This is an accepted version of the following article.

Kita, S., van Gijn, I., & van der Hulst, H. (1998). Movement Phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth & M. Fröhlich (Eds.), *Gesture and sign language in human-computer interaction, International Gesture Workshop Bielefeld, Germany, September 17-19, 1997, Proceedings. Lecture Notes in Artificial Intelligence* (Vol. 1317, pp. 23-35). Berlin: Springer-Verlag.

Movement Phases in Signs and Co-Speech Gestures, and their Transcription by Human Coders

Sotaro Kita[1], Ingeborg van Gijn[2], Harry van der Hulst[3]

[1] Max-Planck Institute for Psycholinguistics
PB 310, 6500 AH Nijmegen, The Netherlands
kita@mpi.nl

[2] Leiden University, Department of Dutch Language
(Current affiliation, University of Amsterdam, Dept. of General Linguistics, Spuistraat 210, 1015 VT Amsterdam, The Netherlands)

[3] Leiden University, HIL / ATW
PB 9515, 2300 RA Leiden, The Netherlands
HULST@rullet.LeidenUniv.nl

Abstract. The previous literature has suggested that the hand movement in co-speech gestures and signs consists of a series of phases with qualitatively different dynamic characteristics. In this paper, we propose a syntagmatic rule system for movement phases that applies to both co-speech gestures and signs. Descriptive criteria for the rule system were developed for the analysis video-recorded continuous production of signs and gesture. It involves segmenting a stream of body movement into phases and identifying different phase types. Two human coders used the criteria to analyze signs and co-speech gestures that are produced in natural discourse. It was found that the criteria yielded good inter–coder reliability. These criteria can be used for the technology of automatic recognition of signs and co-speech gestures in order to segment continuous production and identify the potentially meaning-bearing phase.

1. Introduction

This paper discusses some aspects of form characteristics of signs in sign language and co-speech gestures. Co-speech gesture ("spontaneous gesture" in [1]), henceforth "gesture", is the body movement that is produced spontaneously in the context of speaking, and that has meaning connections to the concurrent speech at the semantic, pragmatic, and discourse levels. Sign language and gesture are both semiotic systems (despite gesture not being a linguistic system), in which certain forms are associated with certain functions in communication ([1]). However, issues concerning the form-function mapping are backgrounded in this paper. The organization of forms will be in focus.

There are different levels of the organization of forms in signs and gestures. In this paper, we focus on the level, where the maximum unit is the movement that starts at the moment of the hand's departure from its resting position (e.g., on the lap) and ends at the moment it returns to its resting position. The excursion of the hand away from the resting position can be segmented into a sequence of discrete phases of different types ([2], [3], [4], [5], [6]). Fig.1 illustrates such segementation of an excusion into different phases (the definitions of different phases will be given in Section 3 and 4). The informant is producing a sign, KLAAR ('ready'), in Sign Language of the Netherlandts, henceforth "SLN". The hand makes the movement that constitutes the word KLAAR in Picture 2, 3, and 4.

