



Deconstructing beat gestures: a labelling proposal

Pilar Prieto^{1,2}, Alice Cravotta³, Olga Kushch², Patrick Louis Rohrer^{2,4}, Ingrid Vilà-Giménez²

¹ ICREA (Institució Catalana de Recerca i Estudis Avançats), Catalonia

² Department of Translation and Language Sciences - Universitat Pompeu Fabra, Catalonia

³ Dipartimento di Studi Linguistici e Letterari (DiSLL) - Università degli Studi di Padova, Italy

⁴ Laboratoire de Linguistique de Nantes (LLING) - Université de Nantes, France

pilar.prieto@upf.edu, alice.cravotta@phd.unipd.it, olgakushch@hotmail.es, patricklouis.rohrer@gmail.com, ingrid.vila@upf.edu

Abstract

In this paper, we advance a comprehensive gesture labelling proposal which highlights the independence of the prosodic and semantic properties of different gesture types and at the same time challenges a simplistic definition of beat gestures as biphasic rhythmic non-meaningful gestures (e.g., [1][2]). Following McNeill's [3] original proposal on gesture dimensions, we defend that all gesture types can associate with prosodic prominence, and even though beat gestures typically display this rhythmic behavior, this is also the case with other representational and pointing gestures too. Second, with respect to meaning, while beat gestures do not represent referential nor metaphoric content, they can serve a range of meaningful pragmatic and discursive functions in speech, which deserve to be further investigated. From a practical point of view, we propose that all non-referential gestures be initially classified as forms of beat gestures with a set of associated properties related to gesture form, prosodic form and pragmatic form. This gesture labelling proposal independently codes for (a) the form of gestures, (b) their properties of temporal association with prosodic prominence, and (c) their pragmatic meaning. We claim that this move allows for a more complete analysis of gestures in large-scale studies and opens the way for more comprehensive assessments of the interaction between gesture forms, prosodic forms, and semantic forms using labelled corpora.

Index Terms: beat gestures, prosodic prominence, gesture-speech association, gesture labelling

1. Introduction

In one of the most used gesture classifications, McNeill [3] proposed a classification scheme for gestures into four main dimensions, namely iconics, metaphoric, beats, and pointing gestures. He described beats as gestures in which "the hand appears to be beating time. As forms, beats are mere flicks of the hand(s) up and down and back and forth, zeroing in rhythmically on the prosodic peaks of speech" ([4], p. 301). In addition, the traditional definition of beat gestures has restricted their gesture forms to short and quick flicks of the hand with a non-complex phasing structure, that is, up-and-down, or back-and-forth movement ([3]). McNeill's traditional definition of beat gestures highlighted the rhythmic and prosodic properties of beat gestures and devoids them of a clear referential function (even though he also pointed out that they were connected to information highlighting). This is probably why recent work on beat gestures has used an oversimplified definition of beat gestures, e.g., as "gestures which represent rhythmic non-meaningful hand movements"

([2], p. 1255) or as "simple, rhythmic gestures which do not convey semantic content" ([1], p.169; see also [5]).

In this paper, we draw on previous experimental findings and on qualitative evidence that challenges a simplified view of beats as non-meaningful biphasic hand movements that act mainly as rhythmic highlighters of speech. Following McNeill [3], we emphasize the independence of the prosodic/rhythmic properties of gestures. It is well-known that speakers integrate their gestures and speech sounds at a temporal level and there exists a temporal synchrony between speech and gesture. Importantly, even though beat gestures have also been called rhythmic gestures, or batons (see [6]) in the sense that they strongly associate with prosodically prominent positions in speech ([7][8][9]), several studies reveal that it is not just beat gestures that are produced in strong association with prosodic prominence, but other gesture types too (e.g., pointing gestures, iconic and metaphoric gestures; [10][4][11]; see section 3 for more details).

