DOI: 10.20396/joss.v14i00.20383

GESTURE STRUCTURE, INSIGHTS FROM ANNOTATION GUIDELINES

BARROS, Camila Antônio[1*]

[1]Freie Universität Berlin – ORCID: https://orcid.org/0000-0002-3776-7606

Abstract: *This paper analyzes the internal structure of gestures, namely preparation, stroke, and recovery, by making a close reading of early (1970's onwards) definitions. Those definitions are centered around segmenting the gesture based on saliency, creating an opposition between the stroke, the most well-defined part of the gesture, and other phases. Although this definition initially relied heavily on the association to speech, the need to operationalize the annotation forced definitions more linked to articulation features (Linguistic Annotation System for Gestures, Bressem et al., 2013). Later studies, mainly linking gestures to phonological structures of speech, started to rely on gesture kinematics to define gesture structure, which are motion-based features such as velocity profiles and movement targets (Kita et al., 1998; Loehr, 2004; Rohrer et al., 2023). The latter provides a qualitative base for the use of motion tracking technology in gesture studies, leading to the interpretation that the stroke can be better defined by its velocity profile than in connection with articulation features. In terms of theory, this impacts how gestures can be segmented and understood in language, without necessarily needing to fully integrate the phonology of the utterance it is associated to. It seems that gesture-speech association depends on the transparency it shows to its speech affiliated, moving from a tight synchrony for less representational and beat gestures, to a loose synchrony for iconic and recurrent gestures.*

Keywords: Gesture; Annotation; Corpus linguistics.

[*] Corresponding author: camila.anbarros@gmail.com

1 Introduction

Gesture studies are centered around how movement of hands and arms can be understood as meaningful and be connected to speech in various ways. There are multiple classifications of gestures, linked to their degree of conventionalization and abstraction, while the internal structure of the gesture remains underdefined. It can be elaborated in terms of form, in which better-defined movement and handshape point to the meaningful part of a gesture. It can also be elaborated in kinematic terms, in which the quantitative analysis of movement enables the analysis of precise synchrony of gesture and speech. In this paper, we argue that kinematic analysis enables a more precise description of the internal structure of the gesture, and by consequence, its relation to speech.

In the following section, we present definitions regarding gestures and their phases, showing how the linear structure was proposedly left underspecified. Anchored in this openness, different classifications emerged to make sense of the meaning created in gestures, as shown in Section 0. In Section 0, we then show two perspectives of analyzing gestures based on their form and their kinematics. Next, both perspectives are compared regarding their treatment of repetitions, in Section 0. In the last section, we conclude that both annotation perspectives lead to an understanding of gesture that is complementary, and it can benefit from automatic tools, although they do not solve all questions. Nevertheless, it becomes clear that competing annotations impact on the way gesture-speech association can be understood.

2 Initial formulations

Understanding the internal structure of a gesture is necessary to comprehend how it can be coordinated with speech, especially in terms of synchronicity, shedding light on which points of a gesture are associated with which points of a phonological grid.

Kendon (1) provided the initial framework for co-speech gestures, by proposing a segmentation guideline. This provided the outline for a structure of the gesture similar to the division found in speech: a larger unit delimited by pauses, internally divided by meaningful units which have different degrees of salience relative to the meaning that is being conveyed. According to the author, gesture and speech are linked to the same idea unit¸ meaning they are tied together not only in time, but also in their semantic meaning, and pragmatic function.[1]

In its first formulation, gesture segmentation starts with units, delimited by rest positions. The gesture unit is the full excursion that hands and arms undertake before resting. Varying from speaker to speaker, rest positions are configurations in which there is little to no tension in hands and arms. Most importantly, it is not possible to recognize any communicative value in a rest position. Within a gesture unit, it is possible to recognize that some parts are more and others less salient. The more salient parts that can be distinguished are the strokes. They are the central part of a gesture phrase in which postures and handshapes "are better defined than elsewhere in the excursion" (3p112). The stroke can (but mustn't necessarily) be framed by preparation, hold, and recovery. The preparation comprises the positioning and tensioning of arms and hands that will lead to the stroke. The hold is when a speaker sustains their hands in the same position and handshape as the stroke, also called a post-stroke hold. Gesture phrases are subdivisions within a

[1] 'Idea unit' is an underspecified concept coined by Chafe (2) that represents a single focus of consciousness, which would reflect in the parsing of the speech stream. As the intention is to understand the argumentation about gestures from the perspective of the authors that first analysed them, the terminology and definitions are kept as faithful to the original as possible. In this case, 'idea unit', is kept as it is pivotal to Kendon's (1972) argumentation.

gesture unit that include the stroke and, when there is a preparation and hold, also include them. They overlap with prosodic phrases, the smallest syllabic group over which a completed intonation tune occurs in Kendon terms.[2] In some cases, the speaker retracts their hands, losing tension and relaxing them in an almost or full return to the rest position. According to Kendon, this is not part of a gesture phrase but is still a part of the gesture unit.

To clarify the internal divisions of gestures, two corpora were used. The examples presented here have been sourced from the BGEST Corpus (6), a Brazilian Portuguese corpus made up by roughly 4000 words and 400 gestures collected amongst ten speakers living in Belo Horizonte, Brazil. It is part of the C-ORAL Brasil corpora. Other examples have also been sourced from Recipe collection (unpublished). This data collection comprises eight speakers telling their favorite recipes, in Brazilian Portuguese (Sao Paulo), European Portuguese (Lisbon), and Spanish (Madrid and Bogotá). Each variety has two speakers in interactions of 30 to 90 seconds.[3]

In Table 1, there is a gesture unit made up of a single gesture phrase, in which different phases and be understood simultaneous to the utterance *depois / <u>grelhado</u> //* (eng. 'then, grilled').[4] In the first frame, there is a rest position, with both hands resting on the speaker's lap. Some tension can be perceived, but it does not seem to have a communicative value. In frame two, the left hand rises until close to the left shoulder, with pursed fingers. During the stroke, in frame three, the left hand makes a straight movement towards the chest. After that, in frame 4, the speaker rests their hand on their lap again in a recovery phase.

