
This article is under copyright. The publisher should be contacted for permission to re-use or reprint the material in any form. For details, see the journal’s website:

https://benjamins.com/#catalog/journals/gest/main
Mapping out the multifunctionality of speakers’ gestures

Kasper Kok (VU University Amsterdam)
Vrije Universiteit Amsterdam. Department of Language, Literature and Communication.
De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands. k.i.kok@vu.nl

Kirsten Bergmann (Bielefeld University)
CITEC - Cognitive Interaction Technology. Faculty of Technology. Universität Bielefeld.
Inspiration 1, 33615, Bielefeld, Germany. kirsten.bergmann@uni-bielefeld.de

Alan Cienki (VU University Amsterdam & Moscow State Linguistic University)
Vrije Universiteit Amsterdam. Department of Language, Literature and Communication.
De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands. a.cienki@vu.nl

Stefan Kopp (Bielefeld University)
CITEC - Cognitive Interaction Technology. Faculty of Technology. Universität Bielefeld.
Inspiration 1, 33615, Bielefeld, Germany. skopp@techfak.uni-bielefeld.de

ABSTRACT

Although it is widely acknowledged that gestures are complex functional elements of human communication, many current functional classification systems are rather rigid, implicitly assuming gestures to perform only one function at any given time. In this paper, we present a theoretical view on the inherent multifunctionality of speakers’ gestures, inspired by frameworks in structural-functional linguistics (Halliday’s Systemic Functional Grammar and Hengeveld & Mackenzie’s Functional Discourse Grammar). Building upon this view, we report on a large-scale internet-based gesture perception study, designed in a way open to the potential for complex multifunctionality of gestural expression. The results provide quantitative support for the view that speakers’ gestures typically contribute to multiple semantic and meta-communicative functions of the ongoing discourse simultaneously. Furthermore, we identify clusters of functions that tend to be combined, as well as correlations between pairs of them. As a whole, this paper achieves some degree of convergence between ecological and experimental views on gesture functionality.

Kasper Kok, is a PhD candidate at VU University, Amsterdam. His research focuses on the incorporation of speaker’s gestures in cognitive and functional models of grammar.

Dr. Kirsten Bergmann is a post-doctoral researcher at Bielefeld University. Her research is on speech and gesture use in human-human- and human-computer interaction.

Prof. Alan Cienki is professor of language use and cognition at VU University, Amsterdam, and is director of the Multimodal Cognition and Communication Laboratory (PoliMod) at Moscow State Linguistic University.

Prof. Stefan Kopp is the head of the Social Cognitive Systems Group at Bielefeld University.
Introduction

Gestures are functional elements of human communication. Over recent decades, various classification systems have been devised to capture their diverse functionality (e.g. Efron, 1972; Lausberg & Sloetjes, 2009; McNeill, 1992; Wundt, 1973). Having enabled the currently flourishing research tradition of quantitative inquiry, these classification systems have undoubtedly been of great value to the field of gesture research. However, the categorization of gestures into discrete functional types implicitly draws on a rather dubious assumption: that gestures carry out only one function at any given time. Ecologically oriented gesture researchers have pointed out that such a ‘pigeonhole’ view on gestural functionality does not do justice to the multifarious nature of gesture use in natural discourse (e.g. Kendon, 2004; Streeck, 2009).

Although many gesture scholars acknowledge this problem, at least in theory, and are aware that the exhaustive classification of gestures into a small number of all-or-nothing categories is no more than a “convenient fiction” (Loehr, 2004, p. 128), operational alternatives are sparse. Less discrete views, such as McNeill’s (2005, pp. 40-41) proposal to regard iconicity, metaphoricity, deixis and temporal highlighting not as categories but as ‘dimensions’ of meaning that may be mixed in a single gesture, have rarely been adopted explicitly in quantitative research designs. Instead, solutions to the issue of gestural multifunctionality often go no further than the inclusion of mixed categories like ‘iconic-deictic’ or ‘beat-metaphoric’ in a coding scheme. Acknowledgement that gestures can be seen as having a primary function next to various secondary functions (Müller, 1998) addresses the basic theoretical issue at stake, but in practice, operationalization of this approach usually results in just coding the primary function of any given gesture, leaving the remaining functions in the dark. Consequently, it is not clear to
date how the (quantitative) extents to which gestures manifest different functional categories simultaneously can be analyzed in a systematic fashion.

In this paper, we attempt to achieve some degree of convergence between ecological and experimental views on gesture functionality. We first provide, as theoretical background, a brief overview of some influential (structural-)functional models of language and their application to gesture. Subsequently, we present the quantitative results of a large-scale internet-based inquiry into the capacity of gestures to perform different (sub)functions simultaneously.

(Multi)functionality in language and grammar

Functional approaches to language are characterized by the assumption that language structure reflects the goals of language users. Although no full consensus exists as to how these goals are best described, most functional accounts agree that language simultaneously serves a cognitive function (it organizes and represents thought) and a social function (it allows for coordination of one’s behavior with that of other people). This ambivalence was already recognized by Wilhelm von Humboldt (1903, p. 24), who concluded that “there lies in the primordial nature of language an unalterable dualism.”