Regarding the pragmatic dimension, in the literature on manual gestures, a general distinction is often made between referential gestures (e.g., iconic, metaphoric, and pointing gestures) and beat gestures, which do not represent referential nor metaphoric content (e.g., [3] among many others). While it is clear that beat gestures are manual gestures which do not have a clear referential component (that is, they cannot really be classified as iconic and metaphoric) they rather contribute to the pragmatic and interactional interpretation of discourse. Even though the pragmatic correlates of beat gestures have not been extensively investigated, many authors have suggested specific discourse functions for beats, ranging from McNeill's [2] and Loehr's [10] highlighting focus function to Kendon's [12] locution-spanning gesture units and Bavelas et al's [13] topic and interactive conversational gestures. Following Kendon's ([14][12]) distinction between gestures that have either a substantial function (e.g., related to the utterance content) or a pragmatic function (e.g., gestures which guide the discourse organization or "express aspects of utterance structure, including the status of discourse segments with respect to one another, and the character of the "speech act" or interactional move of the utterance" ([14], p.247), we claim that beat gestures are a type of pragmatic gesture that contributes a variety of discourse meanings (see section 4).

This paper is the result of the team's hands-on discussion of manual and non-manual gestures over a 3-month doctoral seminar on beat gesture coding at Universitat Pompeu Fabra. Even though we have worked on gestures from a variety of databases, for ease of reference and style we use video recordings from a 15 m and 52 s academic Ted Talk lecture given by David Keith on the 26th of September, 2007

[15]. As it is well-known, the presence of non-referential manual gestures is pervasive in public discourse, such as in political and scientific speeches. The video files corresponding to the figures can be found as associated video files linked to our *Speech Prosody* paper. In practical terms, the need to provide an exhaustive classification of gestures into the four standard gesture types (e.g., iconics, metaphoric, beats, and pointing gestures) lead us to propose a more inclusive definition of beats than the original definition. In our view, all manual gestures that cannot be classified as being iconic or metaphoric in nature, as well as deictic, be classified as beats. Beat gestures can thus be described as non-referential gestures involving different hand shapes and gesture articulators, as well as complex gesture phases (see section 2), and with clear pragmatic effects.

All in all, this paper argues for a multidimensional view of gestures and proposes a labelling procedure which independently codes five separate tiers, namely, (a) gesture type, according to McNeill's classification; (b) hand gesture form, with further optional tiers to code forms of other articulators; (c) gesture phases; (d) prosodic component; (e) pragmatic component (see section 5).

2. Gesture forms and phases of beats

2.1. Beat gestures may involve different hand shapes and gesture articulators

Even though McNeill described a typical beat as "a simple flick of the hand or fingers up and down; or back and forth..." ([3], p. 15; see Figure 1 for a reproduction of this gesture), the majority of manual gestures found in our Ted Talk come in a great variety of shapes and positions.



Figure 1: *Illustration of a beat gesture by McNeill ([3], p.16).*

Given the need to exhaustively classify gesture types in large databases, we propose a non-strict definition of beat gestures as non-referential hand gestures which cannot be classified as iconic, metaphoric, or deictic. This loose definition has also been used in experimental research (e.g., Biau and Soto-Faraco's [16] analysis of the neural processing effects of beat gestures in political speech across a variety of hand shapes). Often times, when speakers produce a manual gesture (regardless of whether it has referential meaning or not), they accompany their rhythmic hand movements with coordinated head nod and raising eyebrow movements, as well as torso movements and eye openings. Kraemer & Swerts [17] proposed the term "visual beats" to refer to the coordinated production of several articulator movements accompanying hand beat movements, typically head nods and eyebrow raises. Interestingly, the more articulators a visual beat involves (and the larger and intense their movements are), the higher the spoken emphasis of the accompanying word. Independent evidence has shown that, similarly to hand beat gestures, head

nods show a strong temporal alignment with prosodic structure ([18]; for a review on the labelling of head gestures as head nods, head tilts, and head turns, see [19]). Several authors have investigated the co-occurrence of what they call "eyebrow beats" and "head beats" with acoustic prominence ([20][21]). Cavé et al. [22] showed that rapid eyebrow rising-falling in subjects as they spoke were associated with F_0 rises in 71% of the cases. Figure 2 shows four video stills extracted from Keith's [15] Ted Talk at 01m 58s representing the co-occurrence between hand movements and eyebrow movements. The four selected images represent key points at the gesture phases, namely, start of the preparation phase, beginning of the stroke, gesture apex (in this case end of the stroke), and end of the recovery phase:



Figure 2: *Biphasic manual beat.*

Figure 3 illustrates three different types of beat gestures with a variety of articulators and forms extracted from Keith's speech (at 01m 58s, 05m 51s, and 09m 50s respectively). An open palm hand gesture coupled with an eyebrow raise (left panel), a fist handshape (central panel), and a beat being carried out by the whole upper body (right panel).