[2] As it will be debated further in the text, this definition was updated in gesture studies to accompany advances in Phonological Theory. The reader is referred to Grice, and collaborators (4) and to the volume edited by Gussenhoven and Chen (5) for more up-to-date definitions.

[3] The examples were coded as bgest_000 when they are a part of the BGEST Corpus and as REC_SBLP, when they were a part of the Recipe collection. In this collection, the following coding was used: "SB" stands for Andean Spanish, "SM" stands for Castillian Spanish, "PL" stands for European Portuguese, and "PS" stands for Brazilian Portuguese. The abbreviation that follows is an acronym for each speaker.

[4] In order to keep consistency and help the reader, all examples follow the same convention as found in the BGEST Corpus used in the C-ORAL BRASIL Corpus (7): speech segmentation is represented by "/" for a non-terminal break, and "//" for a terminal break. Underlined words stand for parts of the utterance that are synchronous to the gesture.

Table 1: Overview of gesture phases, REC_PLCB[5]

#	Frame	Text
1		silence
2		*depois* eng. ‘then’
3		*grelhado* eng. ‘grilled’
4		silence

McNeill (8) kept similar definitions, stating that the stroke is the “main part of the gesture” (p25), which might be preceded by a preparation and followed by a retraction. When the stroke has a kinetic goal, a point of maximum extension and maximum tension in the fingers, this point is called an apex (9). Here, the nomenclature ‘recovery’ is replaced by ‘retraction’, but the definition of “return to quiescence” (8, p25) stays the same. In both classifications, the definition of the phases (and by consequence, phrases) relies on the perception of movement and the meaning it ties with speech.

Thus, unlike phonological structures, such as syllables, the internal division of gestures consists of a parsing of salient and non-salient units that can be seen as meaningful and be associated with speech. Syllables follow a crescendo/decrescendo of sonority and are linked to

[5] All corresponding videos can be accessed in the supplementary material.

cycles of the jaw opening/closing. The sonority is encompassed in a combination of distinguishable parts (phones) that are linked to an abstract instance (phonemes) combined in grammatical ways (10, 11). The preparation does not create expectations for a given meaning, just for a possible meaning, whereas a syllabic onset due to parallel transmission carries acoustic features that create an expectation for a limited universe of vowels (12). In sign languages, this would also be the case, but instead of vowels, there are certain parameters of handshape, place of articulation, movement, and hand orientation that have to be fulfilled to make a syllable[6] (13). To exclude a movement understood as a preparation, does not impact the stroke – there can be strokes with no preparation. Removing a consonant of a syllable impacts the syllable structure, creating fertile ground for phonological processes that may take place. What does impact the gesture is the absence of sharp salient movement, in which it is understood as an interrupted stroke.

In Table 2, there is a sequence of an incomplete stroke and a complete stroke, during which the speaker utters: *é só comprar batata palha /* (eng. 'you only have to buy shoestring potatoes'). The stroke is underlined, and the first two words do not have any accompanying movement. In frame 1, there is an incomplete stroke during *comprar*, in which the speaker places the left hand with pursed fingers at the center of their torso, but then suddenly retracts the hand towards their chest. This position, frame 2, is held until the next stroke takes place, frame 3, using both slightly tensioned hands to make a movement away from their chest.

[6] In sign languages, there are multiple differences that need to be addressed when comparing their phonology to oral languages. The reader is referred to Brentari (13) and references therein for a deeper understanding that the space here allows.

Table 2: Incomplete stroke and stroke, REC_PLCB

#	Frame	Text
1		*comprar* eng. ‘to buy’
2		silence
3		*batata palha* eng. ‘shoestring potatoes’

This does not mean that certain kinds of patterns of movement and handshape that emerge in the stroke can be associated with specific meanings through conventionalization. Instead, that relation is established in a different way than in the phonological sense: while the association between the Saussurian *signifié* and *signifiant* in words starts with the *signifiant* that is arbitrarily associated with a *signifié*, gesture families (like lexeme families) start at the same *signifié* that is recurrently associated with the same forms, *signifiant* (14). Gesture families can have different degrees of complexity, which makes it possible to recognize different tendencies in the form association or in varying degrees of recurrency. An example of this is the Open Hand Prone down (Figure 1), associated with negation. It is also found in the example below, taken from the BGEST Corpus (6). Here, the speaker sweeps the right hand at the center of their torso from left to right in order to negate any further contact with this specific partner, while uttering *aí / ele parou de falar comigo no ano-novo //* (eng. ‘then / he stopped talking to me by New-Year’).

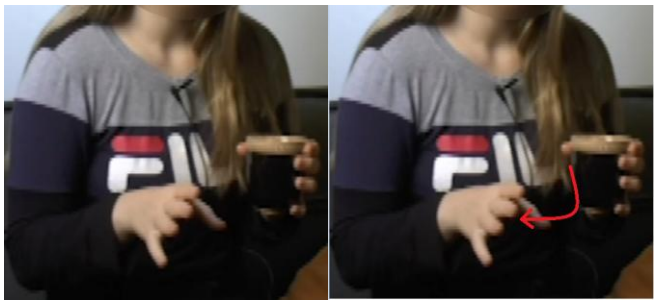


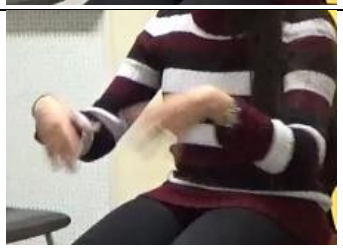
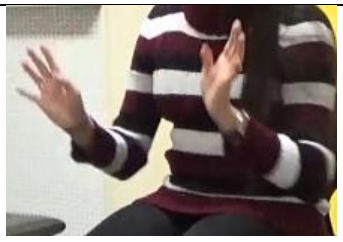
Figure 1: Open Hand Prone down, bgest_007.