Karl Bühler’s (1934/1990) Organon model, which has exerted great influence on modern functional linguistics, advances the view that language manifests a relation between three main components of communicative events: a Sender, a Receiver and a Referent. Linguistic functions can be characterized accordingly: aspects of an utterance that serve to describe objects, situations or mental states have a representational function; those that reveal the speaker’s emotional state or attitude towards the situation described have an expressive function; those that are aimed at
directing or affecting others’ behavior or mental state perform an appeal function. Crucially, these functions are not all-or-nothing, nor are they mutually exclusive. An important tenet in Bühler’s account is that all three functions are present to some degree in every utterance, albeit not always in equal salience. That is, any expression in some way reflects the Sender’s construal of the Referent, his or her affective evaluation of it, and the intended effect of the message on the Receiver.\(^1\) From this point of view, all linguistic utterances are inherently multifunctional.

The work by Bühler and his successors has inspired various models of language structure. Functional models of grammar (Dik, 1989; Halliday, 1985; Hengeveld & Mackenzie, 2008; Van Valin, 1993) hold that linguistic structures are best understood in relation to their functional contributions to the ongoing discourse. Systemic Functional Grammar (Halliday, 1995), one of the most elaborate models in this field, pursues the view that language structure reflects the interplay between different subfunctions. This theory describes grammar on three levels: in terms of its role in organizing experience (the ideational subfunction); in terms of the various social relationships that are acted out through language (the interpersonal subfunction) and in terms of the way these functions are combined into text or speech (the textual subfunction).\(^2\) These subfunctions are further realized by large number of ‘systems’, i.e. sets of choices available to speakers during language production. The eventual surface form of an utterance, accordingly, can be seen as emerging from the interplay of cognitive, social and contextual constraints on expression, which continuously compete for recognition during communication.

\(^1\) Utterances that do not have a referential meaning (e.g. expressives such as *Ouch!* ) are an exception to this.

\(^2\) Note that these three subfunctions do not map one-to-one on Bühler’s functions. The ideational function corresponds to Bühler’s representational function. Both the appeal and expressive functions would fall under interpersonal in Bühler’s model.
Functional Discourse Grammar (Hengeveld & Mackenzie, 2008), an incarnation of Dik’s (1989) Functional Grammar, also acknowledges that language structure simultaneously reflects pragmatic (Interpersonal) and semantic (Representational) factors. Based on extensive typological inquiry, FDG furthermore recognizes a number of hierarchically organized layers on each of these levels of analysis. On the semantic level, this layering reflects the hierarchical embedding of lower order entities (e.g. concrete, tangible entities) into higher-level entities (e.g. propositions and states of affairs). On the Interpersonal level, language is seen as actional by nature: the higher interpersonal layers correspond to the act of expressing a stretch of discourse (which may be linguistically marked by words such as in sum, or secondly); the lower layers correspond to the act of evoking entities or attributes. All non-representational aspects of linguistic meaning are analyzed at this level, including the expression of reduced commitment to the attested information (e.g., hedging-suffixes) and the assumed presence or absence of common ground (e.g., marked by the use of a definite or indefinite article).

Functional models applied to gesture

Various connections and overlapping assumptions exist between functional linguistics and work by Kendon (e.g. 1981, 2004), Streeck (e.g. 2009), Enfield (e.g. 2009) and many others. The work of these scholars concentrates predominantly on gestures’ meaning and function as situated in natural discourse. Nonetheless, explicit reference to, or adaptation of, the functional models discussed above are somewhat sparse. Müller (1998 Ch. 2; 2013, p. 214) takes on Bühler’s model to support the view that gestures can be understood in terms of their representational, expressive and appeal-related qualities. She maintains that gestures can fulfill each of Bühler’s
functions as their primary purpose: “gestures are in principle functionally comparable to verbal signs: […] they may re-present something other than themselves while at the same time expressing some inner state, being addressed towards somebody, and […] executing speech acts and other communicative activities.”

Only few concrete attempts have been made at incorporating gestures in functional models of grammar. Martinec (2004) and Muntigl (2004), adopting a Systemic Functional Linguistic view, have argued that some gestures can be analyzed in terms of a network of competing functional choices. The models they propose are rather programmatic, however, and based on too sparse data to validate the generalizability across gesture types. Taking Functional Discourse Grammar as a point of departure, recent work has further corroboration that a stratified functional model of language structure aligns well with the ways gestures have been studied in the literature (Connolly, 2010; Kok, in press). As demonstrated by Kok (in press), each of the main layers of semantic and pragmatic organization that are recognized by FDG can receive gestural expression. Higher level pragmatic layers can be modified by discourse-parsing gestures (Kendon, 1972, 1980, 2004) as well as interaction-management gestures such as those described by Bavelas (1992) and certain emblems (Efron, 1972); lower-level pragmatic layers are relevant for certain gestures that signal reduced commitment to some attribution or assign focus to a referent. On the Representational Level, gesture sequences that represent temporally coherent sets of states of affairs relate to higher layers (the Episode layer, in particular, cf. Müller, Ladewig, & Bressem, 2013), whereas lower semantic layers can receive (co-)expression by representational gestures, e.g. those that represent the shape or size of an object.

