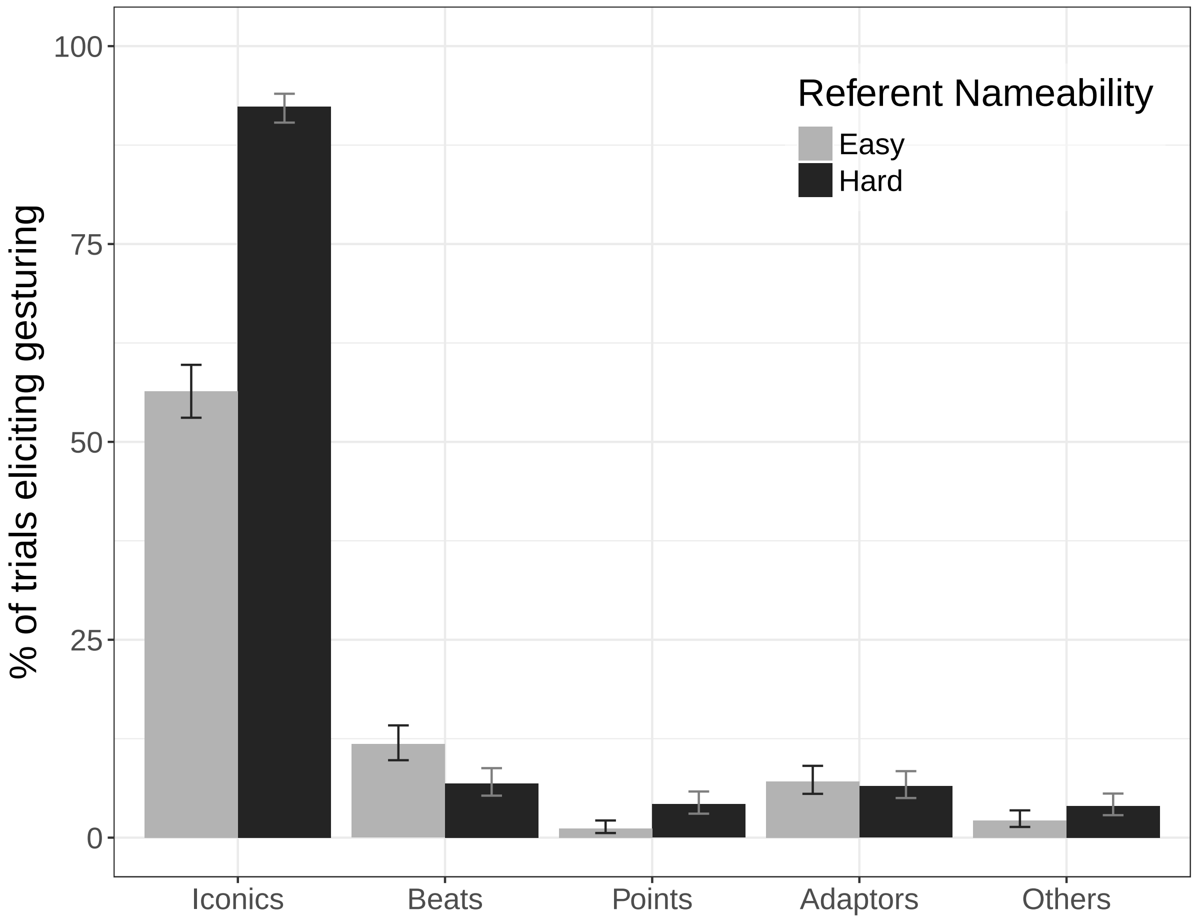

Figure 2. Occurrence of different types of gesturing by verbal-codability

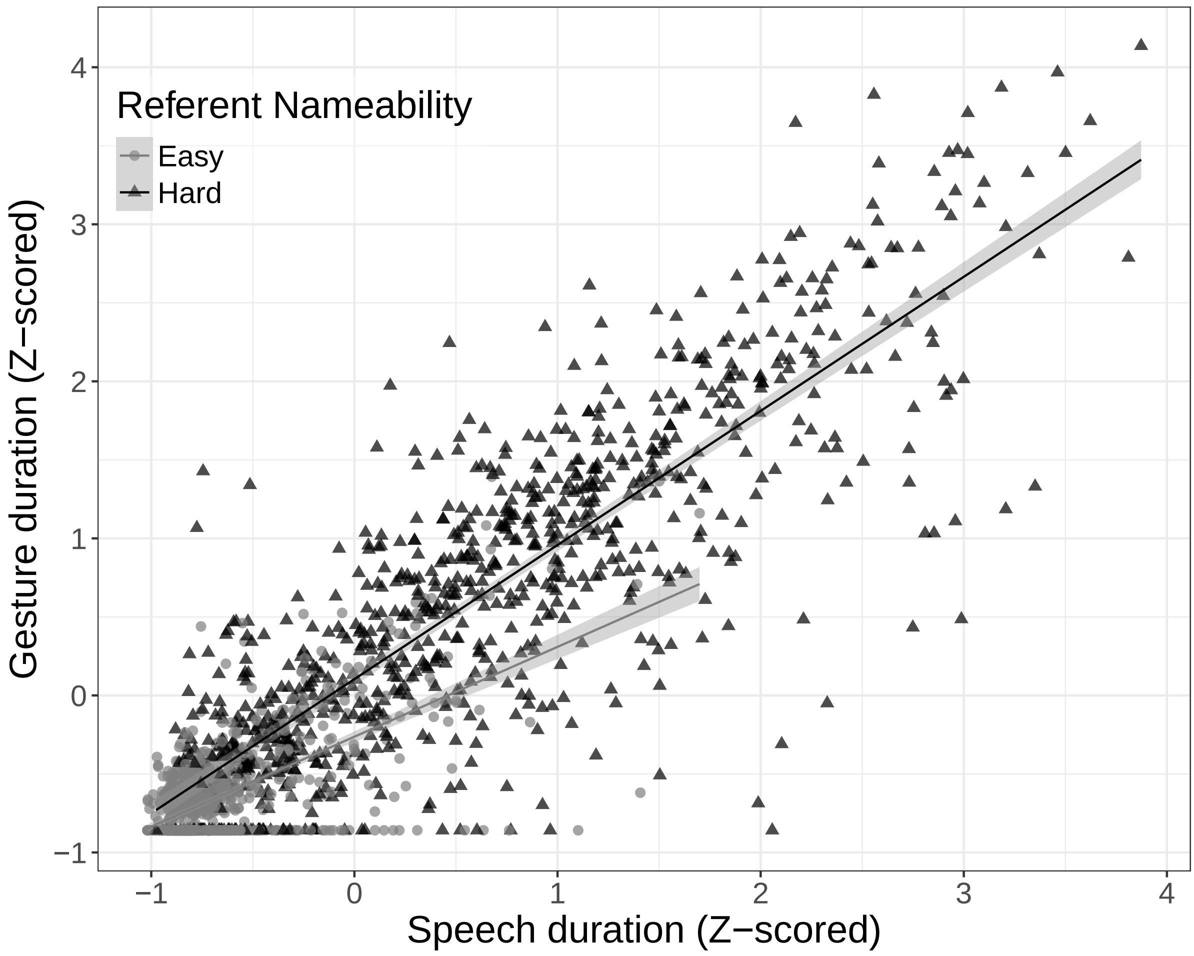