In these highly influential publications, there is a clear subjective factor intrinsic to the delineation of the internal structure of a gesture, as the saliency of the stroke depends on the perception of the annotator. Kendon (3) reports that there is an agreement on the detection of deliberate movements, gestures, when a group of people evaluate a video, indicating that there is a qualitative difference that allows one to recognize it. When deliberately expressive, the movement is characterized by having sharp boundaries of on- and offset, as well as an excursion. Therefore, it is different from a simple change of positions. Furthermore, the excursion was understood as a part of what the speaker was saying rather than a manipulation of an object, an adjustment of body position or as incidental.

The gesture's internal structure presented so far is compelling, although underspecified. McNeill (8, p22) justifies it highlighting that by not having a well-formedness, i.e., not having a shape to adhere to, gestures can express "just those aspects of meaning that are relevant and salient to the speaker and leave out aspects that language may require but are not relevant to the situation". This explains why the stroke itself was not further specified, as its meaning emerges from the association with speech.

However, by not specifying the stroke, the means by which the association to speech can be made and perceived by speakers relies on a complicated synchrony, as shown in the example in Table 3.

Table 3: Slight asynchrony of iconic gesture, REC_SBLG

#	Frame	Text
1		*se pone [el tamal] en una hoja con* eng. ‘you put [the tamal] in a pan’
2		*con* eng. ‘with’
3		*água* eng. ‘water’
4		*e ahí* eng. ‘and there’
5		*se pone* *en. and it is set*

In this example, the stroke is not synchronous to what it is referring to but slightly precedes it. The speaker utters *se pone [el tamal] en una hoja con agua* (eng. ‘you put it [the tamal] in a water pan’) synchronous to two iconic strokes. The first frame is the stroke with both hands open, palm down, in a downward movement. This is immediately followed by a second stroke (frame 2), a circular movement with both hands, depicting a pan filled with water. The word *agua* (eng. ‘water’), however, is uttered during a recovery phase, revealing a slight asynchrony of gesture and affiliated meaning. It is followed (frame 4) by an abstract deixis, marked by a gesture with both hands indicating where the tamal should be placed. In frame 5, there is a recurrent gesture, a throwing away gesture (15), which bears a semantic core of exclusion. Here, it can be interpreted as once the *tamal* is in the water, there is not much more to be done. This example implies that

the synchrony rule[7] proposed by McNeill (8) applies differently, sometimes meaning a complete overlap, other times, a slight displacement is also possible.

Different continua have been built to describe how gestures can be ascribed to language. Initially, it was proposed in terms of conventionalization (as in Kendon's continuum by McNeill 8, 16, expanded by 17), tackling how the imagistic and mimic components of a gesture can explain its relation to speech. In this perspective, the bulk of gesture studies should focus on gesticulation, "idiosyncratic spontaneous movements of the hands and arms accompanying speech" (8, p37). Later it was elaborated to make sense of the variation of form and meaning in gestures (18), tackling how historically gestures are built with speech as an emergent hybridization in terms of both conventionalization and compositionality (p14). It is also possible to separate gestures in terms of their representationality (19, 8), dividing between (a) representational, such as iconic and metaphoric gestures, and (b) an undefined category of non-representational gestures, in which there is no immediate similarity between the gesture and its affiliated phrase.

Table 4: Comparison of different continua

Convention	-			+
Kendon's Continuum (8, 16, 17)	Gesticulation	Pantomime and deictics	Emblems and metaphorics	Sign language and iconics
Form and meaning	variable			fixed
Compositionality and conventionalization (18)	Singular gestures, variable relation of form and meaning	Recurrent gestures, stabilization of form-meaning	Emblems, fully conventionalized in fixed form-meaning unit, not compositional	Signs, fully conventionalized in a fixed form-meaning unit and able to be used compositionally
Imagistic	-			+
Representationality (8, 19)	Broad (and not fixed) category that encompasses various types of gestures, including beats, deictic, emblems, recurrent, and others. The meaning of these gestures is not immediately apparent from their form.			Iconic and metaphoric gestures

In the following section, we argue that articulation-based annotation schemes seem to be more directed towards explaining the representational gestures, with variable degrees of form-meaning relations and conventionality. Kinematic annotation schemes, on the other hand, seem to enable researchers to shed light on non-representational gestures. This method can show if it is true that less conventionalized gestures tend to be more sensible to phonological synchrony, indicating an increase of linguistic properties in gestures, such as prosodical features. This might entail different semantic and pragmatic nuances, depending on the temporal placement of the apex. More representational and conventionalized gesture synchrony, on the other hand, is not exactly time-constrained, but rather structurally dictated. In those cases, the combined meaning

[7] Synchrony rule (8) stands for a tripartite property of gestures, which states that the stroke phase is integrated into the phonology of the utterance" (Phonological synchrony, 8, p26), gesture and speech "present the same meanings at the same time" (Semantic synchrony, 8, p27) and "perform the same pragmatic functions" (Pragmatic synchrony, 8, p29).

of verbal and visual message stems from a combination of gestural depictions that requires an overlap with a certain stretch of speech – a word, an utterance – to create an affiliation.