As argued above, however, these functions cannot be expected to map onto any particular gesture token in a one-to-one fashion. Motivated by this consideration, the remainder of this
paper presents an empirical examination of the ways in which, and the degrees to which, speakers’ gestures are multifunctional. The design of the study is loosely based on functional linguistic models such as SFG and FDG. Its focus lies on those representational and interpersonal functions that occur in route directions, a type of discourse that is known to involve descriptive as well as directive communication (cf. Denis, 1997; Lücking, Bergman, Hahn, Kopp, & Rieser, 2013). In particular, we focus on the functions listed in Table 1 and their (co-)occurrence in a large corpus of natural and spontaneous direction-giving dialogues, the Bielefeld Speech and Gesture Alignment corpus (SaGA, Lücking et al., 2013). This set of functions was selected on the basis of theoretical considerations (they relate to various levels of linguistic organization) and operational criteria (they have a high frequency in the corpus). That is, the list was prepared by assessing which of the functions that have been discussed in the literature on linguistic approaches to gesture, discussed above, are recurrent in the SaGA corpus. As a consequence, it is not exhaustive: the existence of additional functions of gestures needs to be acknowledged (e.g., their role in discourse segmentation and expression of illocutionary force; Kendon, 1995), but these do not occur frequently enough to be worthwhile including in the research design.

Table 1. Examples of (sub-)functions of speech that can be realized or modified gesturally and were included in the present study.

<table>
<thead>
<tr>
<th>Representational functions</th>
<th>Examples in speech</th>
<th>Examples in gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refer to a concrete, tangible entity</td>
<td><em>The book</em>; <em>she</em>; <em>Mary</em></td>
<td>Certain pointing gestures; gestures that represent an object; catchments (reiterated deictic reference to an object associated with an object)</td>
</tr>
<tr>
<td>Refer to a location</td>
<td>At your left-hand side; in London; here</td>
<td>Certain pointing gestures and catchments</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>Describe a physical property (e.g. size or shape) of an object or person</td>
<td>Big; round; shaped like the leaf of an oak</td>
<td>Gestures that depict the size or shape of a referent, e.g. by drawing its contours in the air</td>
</tr>
<tr>
<td>Describe the movement of an object or person</td>
<td>is spinning; turns left; rolls down</td>
<td>Gestures that trace an objects’ movement trajectory in space</td>
</tr>
<tr>
<td>Designate the amount of the referent</td>
<td>A; five; a couple of</td>
<td>Conventional number gestures (extending a certain number of fingers vertically); the use of one or two hand when referring to one or two objects.</td>
</tr>
<tr>
<td>Locating an event in (real or fictive) space</td>
<td>[the accident happened] over here; [we watched TV] in the living room</td>
<td>Gestures that depict an event in a certain region of space. Certain pointing and placing gestures (e.g., a pointing gesture co-occurring with the accident happened over here)</td>
</tr>
</tbody>
</table>

**Meta-communicative and textual functions**

<table>
<thead>
<tr>
<th>Function</th>
<th>Examples in speech</th>
<th>In gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal (lack of) commitment to the accuracy of some ascribed property</td>
<td>Ten-ish people Sort of big</td>
<td>Swaying gestures (e.g. wiggling the lax hands at the wrist); shoulder shrugs</td>
</tr>
<tr>
<td>Signal importance of some element of the utterance</td>
<td>Word order, prosodic prominence, it-clefts: It was John who went to the university</td>
<td>Gestural expression in general (Levy &amp; McNeill, 1992). In particular certain beat gestures (e.g. those deliberately used in political speeches).</td>
</tr>
<tr>
<td>Indicate that one is having trouble finding the right words</td>
<td>Filled pauses such as uh and uhm</td>
<td>Finger snapping; cyclic gestures</td>
</tr>
</tbody>
</table>

Note that, in accord with the functional linguistic tradition, the functions in Table 1 are defined relative to the utterance as a whole. Hence, this listing is not meant as a full taxonomy of gesture functions independent of speech, but as a (non-exhaustive) set of examples of functions of linguistic elements that can be (co-)performed or modified gesturally. In addition, and more crucially, these functions are non-exclusive, i.e., they can be expected to occur in combination.
In the following, we aim to empirically investigate the occurrences and correlations of the functions listed in Table 1. The conflation of different types of semantic information in a single gesture has already been examined in previous experimental research that employed a semantic feature-based approach (Beattie & Shovelton, 1999, 2001; Bergmann, Aksu, & Kopp, 2011; Bergmann & Kopp, 2006; Holler & Beattie, 2002). Others have demonstrated that the representational qualities of gestures are closely linked to pragmatic factors, such as the information structure of the spoken discourse (Levy & McNeill, 1992; Parrill, 2010), the negotiation of common ground (Holler & Wilkin, 2009) or the disambiguation of lexical elements (Holler & Beattie, 2003). However, to our knowledge, a more comprehensive and empirical investigation of gestural multifunctionality, whereby a wider set of functions is explored simultaneously and via analyzing the judgements of a large number of human non-expert raters, has not been carried out before.

