

Dimensionalizing co-speech gestures

Stefanie Shattuck-Hufnagel¹, and Pilar Prieto^{2,3}

¹Speech Communication Group, Research Laboratory of Electronics, MIT, Cambridge MA USA

²Institució Catalana de la Recerca i Estudis Avançats, ICREA, Barcelona, Catalonia

³Dept. of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Catalonia

ABSTRACT

The manual gestures that accompany spoken utterances come in many different forms, and serve a number of different purposes. McNeill's classic (1975) descriptive types, with referential categories (iconic, metaphoric and deictic) distinct from a rhythmic category (beat-like), have become standard, but this unidimensional view is inconsistent with parallel observations that referential gestures can have a rhythmic component aligned with spoken prominence, and beat-like gestures can also signal pragmatic meanings. Some researchers have argued against a 1:1 relationship between gesture form and function, and proposed to 'dimensionalize' analysis of co-speech gestures, on the basis that all gesture types may have both a prosodic component and a pragmatic component, and can express a range of functions. Here we argue for a dimensionalized system for labelling co-speech gestures as well as their prosodic and pragmatic characteristics, and illustrate how such a system can reveal significant aspects of the speech-gesture relationship.

Keywords: co-speech gesture, gesture labelling, beats, dimensionalization, discourse structure

1. INTRODUCTION

It has long been recognized that when a speaker produces an utterance, many other parts of the body besides the vocal tract are in motion, and that some of these extra-vocal-tract movements are related to the speech act, contributing to the listener's understanding of the speaker's intent (see Kendon [1], [2] for reviews). Co-speech gesture has been observed to relate to the speaker's communicative intent in a wide variety of ways, including, for example, visually illustrating an aspect of the propositional content of the speech or adding components to the propositional content, enhancing rhythmic prominences in the speech, marking aspects of discourse structure, signalling pragmatic and semantic meaning, signalling aspects of the social interaction such as politeness, etc. (Kendon [3], [1], [2]; McNeill [5], [6]; Prieto et al. [7] and many others). This wide range of observations calls

for the analysis of large gesture/speech samples, to quantify aspects of their form and function, and to test emerging hypotheses.

To facilitate such analyses, McNeill and colleagues [4], [8] proposed a categorization scheme for labelling co-speech gestures, which divided gestures into various types. These included (roughly speaking) referential gestures (such as iconic, metaphoric and deictic), which visually illustrate an aspect of the spoken utterance, and non-referential gestures, called beats, generally simpler in form and aligned with important words. This categorical scheme allowed McNeill and colleagues to categorize all of the gestures in their samples, and inspired a rich set of subsequent studies using the proposed typology, which has since become standard methodology in the field. This typology was largely based on the relation between a gesture and the propositional content of the speech (referential gestures), but in part also on the prosodic structure of the speech (beats, sometimes described as marking out the rhythm of the speech). There were also a number of supplemental categories, such as emblems (conventional gestures with fixed meanings) and cohesives (gestures indicating relationships among parts of a communicative interaction), some of which appeared to relate to pragmatic meanings, such as aspects of discourse structure. Thus, from the earliest days of its widespread application, the gesture categorization proposal has been accompanied by a parallel set of observations and hypotheses that undercut---or perhaps better, enrich---the unidimensional categorizing approach. Many examples in the literature suggest that a sharp distinction between meaning-based referential gestures aligned with meaning-based elements in the speech, on the one hand, and rhythm-based beat gestures aligned with prominences in the speech, on the other, may need re-thinking. This is because many gestures have both prosodic and meaning-related characteristics, suggesting the need to label a number of different dimensions for each gesture rather than allocating it to a single category.

McNeill himself has suggested that a dimensional rather than a categorizing approach would be an

improvement, noting that “none of these ‘categories’ is truly categorical. We should speak instead of dimensions... iconicity, metaphoricity, deixis, ‘temporal highlighting’... social interactivity... The essential evidence that these semiotic properties are dimensional and not categorical is that we often find iconicity, deixis and other features mixing in the same gesture” ([5], pp. 41-42; see also [9]). However, this insight has not been generally incorporated into standard labeling conventions. In this paper we provide explicit examples of how referential gestures often align with spoken prosody, and how beat-like gestures do not always do so. Such observations support the need for a dimensionalized system for labeling co-speech gestures, to explore the hypothesis that all gesture forms have the potential both to reflect prosodic characteristics and to signal meanings, whether propositional or semantic-pragmatic.