3 Annotation schemes

As gesture studies progressed, different annotation proposals appeared to ease this laborious task, providing guidelines on how to consistently analyze it. No annotation system is theoretically neutral, and gesture studies are no exception to this rule (20). There are at least three main perspectives that can impact gesture annotation schemes: movement perspective, form-based perspective, and kinematic perspective.

The first refers to centering the analysis around only bodily movements that regardless of whether it is linguistically understood or not. This was, for instance, used for the annotation system NEUROGES (21), a method of movement assessment that can provide a repertoire for coding different situations, from co-speech gestures to dancing. Here, the linguistic role of the movement, as in gestures, does not play a role for it to be considered in the annotation.

The second perspective refers to a focus in a form-based approach to gestures, creating "a description of the structural and functional properties" (22). In this conception, put forward by the Linguistic Annotation System for Gestures (22, LASG), the gestural meaning emerges from its form, as regularities emerge revealing *Gestalts*, "meaningful wholes" (p1100), that will reflect cognitive aspects, such as embodiment. Hence, the gestural meaning emerges from gesture form, a combination of articulatory features that involve almost the same parameters as sign language (handshape, movement, location, and palm orientation) connected to an underlying meaning provided by an overlapping word or utterance.

The third perspective, kinematics, is also based on forms, but instead of looking specifically for *Gestalts*, the cognitive link is based on the understanding that the organization of the movement provides a blueprint for synchrony to speech. The integration to the phonology of the utterance stems not from its overlap, but from specific velocity modulations that, when aligned to certain parts of the utterance, are able to modify the perception of events that are normally restricted to acoustic cues (23–25) and create pragmatic nuances (26, 27). For the purpose of this article, we are going to explore the last two types, which were more extensively explored in linguistic works.

3.1 Form-based approach

In the LASG, the segmentation of gesture units and phases is directly taken from Kendon's seminal publication (1), as a way of establishing the relation of these chunks to locution clusters in speech. The bulk of the annotation relies on gestural form, making it explicit in the description of orientation, movement, hand shape, and position in the gesture space (i.e., torso), and motivation of form. Here, the gesture is qualitatively characterized based on its ephemeral shapes, objects, and trajectories that are created, in order to assess its basic meaning. The idea is to understand which motoric patterns from everyday actions are schematized to evoke specific meanings. A straightforward example of this is the use of circles, either with the fingers or creating the form in the air, to evoke the meaning of a cycle.[8] Aside from the annotation being dedicated to gesture form, this scheme also includes tiers dedicated to the annotation of speech and gestures in relation to speech.

[8] The reader is referred to Kendon (3) and Ladewig (28) for a description of the meanings linked to the cycle gesture.

This approach is the result of a number of studies, which analyzed gestural meaning by its form, looking at how movement articulation related to mundane actions (29–31). A pervasive way of understanding gestures is by using this broad definition of its phases, and using the meaning conveyed through form. This allows to classify co-speech gestures regarding their degree of conventionalization, building a continuum between *ad hoc* creations that "encode meaning in a holistic fashion" (32, p134)[9] and "fixed-form meaning relation" (18, p4) that has gained meaning through a historical process of lexicalization, creating a progressively arbitrary sign.

This continuum is framed by an understanding of language as being fundamentally multimodal, usage-based and dynamic, differentiating singular, recurrent, and emblematic gestures (33). Singular gestures are based on a repertoire of techniques of imagistic depiction culturally determined and inserted in the interaction as a free manifestation of the speaker. Recurrent gestures rely on *Gestalts* that are used at a certain frequency in a single context, so that an ongoing conventionalization takes place, so that its use is constrained to a meaning yet requires speech to be fully understood. Emblematic gestures are fully conventionalized, being able to perform acts of speech, also when in the absence of speech. In this classification, the pragmatics leaves a place for semantic meanings when moving from singular to emblematic gestures. This classification is the result of an interpretation of both Kendon (3) and McNeill (8) put forward by Müller (18), opposing a strong divide between gesture and sign, as defended by Goldin-Meadow and Brentari (34).

Another way of classifying gestures is by regarding their degree of abstraction, going from highly imagistic gestures to flicks of the hand. In this perspective, the interplay of level of representationality and pragmatical meaning is decisive to decide the semanticity of the gesture. This does not entail that a pragmatical gesture is less imagistic, rather indicates that how the same movement, e.g., pushing away gesture, will be interpreted depends on the word it is affiliated with. If it overlaps with a 'no', it will carry a pragmatic meaning of negation, whereas if it overlaps with something more iconic, such as 'push the books to the side', it will carry a semantic meaning. In McNeill (8) a gesture classification is proposed, defining the gestures as "idiosyncratic spontaneous movements of the hands and arms accompanying speech" (p37) based on the necessity of the presence of speech. In this differentiation, there are iconic gestures, depictions of concrete concepts by formal similarity to the semantic content; metaphoric, depictions of abstract concepts, e.g., time as space; cohesive gestures, without a depiction pattern, which aid the conversation rather than inserting meaning to it; and beats, gestures that follow speech in some sort of rhythmicity.

Taking together the classifications of gestures in different continuums informs us about how it is possible to make sense of hand and arm movement looking at it holistically, i.e., it is the whole gesture that creates its meaning, and not specific internal details. This understanding is at the core of the proposal of McNeill's on how gestures do not convey meaning as language, as they do not build meaning compositionally. Yet, as research in the field progresses, it is necessary to understand that there is a patterning in gestures, which allows an association to speech. In this paper, we argue that the association to speech depends on the internal structure of the gesture.