Methods

A large set of recordings of gestures and co-occurring speech, attested in a relatively natural setting, was subjected to a web-based perception survey. Participants were presented with a set of video snippets and a number of statements about one gesture at a time, corresponding to each of the functions listed in Table 1 (e.g. *The hands of the speaker refer to a specific object or person*; see Table 2 for the full list of questions). Their task was simply to indicate, on a seven-point Likert scale, whether or not they agreed with each of these statements. Because the number of participants and the stimulus sample were rather substantive (462 stimuli, 18 raters per stimulus), this setup allows for gaining detailed, quantitative insights into the (multi-
functionality of single gestures, as well as into the patterns and clusters that exist across the stimulus set.

Materials

The stimuli were video snippets taken from the Bielefeld Speech and Gesture Alignment corpus (SaGA, Lücking et al., 2013). This corpus contains a large collection of German-spoken recordings of participants engaged in a route-description task, whereby one person explains a route through a virtual city to another person. For the current study, short video snippets, containing recordings of the person explaining the route, were extracted from five of these dialogues and used as stimuli. The start and end points of these snippets were determined by the nearest moment where the hands were in rest position or the speaker paused his or her speech (so that the stimulus videos contained relatively isolated units of discourse). Because the speakers often made successive gestures without returning the hands to rest position, some video snippets contained more than one gesture. Therefore, numbers appeared on the screen during the stroke phases, so that each gesture of interest could be referred to individually (see Figure 1). Crucially, no a priori filtering was applied to the stimulus set. That is, to prevent the content of the stimuli from being biased towards a certain functional type, virtually all gestures performed by the route givers were used as stimuli. The only exceptions were complex gesture sequences with more than six strokes performed in quick temporal succession. These were discarded to prevent the stimulus videos from being too long for the online survey.
The application of this procedure resulted in 174 videos containing a total of 462 individually marked gestures. The average length of the videos was 10.9 seconds (σ=3.50). For the current study, all videos were played with the sound on.

Participants

A total of 366 participants were recruited through the online platform Crowdflower\(^3\) and received a monetary reward for their participation. After applying the filtering procedures described in the section Performance Diagnostics and Filtering below, 260 participants remained. Their reported ages ranged between 18 and 69 (M = 38.2, σ=11.7). All were present in Germany at the moment of participation and reported to be fluent in German.\(^4\) The participants (reportedly) had diverse daily occupations; only 42 of them (16.2%) had an academic affiliation.

Procedure

To ensure that participants were able to play the video and sound, a visual and auditory ‘captcha’ was implemented where participants had to type in a word that was presented via audio and a word that was presented visually. Participants who passed this control were given the following instructions (translated from German):

In this survey you will be asked to answer questions on short video segments. The questions concern the relationship between the speaker’s gestures and what he/she says.

---

\(^3\) http://www.crowdflower.com

\(^4\) This was confirmed on the basis of ip-addresses. It is not very likely that speakers with low German skills participated, because they would have been unlikely to pass the test questions, which were asked in German.
Every list of questions concerns one single gesture. When multiple gestures occur in a video segment, they will be marked with numbers.

Please read the questions carefully and answer on the basis of your own perception. Be aware that for some videos none of the options is applicable, while for other videos, multiple options may apply.

In the video segments, someone explains a route through a city to his/her interlocutor.

You can play the video as often as you want.

Subsequently, participants saw a webpage with an embedded video on the top of the screen and the nine statements listed in Table 2 below. The following question was asked: ‘Do the following statements apply to gesture number X?’ where X was the number that appeared in the video in concurrence with the gesture of interest.\(^5\)\(^6\) On the right side of each question a 7-point Likert scale appeared with labels on both extremes: trifft sicher nicht zu ‘certainly does not apply’ and trifft sicher zu ‘certainly applies’.

Table 2. Statements presented to participants (translated from German).

<table>
<thead>
<tr>
<th>Question label</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depict-Shape</td>
<td>The hands of the speaker depict the size or shape of an object or person.</td>
</tr>
<tr>
<td>Signal-Prominence</td>
<td>The hands of the speaker show that the provided information is important or deserves special attention.</td>
</tr>
</tbody>
</table>

\(^5\) The pronouns were adjusted to the speaker’s gender.
\(^6\) Two additional questions were asked, but these were discarded from the analysis because of very low agreement among participants. The questions were: ‘The hands of the speaker show that the depicted object is unknown to the addressee’ and ‘The hands of the speaker show for how long the described event takes place.’
To get accustomed to the task, participants were first given a practice trial (the same for all participants). Subsequently, participants were presented with a block of twenty gesture videos, presented one at the time, which were randomly sampled from the total collection of stimuli. Because each block contained a different subset of the total collection (20 out of 462) and all analyses of interest are item-based, participants were allowed to take part in the study multiple times, with a limit of five. All redundant data points (cases where the same participant had by chance been assigned the same video twice) were excluded from the analysis post-hoc. In order to be eligible for the monetary reward, participants were required to complete the entire block of twenty stimuli. The order of the questions was randomized for each stimulus. For one out of every four videos on average, a test question was added to the survey to ensure that participants were actually reading and paying attention to the questions. These test questions simply asked to tick one of the seven boxes (e.g., the second one from the right).