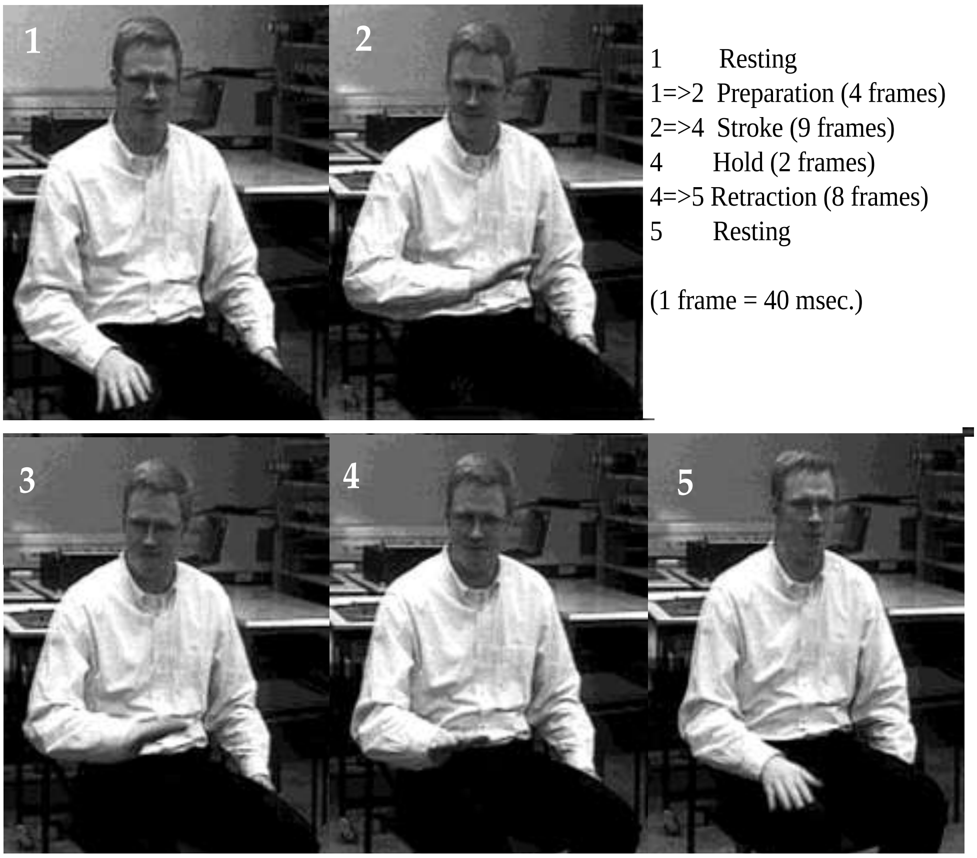

Fig. 1. Movement phases in the SLN sign, KLAAR ('ready')

Previous work on form characteristics of movement phases mostly focused on the meaning-bearing phases, especially in sign language studies. A large body of sign phonology literature focused on the lexical part of the body movement. Some studies on the form characteristics of gestures also focused on meaning-encoding phases ([7], [8], [9]). This paper, in contrast, takes the entire sequence of movement phases into account, and the formal syntagmatic relationships among movement phases are discussed. We will also discuss the methodology of coding movement phases on the basis of video recording of body movements. The coding involves segmenting a stream of body movement into movement phases, with different dynamic characteristics, and identifying the types of phases. We will report on the reliability of this coding across trained human coders. The coding system developed here is in principle implementable on a machine. Thus, this study is relevant to the technology of machine-recognition of signs and gestures, especially when the continuous production of signs and gestures is involved.

This paper is organized as follows. In Section 2, previous studies on movement phases are reviewed. In Section 3, we propose a syntagmatic rule system for movement phases, which holds for both sign language and gesture. In Section 4, we make explicit the coding criteria, with which an observer of video recording can segment a stream of body movement into movement phases and identify phase types . In Section 5, we report on the inter-coder reliability of the coding based on these criteria.

2. Previous Studies on Movement Phases

2.1 Sign Language

In sign phonology, movement phases that represent a lexical item have been the main focus of investigation (e.g., Pictures 2-4 in Fig. 1). Since [10], many sign phonologists have claimed that a sequence of three phonological elements are necessary to account for the form of a sign in which the hand moves from one location to another (the claim has been made at least for American Sign Language (ASL) and SLN). The sequence is Hold--Movement--Hold, using the terminology of [4]. A hand movement is preceded and followed by a holding period, in which the location of the hand does not change. An example of such a Hold was seen in Picture 4 in Fig. 1. Other researchers had overlapping but different conceptions of each element; thus, they used different terminology such as Location--Movement--Location ([5]) and Position--Movement--Position ([6]). Despite their theoretical differences, the basic insight is that a sign in which a hand changes its location consists of 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[1]. Each of the three elements is phonetically realized at least under certain morphological processes (e.g., the phrase-final lengthening in ASL realizes the final hold, [11]).