The following two sections of this paper briefly summarize some of the empirical evidence supporting the twin arguments that (a) the referential/non-referential distinction is not equivalent to non-prosodic/prosodic; instead, referential gestures can sometimes be prosodic, and non-referential gestures are not always purely prosodic, and (b) the referential/non-referential distinction is not equivalent to meaning-bearing/non-meaning-bearing; instead, non-referential gestures can sometimes signal meaning.

We propose that all gesture types can in principle associate with two broad and independent dimensions, namely (a) rhythmic and prosodic prominence properties; and (b) a variety of semantic and pragmatic properties. While we adopt McNeill’s original beat or rhythmic dimension, we argue that it is more complex than originally envisioned. We include the iconicity, metaphoricity, and deixis components in the referential dimension, focussed on propositional content, and propose, in concert with the literature, an additional broad ‘semantic-pragmatic dimension’, to include non-referential aspects of meaning e.g. pragmatics and discourse structure. On this view, prosodic and pragmatic characteristics constitute independent layers which can be added to all gesture types. Examples are extracted from a study of American English public speech, where co-speech gestures of public discourse sessions were coded. Prieto et al. [7], examined a 16 minute TED talk (given by David Keith on Sept 26 2007). Similar findings were noted in Shattuck-Hufnagel and Ren (2018).

2. EMPIRICAL EVIDENCE

The following sections provide evidence that referentiality and rhythmicity do not define distinct types of gestures, which argues for a dimension-based annotation system that allows for the capture of a wide range of combinations of meaning and rhythmicity.

2.1. Gestures cannot be neatly divided into those that are prosodic vs those that are referential

2.1.1. Referential gestures (iconics, metaphoric, and deictics) often align with prosodic prominence in speech.

For beat gestures, the relation to spoken prosody is often quite clear. Many studies have shown not only that the most prominent part of a beat gesture often temporally associates with a pitch-accented prominent syllable in speech ([10][11][12]). However, studies have also shown that other gesture types such as iconics (e.g., [13]), and pointing gestures can do the same (e.g., [14][1]). Figure 1 shows one such example extracted from [7] which illustrates the iconic gesture for “sunk below” associated with prosodic prominence (expressed through pitch accentuation) in speech.

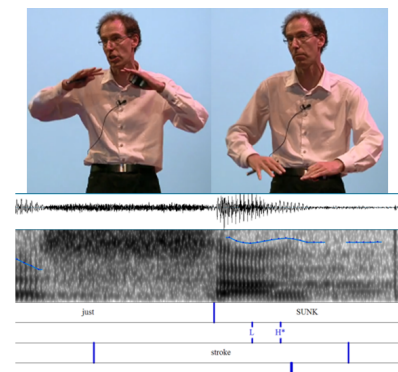


Figure 1: Iconic gesture prosodically associated with a pitch accent “it just **sunk** below...” (**bold** indicates pitch accented syllable). Figure extracted from [7].

2.1.2. Non-referential gestures (beats) do not always align with prosodic prominences

Beat gestures are not invariably associated with prominent positions in speech. For example, recent work by Rohrer et al. (submitted to this conference), reporting on a gestural and prosodic analysis of a French TED Talk, shows that from a total of 654 non-referential beat gestures, only 56.11% were strictly temporally aligned with a pitch accented syllable. Further inspection of the non-aligned beats showed that over 87% of them were associated to non-prominent syllables at edge-initial AP

boundaries, showing that beats in languages other than English might be non-rhythmic in nature. This ‘misalignment’ of (beat) gestures may occur also during lexical searches (Butterworth beats, as per Stam [15]), in speech errors, hesitations, or restarts, etc.

Taken together, these results highlight the fact that more work is needed on the prosodic affiliation of both referential and non-referential gestures, and their potential role in marking higher-level prosodic constituents and speech production planning.