**Results**

*Performance diagnostics and filtering*
Because the internet-based method that was employed lacks full experimental control, we undertook several steps to validate the reliability of the data. Four performance thresholds were established to filter out those participants who gave a strong impression of having completed the task without paying attention to the instructions or taking the assignment seriously. Trials were excluded if the participant (1) had failed to respond correctly to 20% or more of the test questions, which simply asked them to tick one of the seven boxes, (2) had taken less than seven minutes to complete the entire survey (average completion time was 30.3 minutes, $\sigma=13.2$), (3) had a variance in their ratings of less than 0.60 ($M=1.98$, $\sigma=.43$), or (4) had given only one or two unique ratings to all questions for five stimuli or more (non-attentive internet participants are known to complete surveys by clicking consistently on just one or two boxes). After having filtered out the participants who had failed these criteria, new participants were recruited in order to balance the number of trustworthy participants per stimulus. This procedure was repeated until exactly 18 participants had judged each video.

To get an impression of the reliability of this sample size with respect to larger populations, we analyzed the mean scores on each of the nine questions with respect to the practice video, which had been assigned to all 260 participants. A Pearson test revealed that the average ratings on each of the questions after 18 participants provide a good estimate of the average scores as they stabilize over time. That is, the mean ratings of the first 18 participants on the practice item correlates strongly with the mean ratings of all remaining 242 participants on the same item ($r=.96$, $p<.001$).

_The multifunctionality of a single gesture_
Before examining the more general patterns that occur in these data, we show how the functionality of a single gesture can be characterized according to the current approach. In example (1) and Figure 1, the speaker talks about two towers above the side aisles of a church. Meanwhile he moves both hands up, with the palms of the hands facing each other, all fingers curved. The timing of the gesture relative to the speech is represented below the speech transcription, following the conventions described by Kendon (2004 ch. 7), and the capital letters above the transcript correspond to the timing of the video stills.

Figure 1. A gesture performed concurrently with the utterance in (1)

(1) die Kirche hat halt ein ein Spitzdach und zwei Türme an diesen Seitenschiffen
LH/RH

‘the church has well a a pitched roof and two towers at those side aisles’
Figure 2 shows the means and standard errors of the ratings assigned by the participants to each of the nine function-questions with respect to this stimulus. With regard to four out of nine potential functions, there is a high degree of certainty among the 18 participants (mean rating 5 or higher). That is, strong consensus exists that the gesture refers to an object, refers to a place, describes the size or shape of an object, and provides information with respect to its amount. In addition, there is rather strong agreement that the gesture signals prominence, i.e. indicates that the provided information is noteworthy (mean rating > 4.5). The remaining four functions consistently receive low scores: participants are generally certain that the gesture does not show that the speaker is having trouble finding words, is uncertain about what he says, depicts movement or localizes some event in space.

7 The treatment of Likert-like items as continuous data requires a note of caution. Because the underlying construct might in fact not be continuous, no absolute comparisons can be made between the distances between pairs of survey scores (the distance between a mean score of 2 and 3 on a certain question is not necessarily comparable to the distance between two gestures that score 4 and 5 on the same question).
These data suggest that the gesture in question is rich in meaning. If the same functionality were to be expressed verbally (e.g., in German), at least five lexical or grammatical elements may have been needed: one for referring to an entity (a noun phrase or pronoun, presumably), one for indicating its amount (e.g. a numeral or inflection), one for referring to its location (e.g. an adverb of place or a prepositional phrase), one for describing its shape (e.g. an adjective) and one for marking that the given information is noteworthy (e.g. a word order or intonation contour associated with discursive prominence).\textsuperscript{8} In the following, we take this layered-gradient view on gesture functionality as a starting point to look at the commonalities that exist between the functional profiles of the 462 gestures for which we have gathered comparable data. In particular, based on correlational patterns in the data, we look further into the general tendencies of specific functions to be co-expressed.