2.2. Gesture

Unlike sign phonologists, researchers of gesture have taken the entire sequence of movement phases into account. [2] observed that a gesture consists of a series of qualitatively different movement phases, such as "movement to a position", "hold", and "transitional movement". The typology of the movement phases was elaborated and the following hierarchical organization of phases was proposed in [3]. A "gesticular unit" (G-Unit) starts at the point when a limb is lifted away from the body and it ends when the limb moves back to the "resting position", such as on the lap or on the arm rest of a chair. A G-Unit consists of a series of a "gesticular phrase" (G-Phrase), which in turn consists of a sequence of "preparation" -- "stroke"-- "hold" --"recovery". A stroke is an obligatory element in a G-Phrase, and the others are optional. In a stroke, the limb makes an "accented

[1]The signs in which the hand stays in one place consists only of Hold/Location/Position.

movement" with a distinct peaking of "effort" in the sense of Laban ([12]) (Cf. footnote 2). In preparation, the limb moves to the initial position of the stroke. In a recovery (or a return), the limb moves back to the resting position to conclude a G-Unit. When multiple gesticular phrases are concatenated, the recovery can be cut short and becomes a "partial retraction" or completely eliminated.

[3] found that the different levels of the movement hierarchy are functionally distinct in that they synchronize with different levels of prosodic structuring of the discourse in the speech. Furthermore, [3] proposed that a G-phrase is a manifestation of an "idea unit".

[13] (cited in [1] and [4]) pointed out that a stroke can be preceded by a hold, and named such a hold "pre-stroke hold". The hold following a stroke was called a "post-stroke hold". He distinguished the two kinds of holds on the basis of their functions. He also stipulated that there can be a G-phrase without a stroke; namely, the sequence of preparation-hold-retraction is possible. Such a hold was called "stroke-less hold".

The functional motivation to distinguish pre- and post-stroke holds is the following. [13] found that a pre-stroke hold is more likely to co-occur with discourse connectors such as pronouns, relative pronouns, subordinating temporal adverbials (e.g., *while*, *when*, *as*), than a post stroke hold. In an utterance, the discourse cohesion is typically established in the initial part, and the new information is presented in the final part. The interpretation of a gesture tends to semantically overlap with the new information of an utterance ([1]). [13] suggested that a pre-stroke hold is a period in which the gesture waits for speech to establish cohesion so that the stroke co-occurs with the co-expressive portion of speech (See also [14]). [13] also found that a stroke with repetitive movement is less likely to be followed by a post-stroke hold. It was suggested that a post-stroke hold was a way to temporally extend a single movement stroke so that the stroke and the post stroke hold together will synchronize with the co-expressive portion of speech (This idea was first put forth by [15]; See also [14]).

Functional justification of "stroke-less hold" was presented in [16], where it is demonstrated that a hold was a gestural expression of the idea of being in a certain state for a stretch of time. It also proposed that some of the pre- and post-stroke holds were expressions of such an idea, rather than being merely regulatory mechanisms for speech-gesture synchronization.

[1] proposed the rules for movement phases, as in (1) (See also [17] for essentially the same proposal). Here, like [3], a stroke was an obligatory element in a gesture phrase.

(1) Gesture Unit = Gesture Phrases*
Gesture Phrase = (Preparation) ⇒ Stroke ⇒ (Retraction)
Preparation = Preparation ⇒ (Pre-stroke hold)
Stroke = Stroke ⇒ (Post-stroke hold)
Retraction is optional when another Gesture Phrase follows.

<u>Notations</u>:

X = Y	X consists of Y
*	One or more occurrences of the element
⇒	Discrete transition
( )	Optional

[17] proposed that another function of a post-stroke hold might be to maintain the phonological synchrony rule that the stroke co-occurs with the phonologically prominent

syllable of the concurrent intonational phrase. When the stroke is completed "too early", the hand is held until the phonological peak.

Among the above studies, [1] was the only work that made explicit how to transcribe different movement phases from video recording of gestures. Different types of phases can be identified by different foci of "effort" in the sense of Laban ([12])[2]. In a preparation and a retraction, the effort is focused on reaching their respective end points (the beginning of the following stroke and the resting position). In contrast, in a stroke, the effort is focused on "the form of the movement itself -- its trajectory, shape, posture" ([1], p.376). [1] defines the stroke both on the formal and functional grounds. Functionally, the stroke is the "content-bearing part of the gesture".

The previous studies of movement phases in gesture and sign have the following problems. First, the rules for movement phases do not cover the whole range of possibilities observed in the behavior. For example, sign language literature does not discuss the preparation and retraction phase. In the gesture literature, a gesture phrase without a stroke is often not recognized. Second, coding criteria for phases are often not reported in a detailed enough way to allow other researchers to replicate the coding. In the remaining of the paper, we will address these issues.