Figure 3: *Three types of beat gestures involving various articulators and forms.*

In sum, beat gestures have the potential to be realized in a coordinated way by a wide range of articulators, and a comprehensive gestural labelling of visual beats needs to include these articulators.

2.2. Beat gestures can have more complex phasic movements

As mentioned before, beats have been traditionally described as baton movements which are characterized by biphasic, open handed movements (e.g., up/down, left/right) (see [3][8] for a review). However, even though it has been claimed that the complex phasing structure that is typically found in iconic gestures is not found for beat gestures, several studies have documented cases of more complex structures. Shattuck-Hufnagel et al. [9] documented many cases of beats having a stroke phase that is "often accompanied by one or more of the optional phases of preparation, hold or recovery" (p. 838). While Figure 3 illustrates beats with a typical biphasic

structure, Figure 4 shows one example of a complex beat gesture extracted from Keith [15] at 06m 30s. The six selected time points represent key points in the phase structure, namely, start of the preparation phase, pre-stroke hold, beginning and end of the stroke, post-stroke hold, and end of recovery phase.



Figure 4: Illustration of a beat gesture with complex phasing structure.

Summarizing Section 2, we claim that beat gestures may involve not only hand movements, but also head, eyebrow, and torso movements, and they can have complex phasing structures.

3. Not only beat gestures are typically associated with prosodic prominence

As it is well-known, speakers integrate their gestures and speech sounds at a temporal level. Studies in the last decades have shown that not only the most prominent part of beat gestures temporally associates with pitch accented prominent syllables in speech ([7][8][9]), but also other gesture types such as iconics (e.g., [10][4]) and pointing gestures (e.g., [11]). This rhythmic/prosodic/prominence component that combines with different gesture types has been described as a “superimposed beat” by some authors ([3], pp. 198, 242; [23][24][25][26]). Figure 5 shows one such example extracted from Keith [15] at 06m 39s which illustrates an iconic gesture representing “down/sunk” associated with prosodic prominence in speech.

On the other hand, beat gestures are not always associated with prominent positions in speech. Yasinnick et al.’s [7] study found that most beats but not all overlapped with pitch accented polysyllabic words, at 90% of the time. Figure 6 shows an example extracted from Keith’s Ted Talk at 03:34 and illustrates a multi-stroke manual beat gesture where not all individual strokes associate prosodically with an accented syllable (e.g., see the second stroke which is not associated with a pitch accent). The misalignment of (beat) gestures may occur also during lexical searches (Butterworth beats as per Stam [24]), in speech errors, hesitations, or restarts, etc. These misaligned beats are of particular interest when studying disfluent speech, L2 acquisition, and modelling speech production.

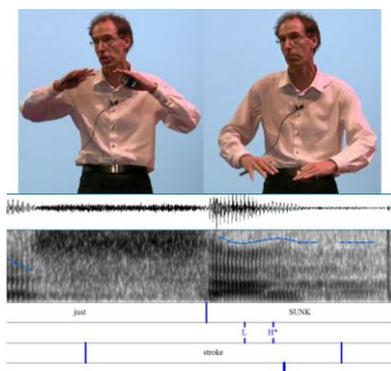


Figure 5: Iconic gesture prosodically associated with a pitch accent “it just **sunk** below...” (**bold** indicates pitch accented syllable).

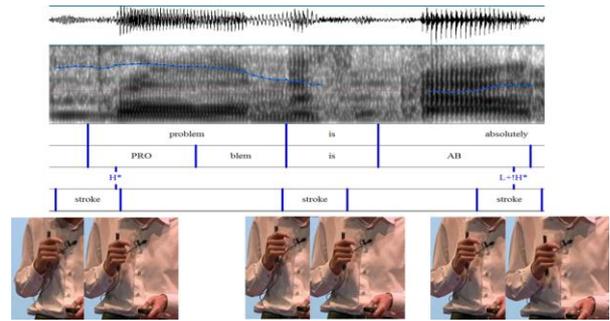


Figure 6: Multi-stroke beat gesture where not all three strokes prosodically associate with prosodic prominence, e.g., pitch accents (as indicated by the English ToBI transcription).