2.2. Non-referential gestures can sometimes bear meaning as well

The evidence presented in section 2.1 suggests that referential gestures can sometimes have a rhythmic or prominence-related quality. Conversely, what have been termed beats or non-referential gestures can sometimes signal important information beyond the marking of spoken prominence. Many authors have highlighted the fact that beat-like gestures can signal pragmatic functions within discourse. One of the clearest of these is focus highlighting and information status marking (e.g., Loehr [13]: p.84; see also [1], [5]). McNeill ([5]: p.15) stated that a beat gesture, in this case defined as a single ‘flick’ of the hand or finger, “indexes the word or phrase it accompanies as being significant [...] for its discourse pragmatic content” and that beats can mark “the introduction of new characters, summarizing the action, introducing new themes, etc.” Shattuck-Hufnagel and Ren [10] also reported such overlap in a study of more than 1300 gestures in a sample of American English academic-style speech, labelled as Referential, Prosodic/Rhythmic or Mixed. Their results showed that a number of gestures were labelled as Mixed, i.e. were seen as having both Referential and Prosodic characteristics.

The examples in Figure 2 from Prieto et al. also show that beat-like gestures can convey a variety of intentions and illocutionary forces which can distinguish between asserting, ordering, exclaiming, etc. The two panels show two examples that illustrate the epistemic functions of beat-like gestures. While the open palms facing the speaker (left panel) show low certainty and low degree of imposition, fist beat gestures (right panel) represent high certainty and high imposition.



Figure 2: Left panel: “oh **yes**, I think it’s bad” referring to climate change. Right panel: “That **report** that landed on President Johnson’s desk...” (**bold** indicates pitch accented syllable).

While the hand fist gesture illustrated in the right panel of Figure 2 conveys a directive/imperative intent which expresses the goal of inducing action for the addressee, other hand gestures can express stronger or milder degrees of the directive function such as expressing a wish, to permit, concede, offer or invite, among others. For example, [16] found that the use of a raised index finger in political discourse enacts a strong directive intent involving a higher power status. [17] showed how the “precision grip gesture” in political discourse enacts not only information focus but also a second-degree pragmatic resource for performatively “making a sharp, effective point”. From a perception point of view, [18] showed experimentally that different types of hand gestures strongly affect the evaluation of persuasion in speech. Specifically, they found that hand gesture type affected four measures of the audience evaluation, namely speaker composure and competence, speaker communication style effectiveness, and message persuasiveness. Discourses with referential gestures (e.g., iconic, metaphoric, and deictic gestures) and discourses with non-referential beat gestures (cohesive and rhythmic) were better than non-linked-to-speech gestures (adaptors) or no-gestures in affecting message evaluation and judgements about the speaker by the receivers. In sum, a variety of epistemic effects are obtained from the use of non-referential hand gesture types such as the fist or the precision grip gestures, and more systematic work is needed on the potential effects of the use of these gestures on the discourse epistemic assessment. Such studies will be enabled by a labelling system which separately annotates the several dimensions.

Some authors, like Bavelas & Chovil [19], have claimed that hand gestures can be used to signal interactive conversational meanings such as common ground. They describe interactive hand gesture produced with the index fingers, marking knowledge that is already shared (the equivalent of “as you know”; see also [20: p. 395]).

Another area of pragmatics which is strongly affected by gesture realization is politeness. In English for example it has been shown that in conversation adults display an array of gestural mitigation cues such as raised eyebrows, direct body orientation, a tense, closed position with small gestures accompanied by a soft voice ([21]), which influence participants' perceptions of politeness in a number of ways.

On a related note, a recent study by Yap & Casasanto (2016) has corroborated the idea that beat-like gestures also encode semantic and pragmatic information (see also [23]). Their study showed that beat gestures produced in a story reflect the spatial semantics of speech both for literal or metaphorical space (e.g., moving the hands higher to express that temperatures are rising, among others).