3. Syntagmatic Organization of Movement Phases in Signs and Gestures.

We propose that n both signs and gesture, the organization of movement phases for one articulator can be described by (2).

(2) Movement Unit = Movement Phrase*
Movement Phrase = (Preparation) ⇒ Expressive Phase ⇒ (Retraction)
Expressive Phase = Independent Hold
Expressive Phase = (Dependent Hold) ⇒ Stroke ⇒ (Dependent Hold)
Preparation = (Liberating Movement) ⇒ Location Preparation >> Hand Internal Preparation
Retraction (when followed by another movement phrase) = Partial Retraction

Notations (See (1) for other notations):
\>> Normally blended but occasionally discrete transition,

The *movement unit*, *stroke*, *holds*, *preparation*, *retraction*, and *partial retraction* are defined in essentially the same way as [1], [3] and [17]. More specific ways to identify each unit will be described in the next section.

2[1] quotes the following passage from ([12], p.11) on body movements in dance:

"When someone moves, you perceive it as more than a change of place or change in the mover's body shape. Movement does not flow along in a monotone -- you see swellings and subsidings, quick flashes, impacts, changes in focus, suspension, pressures, flutterings, vigorous swings, explosions of power, quiet undulations. All this variety is determined by the way in which the mover concentrates his exertion or effort. This exertion or effort may, at any particular moment, be concentrated in the changes in the quality of the tension, or the flow of movement; it may be concentrated in changing the quality of the weight, or the quality of time in the movement; or it may be concentrated in the mover's focus in space"(p.376) .

Unlike [1], [3], and [17], a *movement phrase* does not necessarily contain a stroke. It must contain an *expressive phase.* The expressive phase, in turn, must contain at least either a stroke or an *independent hold.* Thus, an independent hold constitutes an expressive phase by itself. This is in contrast to a *dependent hold*, which flanks a stroke, as in the hold in Fig. 1.

The expressive phase is the semiotically active phase. In an independent hold and a stroke, the form of the body movement is associated with the information to be conveyed. The association can be, for example, on the basis of iconicity in the case of iconic gestures ([1]) and social conventions in the case of signs. Dependent holds are parasitic to the stroke. They arise from the semiotic coordination or modification of the expression in the stroke. For example, it emerges as a result of the coordination between the stroke and the concurrent verbal expression in the case of pre- and post-stroke hold in gestures. In the case of signs, a dependent hold can add other dimensions of meaning (e.g., the end of a sentence and morphological modifications) to the sign expressed by the stroke phase. It is also possible, for sings and gestures, that a dependent hold arises in the coordination of two hands; for example, in a sign where two hands are coordinated, one hand could be held in order to wait for the hand to reach its starting point.

The hold adjacent to a stroke is structurally ambiguous, and a functional analysis is necessary to resolve the ambiguity. Such a hold can be dependent on the stroke, and is subsumed under the same expressive phase as the stroke. Alternatively, such a hold could be the beginning of a new expressive phase (a hold representing a distinct idea from the stroke along the line of [16]).

The preparation starts out with an optional *liberating movement*, which makes the hand free from a constrained position such as undoing of the interlocking of fingers by two hands. The *location preparation* brings the hand to the starting position of the expressive phase, and the *hand-internal preparation* is a hand-internal movement, in which the hand shape and the orientation of the hand are set to the starting values of the expressive phase. Unlike the discrete transition between other types of phases, the transition between other types of phases, the location preparation and the hand-internal preparation overlap in time, in the majority of the cases. Furthermore, the onset of the hand-internal preparation does not precede the onset of the location preparation.

(3) is for one articulator, which is typically an arm and a hand (the articulator could be the head, a leg, or the entire torso; however, we will limit our discussion to the movements of the hand). When two articulators' movement units overlap, there is no constraints on what phase types can overlap. For example, the right hand's stroke can overlap with the left hand's preparation for the next expressive phase. However, when two strokes coincide or when a stroke and a hold coincide, there are restrictions as to what two limbs can do. When two strokes by two hands coincide in sign language, the movements obey the well-known Symmetry Condition, which states that the movement trajectory, the hand orientation, the hand shape, and the hand-internal movement have to be either the same or symmetrical (see, e.g., [18]). [19] and [20] have shown that the Symmetry Condition also holds for gestures. When a stroke and an independent hold coincide, it is well-known that in sign language the choice of the holding hand's hand shape is severely limited (see, e.g., [18]).