[9] Being strict about the terminology, it is necessary to say that Kendon's formulation use the term *gesticulation* to define movements that seem "deliberate, conscious, governed by an intention to say something or to communicate" (3, p112) in opposition to other movements. McNeill used this term to define the gestures that needed speech to have a meaning, in opposition to emblems, highly conventionalized gestures that can happen regardless of speech. McNeill then proposed *gesture* to be used instead of *gesticulation.* After that, the any deliberate hand movement was caught under this umbrella term, therefore Kendon's distinction gesture/gesticulation is not used here. In cases in which the type of gesture is necessary, it is also specified.

3.2 Kinematic approach

The development of the kinematic perspective is tied to the understanding of how gesture can be described first by its movement, i.e., velocity, vertical amplitude (highest point during a movement), submovements, and number of holds. This combination then sets the scene for the associations that can be made in relation to the utterance it is affiliated with.

Kita, van Gijn, and collaborators (35) tried to differentiate the phases of a gesture in a way that would also be valid for sign language, i.e., with a syntagmatic rule system. In sign language, five parameters must be fulfilled to make up a lexical sign, namely, handshape, movement, location, palm orientation, and non-manual markers (13). Alterations in these parameters can make up minimal pairs. For lexical signs with a change in location, Kita, van Gijn, and collaborators (35) observed that there is a "sequence of three elements, each of which can be independently affected by morphological processes and phonological rules, and to each of which other form features (e.g. hand shape, hand orientation) can be associated." (35, p25).

For gestures, unlike for signs, the whole movement is divided into phases, not only the obligatory stroke. This division also can include holds, meaning the absence of movement before or after a stroke, which can be interpreted as a way of sustaining synchrony with speech, as "[w]hen the stroke is completed 'too early', the hand is held until the phonological peak" (35).[10]

For the authors, the stroke is the "content-bearing part of the gesture" (35, p27), in which there is an effort on the movement form. While strokes convey a meaning in functional terms, formally, they have a more defined form (shape, trajectory, posture) than elsewhere in the movement stream.

Both signs and gestures can be then segmented into movement phrases, which require not solely strokes, but fundamentally an *expressive phase*, a semiotically active phase in which movement and conveyed information are associated with each other. Here, *an independent hold* can also be expressive. Preparation can be underspecified by a liberating movement, location preparation and hand-internal preparation.

The key feature of this segmentation is the attention to *velocity profiles* as a cue to discontinuities, which can also be marked by a change of direction in the hand movement. The hand movement can be hand internal, a change in the hand shape, palm or finger orientation, or a path, a change of position of the wrist in relation to the torso. Only the latter is relevant for the segmentation, as it can differentiate if a stroke has one or multiple segments.

Making these differences more concrete, a single-phase segment would be as in Table 1, where there is only one velocity profile and a single direction. If there is only one velocity profile and multiple directions, as in Table 3, there is a multi-phase segment[11]. There is something in between, i.e., a semimulti-segment phase takes place: it includes cases in which the "hand stops abruptly in the first segment, and in the second segment, it bounces back (with a different velocity profile) to the opposite direction, tracing back the trajectory of the first segment (not necessarily all the way back)" (35, p30).

[10] Phonological peak is an unspecified term used by McNeill to describe the link between gesture and speech, indicating that the gesture is "integrated to the phonology of the utterance" (8, p27). In a later interpretation of the term, Shattuck-Hufnagel, Yasinnik, and collaborators (36), would say that the rationale behind the timing relationship of gestural strokes and intonational phrases in Kendon (1) in 1972 is related to the final pitch accent of an intonational phrase. This would then later be interpreted by McNeill as gestures being part of an utterance. Nowadays, gesture relations to prominence, not only related to pitch accents, are under scrutiny.

[11] "[A] two-segment movement without velocity-profile discontinuity is coded as one phase even if there is an abrupt change of direction" (35, p29).

The takeaway of this discussion is that both gestures and signs, behave similarly in terms of movement organization, as the specification of the content-bearing unit, lexical sign or a stroke, is defined by velocity profiles.

Loehr (9) made a similar distinction, using the apex as an index for the velocity profiles. In the stroke, there is a specific peak that can be distinguished as having the "kinetic goal of the stroke" (9, p89). In this point, the movement speed is slower. As such, it is possible to distinguish it by a sharp frame, the point of maximum extension, and/or sudden stops or changes in direction (36). This then allows for differentiation between uni-, bi-, and multidirectional strokes. In the first case, the kinetic goal is at the end of the stroke, matching the single-segment phase. Bidirectional strokes would be understood as a stroke in which the direction changes abruptly, that matches both the semimulti-segment phase and the multi-segment phase. Multidirectional strokes have multiple apexes, which include cases in which there are movements that repeat themselves.

The classification was undertaken by Rohrer, and collaborators (37) when elaborating a full manual on how to code gestures, also known as the Multidimensional Multimodal Labelling System (M3D). The author's aim was to create an annotation scheme that conciliates McNeill's and Kendon's views, proposing that gestural functions are not mutually exclusive, but rather operate simultaneously as multi-dimensional. This means that "gestures do not need to fit into particular categories, but rather different formal, semantic, pragmatic, and prosodic characteristics can be assessed as independent dimensions of individual gestures" (37, p9).