These examples illustrate that non-referential beat-like gestures can signal a variety of pragmatic meanings that mark not only the information structure of the discourse, but also its illocutionary force, the epistemic stance adopted by the speaker, and politeness, and other meanings. In fact, it is not only beat gestures which have all these pragmatic dimensions but also referential gestures. For instance, it is well-known that iconic gestures and pointing gestures produced with a beat dimension (i.e. aligned with spoken prosodic prominences) can signal focus or contrastive focus (see [14]). All in all, we believe that further work is needed to investigate the pragmatic dimension of both referential and non-referential gestures which contributes so much to discourse and illocutionary meaning, and one tool that will contribute to this investigation is the development of dimension-based categorization.

3. PROPOSED LABELLING DIMENSIONS

The discussion above suggests that, as a dimensional approach to analysing co-speech gestures is worked out, in addition to its referential component, it should include a prosodic component, to capture e.g. the relationship of the gesture to the prosody of the speech as well as its own rhythmic pattern, a semantic-pragmatic component, to capture e.g. its relation to information structure, speech act type, epistemic status, and affective stance, an interaction component to capture its function in the social interaction, and a kinematic component, to capture aspects such as hand shape, handedness, trajectory shape, location with respect to the body, etc. Using this latter dimension, Shattuck-Hufnagel and Ren [10] have shown that gestures which are

prosodically beat-like can also have complex shapes and phases.

4. CONCLUSIONS

We have argued for the proposal that all gesture types may have both a prosodic component and a pragmatic component, and can carry out a range of functions. This proposal directly builds on and expands previous work, including McNeill's [6] dimensionalization proposal and Kendon's ([1], [2], [3]) observations of both the prosodic and the pragmatic aspects of co-speech gesture, by separating out several dimensions, and offering a set of examples from qualitative analysis and labelling.

Dimensionalizing the analysis of gestures will have both theoretical and practical advantages. The practical advantages include comprehensive labelling which provides a way to annotate non-referential gestures which fail to fit with the definition of purely meaningless beats, but instead contribute pragmatically to the discourse. Theoretical advantages include facilitation of the study of the relationship between all types of gestures and prosodic prominence and phrasing. Positing a pragmatic dimension will also enable further study the pragmatic contribution of gesture to communication. Recent empirical results on the semantic-pragmatic features of beat gestures seem to support this proposal for the relationship of these elements to beat gestures (e.g., [22][23]).

Interestingly, it appears that gesture and spoken prosody work in parallel ways to jointly encode a set of sociopragmatic meanings related to information structure, speech act information, epistemic stance, or politeness ([24]). For example, focus or contrastive focus marking across languages is not only conveyed through prosodic prominence (e.g., with pitch accentuation), but also with the associated use of head nods and manual beat gestures, as well as eyebrow movements ([25][26][27]). Interestingly, recent results on the prosody and gesture patterns of politeness indicate a cross-linguistic tendency to use prosodic and gestural mitigating strategies when the interlocutor is of higher status ([28][29][30]).

In sum, the scientific advantage of dimensionalizing gesture analysis is that it will permit comprehensive assessment of the patterns of integration between co-speech gestural movements, spoken prosodic structure and discourse structure that different gesture types display in large scale corpora.