\textit{Mapping out the gesture functionality space}

In order to obtain a global overview of the relations between gestural (sub-)functions, we applied Principal Component Analysis – a technique for reducing the complexity of high-dimensional data by mapping them onto the axes of biggest variance. The first three principal components, as Table 3 shows, explain about 78\% of the total variance within our data. The difference in

\textsuperscript{8} Note that the gesture \textit{was} in fact performed together with speech, and from the transcription we can see that speech and gesture are largely co-expressive in terms of all functions described. Only the shape-depiction aspect of the gesture is not explicitly mentioned verbally (but one may argue that it is implicit in the meaning of the word \textit{tower}).
informativeness between the third and subsequent components is relatively marginal. The plot in Figure 2 displays the Eigenvector-rotated values of all gesture stimuli on the first two principal components as points, and the coefficients of the survey questions on these components as vectors. Generally speaking, question-vectors pointing in the same direction have a similar response profile across stimuli, and points projected in the direction of any of the vectors represent gestures with high scores on the corresponding questions.

Table 3. Loadings of all variables (the survey questions) on the first three principal components

<table>
<thead>
<tr>
<th>Question</th>
<th>Comp.1</th>
<th>Comp.2</th>
<th>Comp.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal-Prominence</td>
<td>-0.42</td>
<td>0.10</td>
<td>-0.25</td>
</tr>
<tr>
<td>Localize-Event</td>
<td>-0.28</td>
<td>-0.40</td>
<td>-0.38</td>
</tr>
<tr>
<td>Refer-to-Object</td>
<td>-0.33</td>
<td>0.42</td>
<td>-0.14</td>
</tr>
<tr>
<td>Refer-to-Place</td>
<td>-0.41</td>
<td>-0.21</td>
<td>-0.24</td>
</tr>
<tr>
<td>Depict-Shape</td>
<td>-0.23</td>
<td>0.51</td>
<td>-0.03</td>
</tr>
<tr>
<td>Depict-Movement</td>
<td>-0.21</td>
<td>-0.45</td>
<td>-0.19</td>
</tr>
<tr>
<td>Number</td>
<td>-0.18</td>
<td>0.36</td>
<td>-0.26</td>
</tr>
<tr>
<td>Signal-Uncertainty</td>
<td>0.40</td>
<td>0.13</td>
<td>-0.56</td>
</tr>
<tr>
<td>Word-Search</td>
<td>0.41</td>
<td>0.05</td>
<td>-0.55</td>
</tr>
<tr>
<td><strong>Variance explained</strong></td>
<td><strong>41.3%</strong></td>
<td><strong>25.3%</strong></td>
<td><strong>11.1%</strong></td>
</tr>
</tbody>
</table>
Figure 2. A scatter plot of the rotated mean scores of all gestures on the first and second component, and the loadings of all questions on these components plotted as vectors.

Some noticeable patterns emerge from this analysis. For one, the spatial organization of the question-vectors suggests that the gestural ‘functionality space’ comprises three, somewhat orthogonally organized clusters. The first can be characterized as a *representational* dimension, pertaining to the capacity of gestures to refer to objects and their intrinsic properties such as size, shape and number. This dimension subsumes the survey questions Depict-Shape, Refer-to-Object and Number. The second cluster roughly represents a *spatial* dimension of gesture
meaning, corresponding to gestures’ capacity to localize objects and events in space. The third can be described as *meta-communicative*, subsuming the questions Signal-Uncertainty and Word-Search.\(^9\) The only survey question that does not clearly fall within any of these three clusters is Signal-Prominence. The capacity of gestures to indicate that some information is noteworthy correlates with *both* the representational and spatial features (see below for a more detailed analysis of these correlations), but appears orthogonal to their potential to signal uncertainty or word search.

Another noteworthy observation is that the gesture stimuli are widely dispersed throughout the entire plot. This suggests that although some functional clusters exist, most gestures fall right in between these. This is in line with McNeill’s (2005) argument that many gestures simultaneously combine iconic, deictic and pragmatic features. In the next section, we examine some of the relations between the scores on some of the individual questions in more detail.

*A closer look at semantic multifunctionality*

A first type of multifunctionality concerns patterns of co-occurrence between different types of semantic information. Figure 3a displays the mean scores of all gesture stimuli on the question Depict-Shape as a function of the mean scores on Refer-to-Object. Figure 3b displays the mean scores on the question Depict-Movement plotted against the scores on Depict-Shape.

---

\(^9\) Note that in most functional linguistic models, the signaling of word search and the display of uncertainty belong to different functional categories. To the extent that the signaling of word-search is aimed to warn the addressee of an impending delay of the discourse, as in the case of interjections like *uh* and *um* (Clark & Fox Tree, 2002), the broad category label ‘meta-communicative’ used here covers Halliday’s interpersonal *and* textual functions.
Figure 3. Scatter plots of mean scores on two pairs of semantics-related questions.