Regarding the division already presented, unidirectional strokes were inherited from Loehr (9) as a stroke with a single apex placed at the right margin. From this perspective, the number of apexes in a stroke can be defined by the number of hand movements that reach zero velocity, identified by an annotator "in terms of the clarity in the image of the individual frames" (37, p28). This entails a differentiation of types of strokes: unidirectional, with a single apex as an endpoint; bi-directional, with two apexes located at the point of change of direction and/or point of zero velocity; and multi-directional strokes, with multiple changes in direction and/or points of zero velocity. It might be the case that a stroke has a bidirectional movement, a bi-directional stroke, in which the apex will be located at "the point of change in direction and of zero velocity" (37, p29), as is the case with up-down movements with a single change (one peak) in movement direction, or multiple peaks (multi-directional stroke) in which the movement of the gesture is made up of successive changes in direction. It can also be the case that a particular stroke does not bear an apex (e.g., circular movements), as the kinematic goal is the movement itself rather than a target configuration, referred to as continuous gesture by 36, Renwick and collaborators (36). Table 4 shows a summary of the perspectives discussed so far.

Table 5: Comparison of definitions to strokes in different annotation schemes

Multiple definitions of strokes	
Kendon (3, 1)	Saliency
McNeill (8)	Synchrony
Bressem, Ladewig, and Müller (22), LASG	Form-based approach to gestural meaning
Kita, van Gijn, and collaborators (35)	Velocity profiles

Multiple definitions of strokes	
Loehr (9)	Apex
Rohrer, and collaborators (37), M3D	Multidimensionality

The perspectives so far reveal a slow and steady departure from the understanding that gestural meaning necessarily depends on the assessment of its form. It is rather the kinematic nuances that provide a blueprint for the association to speech. Such a perspective can be seen in works that intend to connect gestures, especially those at the more abstract and less conventionalized end, to specific points in the utterance, such as stressed syllables (24), perceived prominence (23), and pitch accents (25). They also offer an insight into a more objective definition of the stroke, linking it to a pragmatic function of movements that bear a communicative intent (38).

4 Discussion aided by repetition

In the sections above, it has been presented that the phases that make up the gesture have an impact on the understanding that one can have from it, influencing the way their meaning can be conveyed. Initially defined through its saliency and synchrony, the stroke, the gestural phase that is responsible for conveying meaning, could be intuitively segmented by seeing which part of the movement as a whole was more defined and seemed to be integrated into the utterance, phonologically, semantically, and pragmatically.

This perspective was put forward by several works, which then culminated in the LASG, an annotation scheme that implemented these subjective stroke definitions, which were compensated by a thorough analysis of form. From the articulation of movement, a *Gestalt* should then emerge that could be used as a base for the meaning conveyed in the gesture.

Kinematic approaches slowly developed more as an operationalization method, clarifying some conundrums that might appear during the annotation. This development, however, impact theoretically on how gestures can convey meaning. Rather than being solely form-based, the meaning can be dissected also by the association to speech by looking at turning points, such as velocity peaks minima and maxima, indicating that the meaning of the gesture is multidimensional: there are phonological points that can be synchronous to speech and, from that alignment, meaningful overlaps to semantic content and pragmatic function emerge. Meaning is then constructed by layers that are associated with speech in different ways.

One point in which form-based and kinematic approaches can be closely compared is in terms of how they handle repetitions in gestures. They can come in two main flavors: a repetition of the same articulation parameters in the same way or the reiteration of some parameters, while shifting others. In the LASG, extensively debated by Bressem (39), this distinction is called, respectively, iteration and reduplication. This definition also contains a semantic nuance by adding complementary semantic information to the utterance.[12] Reduplications convey complex gestural meaning, being restricted to abstract meanings, conveying emphasis through redundancy.

[12] The author also mentions a prosodic function, which is not specified, simply hinting it to a prosodic use of beat gestures, unconventional and abstract gestures that tend to synchronize with tonal events (9), although this notion is under intense debate (41). As this is not sufficiently described in LASG, this will not be taken into account.

Table 6: Iteration and reduplication

	Frame 1	Frame 2	Text
Iteration REC_SBLG			*zanahoria* eng. 'carrots'
Reduplication bgest_003			*por exemplo / se eles [os alunos] vão produzir um statement of purpose* eng. 'for instance, if they [the students] are producing a statement of purpose.'

In the example for iteration, extracted from a recipe description, the gesture overlaps with *zanahoria* (eng. 'carrots'), with a similar amplitude throughout, adding the information that the carrots should be round. The example for reduplication, extracted from an explanation of different text types, *por exemplo / se eles [os alunos] vão produzir um statement of purpose* (eng. 'for instance, if they [the students] are producing a statement of purpose'). The circular movement changes, growing bigger in the second frame, indicating that there are multiple elements that need to be addressed in a statement of purpose (42, p87, adapted).

Kinematics also makes a similar differentiation. Multiple segments and multiple apexes entail that the annotator perceives the repetition as a necessary part of the meaning created in the stroke. According to Kita, van Gijn, and collaborators (35), it is possible to differentiate between cases in which there is no hold in between movement repetitions, called a single repetitive phase, and cases in which one movement is superimposed on a path movement, which would change the duration of the repeated phase, as well as the movement size. Rohrer, and collaborators (37) adopted a similar position, but classifying the first cases as a multidirectional stroke, since the meaning is holistically conveyed by the repetition. Movements that show repetition but differ in the articulatory features are understood as different unidirectional strokes.

Comparing the two perspectives on repetition, it becomes clear that both agree on the idea that gestures qualitatively differ in their repetition. For LASG, this is a result of gestural complexification, leading to "specific structural and semantic aspects characteristic of these gestural patterns" (39, p46). From the kinematic perspective, repetition might serve as a resource for gestural manipulation, defined not by its communicative effect but by its motor characteristics.[13] The communicative effect emerges afterwards.