Figure 3. Relative durations of speech and gesture.

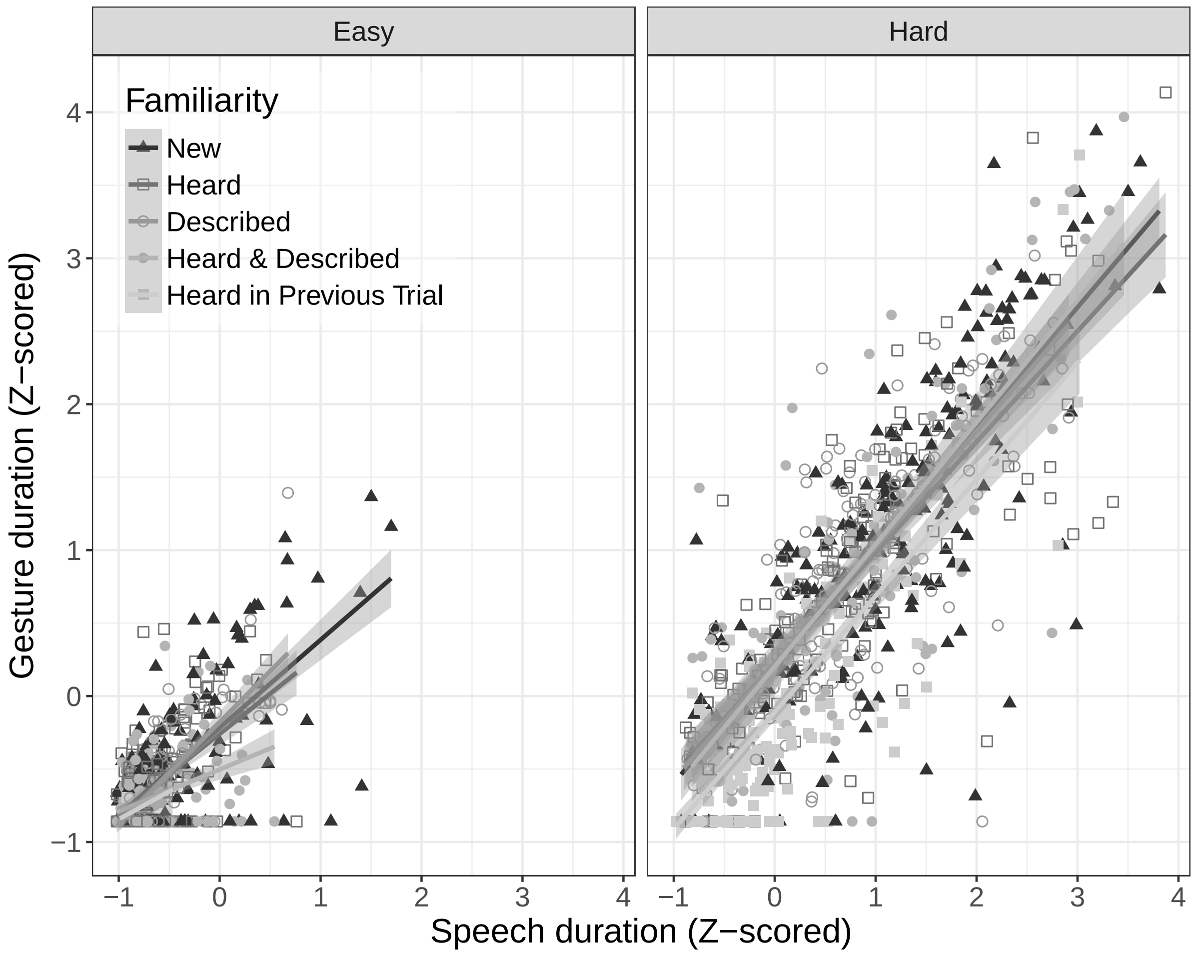

Figure 4. The speech-gesture relationship for different levels of familiarity