We thus claim that the gesture’s association with prosodic prominence is an independent component of gesture realization and that it should be labelled separately from the gesture form component (see section 5 for Labelling proposal).

4. Beat gestures serve a range of important pragmatic functions in speech

As mentioned before, beat gestures have been defined as non-meaningful gestures by some recent work (e.g., [27][2][5]). However, beat gestures can be also claimed to have a pragmatic function, which “relates to features of an utterance’s meaning that are not a part of its referential meaning or propositional content” ([12], p.158). Authors have highlighted clear pragmatic functions for beat gestures. For example, McNeill stated that a beat gesture “indexes the word or phrase it accompanies as being significant [...] for its discourse pragmatic content.” ([3], p. 15) and that beats can mark “the introduction of new characters, summarizing the action, introducing new themes, etc.”. One of the most clear functions of beat gestures at the discourse level is focus highlighting and information status marking (new vs. given, focus, emphasis, [10], p. 84; [3][14][12]).

Moreover, beats can be claimed to have a discourse structure and conversational interaction function, as well as an epistemic function. According to McNeill, beat gestures, together with iconic, metaphorical, or pointing gestures, can have cohesive functions too because they “tie together thematically related but temporally separated parts of the discourse” ([3], p.16). Streeck [28] provides examples where changes in gestural form “indicate that a new speech act or discourse segment is initiated” and he claims that “the overall pragmatic structure of a speaker’s contribution is visualized by the distribution of similar and different pragmatic gestures across the utterance”. Bavelas et al. [29] claimed that beat gestures are subsumed under interactive conversational gestures, which contribute to the “nature of dialogue itself, rather than with the specific topic of discourse” (p. 476). As proposed in [28], changes in the gestural form mark changes in the illocutionary force of the discourse and/or in the stance adopted by the speaker. In our data we have found a variety of manual gestures expressing the speakers’ epistemic stance, e.g., the belief state of the speaker with respect to his or her proposition. The two panels in Figure 7 show two examples that illustrate the epistemic functions of beat gestures. While the open palms facing the speaker (left panel) show low epistemic commitment and detachment; fist beat gestures (right panel) represent high epistemic commitment:



Figure 7: *Left panel ([15], 03m 19s): “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).*

In our view, the pragmatic and interactional effects of a rich set of non-referential manual gestures deserve to be investigated.

5. Labeling proposal for (beat) gestures

The coding scheme presented in this section has been designed to separately code co-speech gestures. Figure 8 shows an example of the coding of a simple one-handed manual beat gesture. The set of five main tiers contain the following information:

Tier 1- Orthographic Transcription

Tier 2- Manual Gesture Type. Here we follow McNeill’s [3] standard classification of referential gestures into iconic, metaphoric, deictic, and use our expanded definition for beat gestures. As mentioned before, from a practical point of view, we propose that all non-referential gestures (even those with other shapes and more complex phasic movements than in the traditional definition) that are prosodically associated or not with prominent speech are beat gestures.

Tier 3- Head gesture type. In this tier, we code for head gestures as a head nod, turn, tilt, protrusion, or slide [19].

Tier 4- Eyebrow movement. Here, we code for eyebrow movements as rising or falling.

Tier 5- Other articulators. Here we code for other articulators, such as the torso (see right panel of Figure 3).

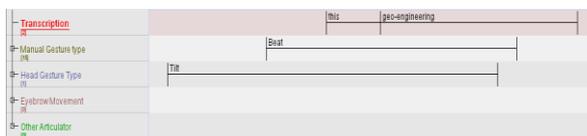


Figure 8: *Example of gesture coding of a one-handed beat gesture ([15] Ted Talk at 03m 37s). Only the 5 main tiers are shown.*

Exploiting the sub-tier option in ELAN, we hierarchically coded for aspects of each gesture articulator, namely gesture form, gesture phases, etc. These aspects are similarly coded as sub-tiers under each of the main parent tiers based on the articulator, as follows:

Sub-tier 1- Gesture form. We take a simplified version of *MIT Coding Manual* (“http://web.mit.edu”) which labels the hand shape, orientation, and trajectory components for the stroke. For non-manual gestures, only trajectory is coded.