In this sense, although departing from different analytical points, both form-based and kinematic approaches lead to similar interpretations of gestures. This is not contradictory, even considering that one uses words and phrases as a landmark, whereas the other uses points in the

[13] In some cases such as bounded back movements or catchments the motor characteristics are tied with the communicative purposes, as pointed by the reviewer.

utterance's phonological grid. It simply leads to the understanding that there are gesture types that tie themselves to speech in different ways. The synchronicity rule proposed by McNeill (8) might then be understood not as a universal theory, in which every gesture is synchronous at all levels. Rather, is the case that one synchronicity sets the path for the others. The keyword here is *domain*. Time span is no longer the most important feature, but rather the domain (phonological, semantic, or pragmatic) to which the gesture is associated with. This association dictates the interpretation of gesture and speech.

A possible path to disentangling those domains and how the associations are built is to go back to conventionalization and abstraction continua. More abstract and less conventional gestures do not provide a basis on which meaning can be directly associated with a landmark. This must be established from the alignment to a phonological instance. Although iconicity or conventionality can be understood from gestures, the span of the association grows into a landmark, as the synchrony serves both gestural and utterance-related meanings.

5 Conclusion

The analysis presented here highlights the complex interaction between gesture phases and their contribution to the construction of meaning. Early perspectives on the subject were centered around saliency and synchrony, emphasizing strokes as being the main vehicle for meaning, tied to phonological, semantic, and pragmatic nuances. Form-based approaches, such as LASG, embarked on subjective definitions, compensating with detailed form descriptions, postulating that meaning is the result of the perception of a *Gestalt* from the gesture. Kinematic approaches based on specific movement features, rather than their form, as velocity peaks and inflection points, to pose a more objective and operationalizable pipeline for gesture analysis, demonstrate that meaning is multilevel and multidimensional.

Both form-based and kinematic approaches, despite their diverging perspectives, converge in interpreting gestures as qualitatively distinct in their repetitions and as dynamic contributors to communication. LASG associates repetition with gestural complexification and semantic nuance, while kinematic methods tie it to motoric characteristics, with communicative effects emerging subsequently. This alignment suggests that gestural synchronicity, as outlined by McNeill, operates selectively across domains rather than universally. To further refine our understanding, the focus shifts to abstraction and conventionalization, where less conventional gestures depend on phonological alignment for meaning, while iconic or conventional gestures expand their associative capacity, intertwining gestural and spoken meanings through layered synchrony.

Acknowledgements

We would like to express our sincere gratitude to the reviewers for their thoughtful comments and insightful discussions, which greatly improved the quality of this paper. Special thanks are extended to Uli Reich and the colloquium of the Romance Philology at the Freie Universität Berlin for their valuable comments and feedback.

On a personal note, I would like to thank Heliana Mello, who first suggested me to study gestures. A one-of-a-kind person who kindly taught me to be as careful and attentive as possible, while also reminding me to give things only as much importance as they deserve. Without her, I would not have gone through half the trouble I did for the sake of research – but I also would have learned far less than I have.

REFERENCES

1. Kendon A. Some relationships between body motion and speech. In: Siegman AW, Pope B, editors. Studies in dyadic communication. New York: Pergamon Press; 1972. p. 177–216 (Pergamon general psychology series; PGPS-7).

2. Chafe WL. Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing. Chicago: University of Chicago Press; 1994.

3. Kendon A. Gesture: Visible action as utterance. Cambridge, New York: Cambridge University Press; 2004.

4. Grice M, Ladd DR, Arvaniti A. On the place of phrase accents in intonational phonology. Phonology 2000; 17(2):143–85.

5. Gussenhoven C, Chen A, editors. The Oxford Handbook of Language Prosody. First edition. Impression: 1. Oxford United Kingdom, New York NY: Oxford University Press; 2020. (Oxford Handbooks in Linguistics).

6. Barros CA, Mello H. The C-ORAL-BRASIL proposal for the treatment of multimodal corpora data: the BGEST corpus pilot project. In: Grajales Ramírez A, Molina Mejía J, Valdivia Martin P, editors. Digital Humanities, Corpus and Language Technology: A look from diverse case studies. University of Groningen Press; 2023. p. 143–62.

7. Raso T, Mello H, editors. C-ORAL-BRASIL: Corpus de referência do português brasileiro falado informal. Belo Horizonte: Editora UFMG; 2012.

8. McNeill D. Hand and mind: What gestures reveal about thought. Chicago: University of Chicago Press; 1992.

9. Loehr DP. Gesture and intonation. Georgetown: Georgetown University; 2004.

10. Meinschaefer J. Sonorität: Sprachstruktur und Sprachverstehen. Tübingen: Narr; 2003. (Cognitio; vol 13).

11. van der Hulst H, Ritter N. The Syllable. De Gruyter; 1999.

12. Fernández EM, Cairns HS, editors. Fundamentals of psycholinguistics. Chichester West Sussex England, Malden MA: Wiley-Blackwell; 2011. (Fundamentals of linguistics).

13. Brentari D. 3. Phonology. In: Pfau R, Steinbach M, Woll B, editors. Sign language: An international handbook. Berlin, Boston: De Gruyter Mouton; 2012. p. 21–54 (Handbücher zur Sprach- und Kommunikationswissenschaft = Handbooks of linguistics and communication science = Manuels de linguistique et des sciences de communication; Bd. 37).

14. Fricke E, Bressem J, Müller C. 123. Gesture families and gestural fields. In: Müller C, Cienki A, Fricke E, Ladewig SH, McNeill D, Bressem J, editors. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2. DE GRUYTER; 2014. p. 1630–40.

15. Bressem J, Müller C. 119. A repertoire of German recurrent gestures with pragmatic functions. In: Müller C, Cienki A, Fricke E, Ladewig SH, McNeill D, Bressem J, editors. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2. DE GRUYTER; 2014. p. 1575–91.