A Pearson correlation test reveals a strong positive trend between the mean answers on Depict-Shape and Refer-to-Object ($r_{460} = .78$, $p < .001$). From the scatter plot, however, it appears that this relation is not fully symmetrical. Whereas none of the gestures in the data set were judged to depict the shape of an object without also referring to a concrete entity, there are some cases of gestures that score high on Refer-to-Object but not on Depict-Shape. Qualitative inspection of this subset of the stimuli, indicated visually in the figure by the dashed ellipse, reveals that this category includes many instances of abstract deixis: gestures that refer to verbally described objects by pointing to a location in interactional space associated with a discourse referent. The reverse pattern – high scores on Depict-Shape with low scores Refer-to-Object – does not occur: according to our participants, those gestures that evoke a physical or spatial attribute necessarily also make reference to some object or person; gestures were never perceived as isolated attributes.
In Figure 3b, we see a rather different picture: there is a weak negative trend between the questions Depict-Size and Depict-Movement ($r_{460}=-.14$, $p=.002$). In line with this trend, the region of the plot corresponding to high scores on both questions is empty. This indicates that the stimulus set did not contain any instances of gestures that were judged to simultaneously depict the shape and the movement of an object. Although there are no reasons why such gestures couldn’t exist in principle (one may imagine a gesture whereby the handshape refers to a pen which is moved through space to represent writing), none of the gestures in the natural spatial dialogues under investigation were judged to have these characteristics.

A closer look at semantic-interpersonal multifunctionality

The semantic multifunctionality analyzed in the previous section pertains to only one of the levels of analysis distinguished in functional linguistics: language’s representational function (or ideational subfunction, in SFG terms). However, there are reasons to believe that gestures additionally often conflate representational (semantic) and interpersonal (pragmatic) functions. In this section, we investigate two of such relations as they occur in our data.
We first look at the question of whether referential gestures are necessarily perceived as indicating discursive prominence. As Figure 4a shows, there is an overall positive correlation between the scores on the corresponding questions ($r_{460} = .59$, $p < .001$). The majority of gestures that were judged to refer to an object or person were also judged to indicate that this act of reference has focal status. This finding corroborates Levy and McNeill’s (1992) hypothesis that gestures are an expression of high communicative dynamism (i.e., they contribute to ‘pushing the communication forward’). A relevance theoretic view (cf. Sperber & Wilson, 1986) on iconic gesturing can be useful to frame this finding: the use of the hands to refer to an object appears to create the expectation of its own relevance as a contribution to the interaction.

**Figure 4.** Scatter plots of mean scores on the questions Signal-Prominence and the questions Refer-to-Object (a) and on the questions Word-Search and Depict-Shape (b).
Finally, we explore the relation between spatial deixis and meta-communicative signaling. In Figure 4b, we see that a negative correlation exists between Word-Search and Refer-to-Place ($r_{460} = -0.50$, $p<0.001$). Thus, the potential of gestures to refer to a location is seldom combined with signaling that the speaker is having trouble finding the right words. As is already apparent from the results of the Principal Component Analysis and Figure 2, the abilities of the hands to refer to a place and to express ongoing management of one’s own speech appear mutually exclusively in our data (as a general trend, at least). Note however, that a few exceptions to this pattern exist: we can see from Figure 4b that some gestures received a modestly high mean score ($5\pm$) on both questions.

**Degrees of functional prominence**

As is evident from the wide dispersion of the data in Figures 3, 4 and 5, a substantial variability exists in the degree of certainty with which the different functions were ascribed to the gestures. In fact, the majority of the mean scores on the questions taken into account falls right in between the ‘certainly’ and ‘certainly not’ poles. Whereas this may to some extent be explained by interpersonal differences in the interpretation of the gestures, it also suggests that the different functions carried out by any given gesture have different degrees of prominence to the observers. Any given gesture may for instance foreground a certain type of information (e.g. the shape of some object) but simultaneously provide information that is of secondary importance (e.g. the location of the object).

This variable prominence may have important ramifications for classification and coding schemes, since it calls for a clear operational definition for regarding a given function as
‘present’ or ‘absent’. In Table 4, we quantify the multifunctionality of our stimuli assuming a relation between the degree of certainty expressed by the raters and the salience of a gesture’s function: questions with mean scores higher than 5 on the 7-point Likert scale are assumed to correspond to a gesture’s primary functions; those with mean scores between 4 and 5 and those with mean scores between 3 and 4 are classified as secondary and tertiary, respectively. Table 4 shows how many of such functions were attributed to the gestures in our stimulus set.

**Table 4.** Number of primary, secondary and tertiary functions per gesture according to degrees of certainty among raters.

<table>
<thead>
<tr>
<th>Function type by rater certainty</th>
<th>Number of functions (as defined on the left) per gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Primary’ functions (mean score &gt; 5)</td>
<td>M=0.87, σ=.97</td>
</tr>
<tr>
<td>‘Secondary’ functions (mean score between 4 and 5)</td>
<td>M=1.74, σ=1.17</td>
</tr>
<tr>
<td>‘Tertiary’ functions (mean score between 3 and 4)</td>
<td>M=2.81, σ=1.63</td>
</tr>
<tr>
<td>Accumulated</td>
<td>M=5.42, σ=1.53</td>
</tr>
</tbody>
</table>

We see that the gestures in the corpus were typically assigned no more than one primary function. Hence, the gesture in Figure 1, which has four primary functions according to the definitions employed here, appears more of an exception than a rule. With regard to secondary and tertiary functions, however, we find that multifunctionality is a much more frequent, if not ubiquitous phenomenon. When accumulating over the three categories, we find an average of over 5 functions per gesture; only 1 out of 462 gestures was ‘unifunctional’ according to these
criteria. Of course, the exact numbers in Table 4 are not very meaningful, as they depend on the amount of questions included in the survey and the way these have been asked (as well as on the operational choices made – recall for instance footnote 7 on the caveats of treating Likert-items as continuous). However, the general pattern in these data underscores another important characteristic of gestures that has often been neglected in experimental work: gestural functions are not all-or-nothing, but come in different degrees of explicitness and salience.