Sub-tier 2- Gesture phases. We follow Shattuck-Hufnagel et al.’s [9] proposal for coding the phase structure of manual gestures. Phases may include preparation phase, beginning and end of the gesture stroke, recovery phase, as well as optional pre- or post-stroke holds. A sub-tier codes for the gesture apex separately. For non-manual articulators, only the apex is coded.

Sub-tier 3- Prosodic Component. As the coding of the prosodic tier is perceptual in nature, raters need to assess whether the target co-speech gesture strokes (for manual gestures) or apexes (for the rest of the articulators) are associated with prosodic prominence or not. If users want to analyze in more detail the gesture-speech temporal alignment they can add another tier with the speech acoustic analyses imported from PRAAT.

Sub-tier 4- Pragmatic Component. This tier specifies the pragmatic, epistemic and discourse functions.

Figure 9 below shows the entire hierarchy for manual gestures using the same example shown in figure 8.

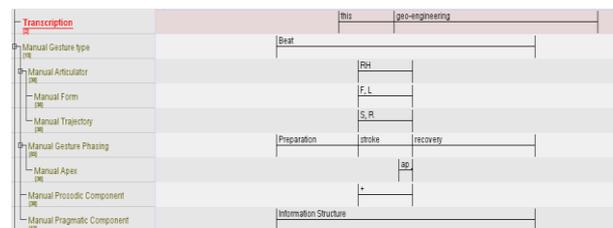


Figure 9: *The Transcription tier and the “Manual Gesture Type” parent tier with the full hierarchy of sub-tiers, taken from [15] at 03m 37s*

6. Conclusions

This paper has proposed a gesture labelling scheme which analyzes all co-speech gestures according to their form, their temporal association with prosodic prominent positions in speech, and their pragmatic function. We claim that a multidimensional approach to the labelling of (beat) gestures has a set of practical and theoretical advantages. First, a practical advantage is that of allowing for a more complete and comprehensive gesture analysis. In the case of beat gestures, a more complex view of these gestures arises when deconstructing their prosodic and pragmatic dimensions. The prosodic tier information (e.g., whether the hand gestures are associated with prosodic prominence or not) will provide important information about the rhythmic properties of gestures, regardless of whether they are beats or not. The pragmatic tier will assess their discourse and epistemic effects of those gestures, which need to be further investigated. Importantly, beat gestures can be viewed as a rich set of non-referential gestures with a variety of prosodic and pragmatic properties which need to be further investigated. All in all, the scientific advantage of dimensionalizing (beat) gesture coding is that it allows us to assess in a comprehensive way the patterns of integration between gesture motion, prosodic structure and discourse structure that different gesture types display in large scale corpora (see [23]).