4. Descriptive Criteria for the Segmentation and Identification of Movement Phases

In order to further the understanding of movement phases, it is necessary for the scientific community to share a set of descriptive criteria for the units of analysis. The purpose of this section is to provide such criteria. They are based purely on the dynamics of the body movement; that is, the meaning and the function of the body movement are not used.

We consider a phase to be a manifestation of a planning unit in motor control, which presumably reflects units at a higher level of cognitive planning for signing and gesturing. The challenge is to identify units in motor control on the basis of video-recording of body movements, without any physiological recordings. The following is the procedure (it partly follows the procedure of [1] (pp. 375-376).

1) Identification of a Movement Unit

The first indication of initiating a movement for a sign or a gesture is the onset of *movement unit*. This usually precedes actual departure of the hand from the resting position. The end of a movement unit is the moment at which the hand makes the first contact with the resting surface. After this moment, some time may pass before the hand finally settles, which is not counted as part of the movement unit. The *resting position* is part of the body or the furniture where the hands can be supported, such as the lap, the chest in the case of folded arms, the arm rest of a chair, and the table top. It could also be the position where self-adapting body movements take place, such as combing the hair with fingers and adjusting clothes, or the position where object manipulation (e.g., grasping a coffee cup) starts. In signing, the hand often rests in the air in front of the chest, in order to be able to quickly launch the movement unit for the next sign. In this case, the elbow is flexed and is very close to the side of the body, and the hand is close to the chest and often has a neutral hand shape with the fingers and the palm naturally curving.

2) Segmentation of phases

2.1). A *movement phrase* is segmented into a sequence of discrete phases. A stretch of movement is divided into two phases if there is an abrupt change of direction in the hand movement, AND there is discontinuity in the velocity profile of the hand movement before and after the abrupt direction change. A short pause constitutes such discontinuity, so does having two stretches of the movement with distinct velocity profiles (e.g., a stretch with increasing speed and a abrupt stop at the end, and then a stretch with a constant speed). The movement can be a so-called path-movement, which changes the location of the wrist in the space relative to the torso, or a purely hand-internal movement, in which the wrist stays in one location but makes a hand shape change, palm orientation change, and finger orientation change[3]. A gradual hand-internal change accompanying a path-movement is not relevant to the segmentation of phases.

A two-segment movement without velocity-profile discontinuity is coded as one phase even if there is an abrupt change of direction. These are called a *multi-segment phase* as

[3] Finger orientation is defined as the vector from the wrist to the knuckles ([19]).

opposed to a *single-segment phase*[4]. A multi-segment phase is observed, for example, when the hand moves as if its finger tip traces the outline of an object with sharp corners such as a square, or when the hand moves as if it enacts an action in which the limb moves in more than one directions with the same velocity profile, such as reaching out for an object and pulling it closer to the body.

However, two segments with velocity-profile discontinuity are considered to be subsumed under one phase if 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). Such a *semi-multi-segment* phase is observed, for example, when the hand moves as if it represents an event in which a moving object hits a hard surface.

2.2). If the same movement, is repeated without any hold in between, the entire duration of repetition is considered to be one *repetitive phase* (e.g., a gesture, in which hammering action is enacted) and is not counted as different phases.

2.3). When a repeated movement is superimposed on a path-movement, the entire duration of the "hosting" path-movement is the phase. Note that the size of the repeated movement may gradually change. When a path-movement is repeated, the value for hand-internal form features may change in the course, but this does not affect the phase segmentation, either.

3) Identification of phase types

3.1). A phase, in which more force is exerted than neighbouring phases, is a *stroke*. Note that acceleration (and deceleration) are good indicator of the exerted force, but sometimes a downward retraction has bigger acceleration than a stroke because of the gravity. A multi-segment phase, semi-multi-segment phase, and a repetitive phase are always a stroke.