16. McNeill D. Gesture and thought. Chicago: University of Chicago Press; 2005.

17. Gullberg M. Gesture as a Communication Strategy in Second Language Discourse: A Study of Learners of French and Swedish [Doctoral Thesis (monograph)]. Lund University Press; 1998.

18. Müller C. Gesture and Sign: Cataclysmic Break or Dynamic Relations? Front Psychol 2018; 9.

19. Kita S. How representational gestures help speaking. In: McNeill D, editor. Language and Gesture. Cambridge University Press; 2000. p. 162–85.

20. Bressem J. Systems of Gesture Coding and Annotation. In: Cienki AJ, editor. The Cambridge handbook of gesture studies. Cambridge United Kingdom, New York NY: Cambridge University Press; 2024. p. 158–81 (Cambridge handbooks in language and linguistics).

21. Lausberg H. The NEUROGES® Analysis System for Nonverbal Behavior and Gesture: The Complete Research Coding Manual including an Interactive Video Learning Tool and Coding Template. Bern: Peter Lang International Academic Publishers; 2019 [cited 2025 Jan 29]. Available from: URL: https://directory.doabooks.org/handle/20.500.12854/121437.

22. Bressem J, Ladewig SH, Müller C. 71. Linguistic Annotation System for Gestures. In: Müller C, Cienki A, Fricke E, Ladewig SH, McNeill D, Tessendorf S, editors. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/1. DE GRUYTER; 2013. p. 1098–124.

23. Krahmer E, Swerts M. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. Journal of Memory and Language 2007; 57(3):396–414.

24. Bosker HR, Peeters D. Beat gestures influence which speech sounds you hear. Proc Biol Sci 2021; 288(1943):20202419.

25. Esteve-Gibert N, Prieto P. Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements. J Speech Lang Hear Res 2013; 56(3):850–64.

26. Trujillo JP, Simanova I, Bekkering H, Özyürek A. The communicative advantage: how kinematic signaling supports semantic comprehension. Psychol Res 2020; 84(7):1897–911.

27. Trujillo JP, Holler J. The Kinematics of Social Action: Visual Signals Provide Cues for What Interlocutors Do in Conversation. Brain Sci 2021; 11(8).

28. Ladewig SH. 121. The cyclic gesture. In: Müller C, Cienki A, Fricke E, Ladewig SH, McNeill D, Bressem J, editors. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2. DE GRUYTER; 2014. p. 1605–18.

29. Müller C. Redebegleitende Gesten: Kulturgeschichte, Theorie, Sprachvergleich. Berlin: Spitz; 1998.

30. Müller C. A Toolbox of Methods for Gesture Analysis. In: Cienki AJ, editor. The Cambridge handbook of gesture studies. Cambridge United Kingdom, New York NY: Cambridge University Press; 2024. p. 182–216 (Cambridge handbooks in language and linguistics).

31. Müller C, Cienki A, Fricke E, Ladewig SH, McNeill D, Bressem J, editors. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 38/2. DE GRUYTER; 2014.

32. Kendon A. How gestures can become like words. In: Potayos F, editor. Cross-Cultural Peerspectives in Nonverbal Communication. Toronto: C. J. Hugrefe; 1988. p. 131–41.

33. Müller C. Wie Gesten bedeuten Eine kognitiv-linguistische und sequenzanalytische Perspektive. Sprache und Literatur 2010; 41(1):37–68.

34. Goldin-Meadow S, Brentari D. Gesture, sign, and language: The coming of age of sign language and gesture studies. Behav Brain Sci 2017; 40:e46.

35. Kita S, van Gijn I, van der Hulst H. Movement phases in sign and co-speech gestures, and their transcription by human coders. In: Wachsmuth I, Frohlich M, editors. Gesture and sign language in human computer interaction: International gesture workshop, Bielefeld, Germany, Sept. 17-19, 1997 ; proc. Berlin: Springer; 1998 (Lecture notes in computer science [including Lecture notes in artificial intelligence]; vol. 1371).

36. Shattuck-Hufnagel S, Yasinnik Y, Veilleux N, Renwick M. A Method for Studying the Time Alignment of Gestures and Prosody in American English: 'Hits' and Pitch Accents in Academic-Lecture-Style Speech. In: Esposito A, editor. Fundamentals of verbal and nonverbal communication and the biometric issue. Washington DC: IOS Press; 2007 (NATO security through science series, 1574-5597. E. Human and societal dynamics; vol. 18).

37. Yasinnik Y, Renwick M, Shattuck-Hufnagel S. The timing of speech-accompanying gestures with respect to prosody. J Acoust Soc Am 2004; 115(5):97–102.

38. Rohrer PL, Tütüncübasi U, Vilà-Giménez I, Florit-Pons J, Esteve-Gibert N, Ren P et al. The MultiModal MultiDimensional (M3D) labeling system; 2023.

39. Trujillo JP, Simanova I, Bekkering H, Özyürek A. Communicative intent modulates production and comprehension of actions and gestures: A Kinect study. Cognition 2018; 180:38–51.

40. Bressem J. Repetitions in gesture: A cognitive-linguistic and usage-based perspective. Berlin, Boston: De Gruyter Mouton; 2021. (Applications of Cognitive Linguisticsvolume 46).

41. Pouw W, Trujillo J, Rutger Bosker H, Drijvers L, Hoetjes M, Holler J et al., editors. GESPIN 2023: Broadening perspectives, integrating views; 2023.

42. Barros CA. A relação entre unidades gestuais e quebras prosódicas: o caso da unidade informacional Parentético [Unpublished MA Thesis]. Belo Horizonte: Universidade Federal de Minas Gerais; 2021.