7. References

- [1] M. W. Alibali, C. H. Dana, and J. Myers, "Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen," *Journal of Memory and Language*, vol. 44, pp. 169-188, 2001.
- [2] D. V. Dimitrova, M. Chu, L. Wang, A. Ozyurek, and P. Hagoort, "Beat that word: How listeners integrate beat gesture and focus in multimodal speech discourse," *Journal of Cognitive Neuroscience*, vol. 28, no. 9, pp. 1255-1269, 2016.
- [3] D. McNeill, *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press, 1992.
- [4] D. McNeill, "Gesture and Communication," in *Concise Encyclopedia of Pragmatics*, J. L. Mey Ed., Cambridge: Cambridge University Press, 2006.
- [5] N. Abner, K. Cooperrider, and S. Goldin-Meadow, "Gesture for Linguists: A Handy Primer," *Language and Linguistics Compass*, vol. 9, pp. 437-449, 2015.
- [6] P. Ekman, and W. V., Friesen, "The repertoire of nonverbal behavior: categories, origins, usage, and coding," *Semiotica*, vol. 1, pp. 49-98, 1969.
- [7] Y. Yassinik, M. Renwick, S. and S. Shattuck-Hufnagel, "The timing of speech- accompanying gestures with respect to prosody," in *Proceedings of the International Conference: From Sound to Sense: +50 Years of Discoveries in Speech Communication, MIT, Cambridge, 10-13 June, C97 – C102, 2004*.
- [8] T. Leonard, and F. Cummins, "The temporal relation between beat gestures and speech," *Language and Cognitive Processes*, vol. 26, pp. 1457-1471, 2011.
- [9] S. Shattuck-Hufnagel, A. Ren, M. Mathew, I Yuen, and K. Demuth, "Non-referential gestures in adult and child speech: Are they prosodic?" in *Proceedings of Speech Prosody 2016*. Boston, USA, May 31-June 3, 2016.
- [10] D. Loehr, "Temporal, structural, and pragmatic synchrony between intonation and gesture," *Laboratory Phonology. Journal of the Association for Laboratory Phonology*, vol. 3, pp.71-889, 2012.
- [11] N. Esteve-Gibert and P. Prieto, "Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements", *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 850, pp. 850-865, 2013.
- [12] A. Kendon, *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University press, 2004.
- [13] J. B. Bavelas, N. Chovil, D. A. Lawrie, and A. Wade, "Interactive gestures," *Discourse Processes*, vol. 15, no. 4, pp. 469-489, 1992.
- [14] A. Kendon, "Gestures as illocutionary and discourse structure markers in Southern Italian conversation," *Journal of pragmatics*, vol. 23, no. 3, pp. 247-279, 1995.
- [15] D. A. Keith, "A critical look at geoengineering against climate change" [online]. TEDTalk available: https://www.ted.com/talks/david_keith_s_surprising_ideas_on_climate_change. [accessed 21-Nov- 2017], 2007.
- [16] E. Biau and S. Soto-Faraco, "Beat gestures modulate auditory integration in speech perception", *Brain and Language*, vol. 124, no. 2, pp.143-152, 2013.
- [17] E. Krahmer and M. Swerts, M, "The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception," *Journal of Memory and Language*, vol. 57, no 3, pp. 396-414, 2007.
- [18] N. Esteve-Gibert, J. Borràs-Comes, E. Asor, M. Swerts, P. Prieto, "The timing of head movements: The role of prosodic heads and edges," *The Journal of the Acoustical Society of America (JASA)*, vol 6. no. 141, pp. 4727-4739, 2017.
- [19] P. Wagner, Z. Malisz, & S. Kopp, "Gesture and speech in interaction: An overview," *Speech Communication*, vol. 57, pp. 209-232, 2014.
- [20] G. Ambrazaitis, and D. House, "Acoustic features of multimodal prominences: Do visual beat gestures affect verbal pitch accent realization?" in *Proceedings of the AVSP*, 2017.
- [21] P. Eckman, "About brows: Emotional and conversational signals," in *Human Ethology*, M. von Cranach, K. Foppa, W. Lepenies, and D. Ploog, Eds., Cambridge: Cambridge University Press, 1979, pp. 163-202.
- [22] C. Cavé, I. Guaïtella, R. Bertrand, S. Santi, F. Harlay, and R. Espesser, "About the relationship between eyebrow movements and Fo variations" in *Proceedings of ICSLP 96*, Philadelphia, pp. 2175-2179, 1996.
- [23] D. McNeill, F. Quek, K. McCullough, "Catchments, prosody and discourse," *Gesture*, vol. 1, pp. 9-33, 2001.
- [24] G. Stam, "Lexical Failure and Gesture in Second Language Development" in *Oralité et Gestualité: Interactions et Comportements Multimodaux Dans La Communication* C. I. Cavé, I Guaïtella & S. Santi, Eds., Paris: L'Harmattan, 2001, pp. 271-75.
- [25] S. McCafferty, "Gesture and the Materialization of Second Language Prosody," *IRAL-International Review of Applied Linguistics in Language Teaching*, vol. 44, no. 2, pp.197-209, 2006.
- [26] S. Harrison, "Evidence for Node and Scope of Negation in Coverbal Gesture," *Gesture*, vol. 10, no. 1, pp. 29-51, 2010.
- [27] W. C. So, C. Sim Chen-Hui, and J. Low Wei-Shan, "Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall?" *Language and Cognitive Processes*, vol. 5, pp. 665-681, 2012.
- [28] J. Streeck, "Gesture in political communication," *Research on Language and Social Interaction*, vol. 41, pp. 154-186, 2008.
- [29] J. B. Bavelas, N. Chovil, D. A. Lawrie, and A. Wade, "Interactive gestures," *Discourse Processes*, vol. 15, no. 4, pp. 469-489, 1992.