7. REFERENCES

- [1] Kendon, A. 2004. *Gesture: Visible Action as Utterance*. UK: Cambridge University Press.
- [2] Kendon, A. 2017. Pragmatic functions of gestures Some observations on the history of their study and their nature. *Gesture* 16(2), 157–175.
- [3] Kendon, A. 1972. Some relationships between body motion and speech. An analysis of an example. In: Siegman, A., Pope, B. (eds), *Studies in Dyadic Communication*. Elmsford, NY: Pergamon Press, 177–210.
- [4] McNeill, D. 1986. Iconic gestures of children and adults. *Journal of the International Association for Semiotic Studies* 62, 1-2
- [5] McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- [6] McNeill, D. 2005. *Gesture and Thought*. Chicago: University of Chicago Press.
- [7] Prieto, P., Cravotta, A., Kushch, O., Rohrer, P.L., Vilà-Giménez, I. 2018. Deconstructing beat gestures. *Proceedings of the 9th International Conference on Speech Prosody*, Poznan, Poland, 13-16 June 2018.
- [8] McNeill, D., Levy, E. 1982. Conceptual representations in language activity and gesture. In: Jarvella, R. J. & Klein, W. (eds), *Speech, Place, and Action*. Chichester: John Wiley and Sons, 271-296.
- [9] Cartmill, E. A., Demir, Ö. E., Goldin-Meadow, S. 2012. *Studying Gesture*. In: Hoff, E. (ed). *Research Methods in Child Language: A Practical Guide*. Oxford: Wiley Blackwell Ltd, 208-225.
- [10] Shattuck-Hufnagel, S., Ren, A., Mathew, M., Yuen, I., Demuth, K. 2016. Non-referential gestures in adult and child speech: Are they prosodic?. *Proceedings of the 9th International Conference on Speech Prosody*. Boston, USA, May 31-June 3, 2016.
- [11] Yassinik, Y., Renwick, M. & Shattuck-Hufnagel, S. 2004. The timing of speech-accompanying gestures with respect to prosody. *Proceedings of the International Conference: From Sound to Sense: +50 Years of Discoveries in Speech Communication*, MIT, Cambridge, 10-13 June, C97 – C102.
- [12] Leonard, T., Cummins, F. 2011. The temporal relation between beat gestures and speech. *Language and Cognitive Processes* 26, 1457–1471.
- [13] Loehr, D. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology* 3, 71-889.
- [14] 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, 850–865.
- [15] Stam, G. 2001. Lexical Failure and Gesture in Second Language Development. In: Cavé, C. I., Guaitella, I., Santi, S. (eds), *Oralité et Gestualité: Interactions et Comportements Multimodaux Dans La Communication*. Paris: L'Harmattan, 271–75.
- [16] Streeck, J.K. 2008. Gesture in political communication. *Research on Language and Social Interaction* 41, 154-186.
- [17] Lempert, M.P. 2011. Barack Obama, Being Sharp: Indexical Order in the Pragmatics of Precision-Grip Gesture. *Gesture* 11(3), 241-70.
- [18] Maricchiolo, F., Gnisci, A., Bonaiuto, M., Ficca, G. 2009. Effects of different types of hand gestures in persuasive speech on receivers' evaluations. *Language and Cognitive Processes* 24(2), 239-266.
- [19] Bavelas, J. B., Chovil, N. in press. Some pragmatic functions of conversational facial gestures. *Gesture*.
- [20] Bavelas, J. B., Chovil, N., Lawrie, D. A., Wade, A. 1992. Interactive gestures. *Discourse Processes* 15(4), 469-489.
- [21] Trees, A., Manusov, V. 1998. Managing Face Concerns in Criticism Integrating Nonverbal Behaviors as a Dimension of Politeness in Female Friendship Dyads. *Human Communication Research* 24(4), 564-583.
- [22] Yap, D. F., Casasanto, D. (2016). Semantics of beat gestures. Talk presented at the *7th Conference of the International Society for Gesture Studies*, ISGS 2016, July 18-22, Paris.
- [23] Vilà-Giménez, I., Igualada, A., Prieto, P. 2018. Observing storytellers who use rhythmic beat gestures improves children's narrative discourse performance. *Developmental Psychology*. DOI: <http://dx.doi.org/10.1037/dev0000604>
- [24] Brown, L., Prieto, P. submitted. Gesture and Prosody as Sister Systems in Multimodal Communication. In: Haugh, M., Kadar, D. & Terkourafi, M. (eds), *Handbook of Sociopragmatics*.
- [25] Dohen, M., Loevenbruck, H. 2009. Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech* 52(2-3), 177-206.
- [26] Kim, J., Cvejic, E., Davis, C. 2014. Tracking eyebrows and head gestures associated with spoken prosody. *Speech Communication* 57, 317-330.
- [27] Kraemer, 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.
- [28] Grawunder, S., Winter, B. 2010. Acoustic correlates of politeness : prosodic and voice quality measures in polite and informal speech of Korean and German speakers. *Proceedings of the 5th International Conference on Speech Prosody*. Chicago, IL, May 11-14, 2010.
- [29] Hübscher, I., Borràs-Comes, J., Prieto, P. 2017. Prosodic mitigation characterizes Catalan formal speech: The Frequency Code reassessed. *Journal of Phonetics* 65, 145-159.
- [30] Winter, B., Grawunder, S. 2012. The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics* 40(6), 808-815.