3.2). A phase, in which the hand is held still, is a *hold*. The hand is rarely perfectly still. The decision is relative to the neighbouring phases. The hand that is drifting about, whose movement has no discernible target direction, is considered to be in a hold phase. Sometimes a hold is performed hand-internally (e.g., with a distinctively 'active' hand shape) at the position that would otherwise be a resting position.

3.3). A non-stroke phase that departs from the resting position is a preparation, and that arrives at the resting position is *retraction*.

3.4). If there is one phase with a moving limb between two strokes, that is a preparation. Occasionally, especially in gestures, one sees that after a stroke the hand makes a non-stroke movement that goes toward a potential resting position, but before reaching the resting position shifts to a preparation for another stroke. The interrupted retraction is called *partial retraction*. In order to be coded as partial retraction, the movement has to exhibit a sign of increasing relaxation as well as the movement toward a potential resting place.

3.5). The preparation phase consists of *liberating movement* followed by *location preparation* (the location of the wrist) and *hand-internal preparation* (the hand shape and the orientation of the palm and the knuckles). Liberating movement frees hands from a constrained position such as the position beneath another hand in the resting position, or the

---

[4] [1] proposes similar distinctions among strokes. (p.380) The multi-segment phase subsumes the "bi-directional stroke". The "Uni-1" stroke, which is "unidirectional, one movement", is the same as the single-segment phase. The "Uni-2" stroke, "two movements, effort exerted in one direction, the other movement returning the hand to starting place"(p.380) is a sequence of two single-segment strokes.

interlocking of fingers with another hand or clothing. The location preparation and hand internal preparation overlaps in time in most of the cases, but occasionally they constitutes two consecutive discrete phases.

5. Inter-coder reliability of segmentation and identification of movement phases

We have conducted a test to assess to what degree two human coders can analyze, in the same way, a stream of video recorded body movement into a sequence movement phases.

5.1 Procedure

Coders. One of the authors (Kita), henceforth "the fist coder", and a research assistant, henceforth "the second coder"[5], were the coders. Both coders had a few years of experience in coding gestures. Prior to this test, they had never coded signs. The first coder had very limited knowledge of SLN, and the second coder had no knowledge of SLN. They were both familiar with the animated cartoon used to elicit gestures an signs.

Data Collection. The video recordings were collected in the following way. Gestures were produced by randomly sampled four Dutch speakers and signs were produced by randomly sampled three signers of SLN. Every informant watched the same short animated cartoon clips[6]. After watching the clips, the informant told the story to another person. The story telling was video taped with a consumer-model PAL Hi-8 cameras. In both cases, the task was explained as a part of research on story-telling. In the case of Dutch speakers, it was not mentioned that gestures were the subject of the study.

Coded Material and Coding Apparatus. Two random stretches of the recordings were selected for each informant. One stretch was coded for the right hand movements, and the other stretch was coded for the left hand movements (for technical reasons only one stretch was coded for one of the signers). Each stretch included 10 to 18 gesture phases in the first coding. The second coder did not know how many phases there are in the first coding.

Coding was performed on the basis of digitized video (QuickTime movies on a Macintosh computer. The movie format was Cinepak with 384*288 pixels). For annotating digital movies with coding information, a software developed at Max-Planck Institute for Psycholinguistics, called MediaTagger ([21]), was used. In this coding environment, the picture quality was greatly compromised compared to the original recording because of digitization into JPEG and data compression from JPEG to Cinepak. However, the movie playing capability of MediaTagger (based on the QuickTime movie playing library) allows easy access to a noise-free stable still-picture of each frame and play-back of a specified stretch of movie.

---

5 We would like to acknowledge Lisette Oliemeullen for participating in the inter-coder reliability test.

6 The animated cartoon was the same as used in [1]. See page 366 to 374 for the detailed description of it. The parts for which we analyzed signs and gestures were from Line 1 to Line 36, and from Line 197 to Line 232.

Coding. The first and the second coders coded movement phases independently according to the descriptive criteria in Section 4. Coding was based only on the visual information (i.e. sound was muted during coding). The coders spent about the same amount of time in coding as they would do in other research projects.