**Discussion and conclusion**

Inspired by objections to the rigidity in current functional classification systems, this paper has advanced, operationalized and empirically substantiated the view that gestures are potentially multifunctional. The results of a large-scale gesture perception study suggest that the functional potential of the gestures judged in the direction-giving discourse segments used in this study involves at least three, somewhat orthogonal components – one pertaining to *reference and representation of objects*, one pertaining to *space and movement*, and one pertaining to *meta-communicative signaling*. Some of these functions can be present simultaneously in a single gesture, but with different degrees of salience.

Note that the three clusters of functions that emerged from our analysis do not strictly reflect Bühler’s three categories. Hence, our data do not directly corroborate Müller’s (1998) claims about how gesture functions can be characterized in the same way that Bühler characterizes the functions of words. On the other hand, given the operational details of the study, our results do not present direct counter evidence to such a view either. As mentioned, the questions in the survey were tailored according to the availability of relevant data in direction
giving dialogues, and consequently not fully geared towards Bühler’s second and third categories in the first place.

Some important notes of caution are in order when interpreting the data presented in this paper. First, the results are certainly not fully independent of the research design and the way the survey questions have been formulated. The set of questions included in our survey may not reflect the full functional potential of gestural expression. In addition, all stimuli came from a route description corpus – a discourse type that involves a relatively large number of concrete referential and spatial gestures and may not be fully representative of everyday dialogue.

With respect to the interpretation of the results, it is furthermore important to emphasize that this study involves the de-contextualized perception of gestures by an idealized comprehender. Raters were unconstrained in time and were allowed to watch the video more than once, while being uninformed about the exact local discourse situation and preceding utterances. Therefore, and because scores were averaged over a group of raters, we should be cautious to infer that the functional profiles of the gestures as investigated here correspond to the real-time processing of the addressee at the moment of the conversation. The characterization of the gestures here is better comparable to canonical (functional) linguistic analysis; it involves a level of description that is abstracted from actual usage and generalized over subjective experiences.

One way of further triangulating the exploratory results presented here, as well as to pinpoint the viability of the individual functions and their combination, is to operationalize them in a predictive or generative model. This would afford simulation in artificial communicators like virtual characters or humanoid robots (Bergmann & Kopp, 2009). A component-based model of
the multifunctionality of gestures also allows for testing the degree to which patterns found in the current study reflect actual on-line perception and interpretation. Likewise, employing virtual agent stimuli would enable exploration of the effects of systematic, possibly minor manipulations in the agent’s gestures on the ascription of different functions, thereby enabling a further refinement of the model.

Notwithstanding the caveats mentioned, the current contribution is among the first to provide quantitatively supported insights into gestures’ pervasive multifunctionality. By exemplifying how a more gradient and layered view of gesture function can be operationalized, it potentially has methodological implications, moreover. A setup akin to the one described here could lend itself to implementation in an experimental design: coding schemes can be endowed with more gradient scales (to allow coders to express different degrees of certainty), and annotation tiers that reflect the layers of a theoretically motivated, stratified model of language – to better capture the gestures’ semiotic complexity. Primary functions of gestures could be determined using a system such as the one described in Bressem et al. (2013), which goes from describing gestural forms to looking at the forms in relation to their functions with respect to the accompanying speech. Their system approaches functions in terms of semantic and pragmatic categories, whereby the former include representational and spatial functions, and the latter include meta-communicative ones (as well as others). Overall, it is the hope that the contents of the present paper, in addition to their empirical value, can inspire quantitative methods to adopt a more refined view on the function(s) of gestures in situated discourse.
Acknowledgements

The first author is grateful to support from the Netherlands Scientific Organization (NWO; grant PGW-12-39) and the German Academic Exchange Service (DAAD; scholarship 91526618 - 50015537). The third author is grateful for research support from Russian Science Foundation grant #14-48-00067. Moreover, this research received support from the German Research Foundation (DFG) in the Collaborative Research Center 673 “Alignment in Communication” and the Center of Excellence 277 “Cognitive Interaction Technology” (CITEC).

References


Lücking, Andy, Bergman, Kirsten, Hahn, Florian, Kopp, Stefan, & Rieser, Hannes. (2013). Data-based analysis of speech and gesture: The Bielefeld Speech and Gesture Alignment Corpus (SaGA) and its applications. *Journal on Multimodal User Interfaces, 7*(1-2), 5-18.