After the independent coding, the two coders viewed the body movements together, and discussed their coding decisions regarding the gross-level segmentation and the identification of the phase type. The gross-level segmentation is recognition of a stretch of movement with a certain directionality (single-segment, semi-multi-segment, or multi-segment) as a phase, regardless of the exact location of the boundaries and the identification of the phase type (preparation, stroke, retraction, partial retraction, or hold). The two codings were regarded as matching on the gross-level segmentation, when it was confirmed that the two coders saw roughly the same stretch of movement as a phase with the same directionality.. When there was an disagreement in the gross-level segmentation or identification, it was checked whether one can agree with the other's coding.

5.2 Results

Inter-coder Reliability. First, we will present the match of the initial independent codings by the two coders. Table 1 summarizes the average number of phases that are recognized for each informant.

Table 1. Inter-coder reliability of the initial independent codings of the number of phases, for four gesturers and three signers

	Average number of phases in the 1st coding	Average number of phases in the 2nd coding	Average proportion of the discrepancy between the two codings to the 1st coding.
Gestures	23.3	24.5	10% (8%)
Signs	23.0	26.0	13% (5%)

Note: The number in ( ) is the estimated population standard deviation across informants.

For each phase recognized in the first coding (20, 22, 24, and 27 phases for the four gesture informants, and 26, 30 and 13 phases for the three sign informants), the following things were checked during the discussion between the two coders. The first is whether there is a phase in the second coding that matches the phase in the first coding at the level of gross segmentation. For such matching phases, then it was checked whether the identification of the phase type matched. The results are summarized in Table 2.

Table 2. Inter-coder reliability of the initial independent codings of gross segmentation and the phase type, for four gesturers and three signers

	Average % of phases, in the first coding, for which the 2nd coding matched on the gross segmentation	Average % of phases, in the first coding, for which the 2nd coding matched on the gross segmentation and the phase type
Gestures	76% (9%)	72% (11%)
Signs	73% (10%)	58% (5%)

Note: The number in ( ) is the estimated population standard deviation across informants.

The main reasons of disagreement in the gross-level segmentation were missing a spatially small movement, the decision on whether to divide of a movement into a preparation and a stroke (as opposed to a long stroke), and grouping of movement segments into a semi-multi-segement and a multi-segment phase. For signs, the match for the phase-type identification was considerably lower than for gestures, probably for three reasons. First, the coders were not experienced in signs. Second, in signs, there are more phases with short duration (see the next subsection). This often makes it difficult to identify a stroke, which is supposed to be more forceful compared to its neighbouring phases. The relative forcefulness is often decided on the basis of the amount of acceleration, which is difficult to assess based on images from a few frames. Third, the signs went back to the resting position less frequently than the gestures (see the next subsection).

For the boundaries of the phases for which the two initial codings matched on the gross segmentation, the exact boundary locations of the two codings were compared (Table 3).

Table 3. Inter-coder discrepancy of the boundaries for the phases, for which the initial independent codings matched on the gross-level segmentation

	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5
G	1%	0%	0%	2%	2%	3%	2%	8%	35%	26%	13%	4%	4%	2%
S	0%	0%	0%	0%	0%	4%	4%	16%	39%	26%	6%	6%	0%	0%

Note: The first row represents the frame number of the 1st coding minus that of the 2nd coding for a given phase boundary (1 frame = 40 msec.). "G" stands for gestures (N=109), and "S" stands for signs (N=80).

For gestures, the two codings located 69% of the boundaries within one frame, and 84% of the boundaries within two frames. For signs, the boundaries matched better. 81% of the boundaries in the two codings fell within one frame from each other, and 91% within two frames. This is probably due to the movements in signs having higher speed (, though there was no means for us to objectively measure it), making the frame of the abrupt change of direction unambiguous.

For the phases in which the two initial independent codings do not match in the gross segmentation and the phase type, the two coders discussed whether one decision is better than the other. As summarized in Table 4, the two coders agreed on this issue for a quite high percentage of phases.

Table 4. Inter-coder reliability after the discussion between the two coders, for four gesturers and three signers

	Average % of phases, in the first coding, for which the 2nd coding matched on the gross segmentation	Average % of phases, in the first coding, for which the 2nd coding matched on the gross segmentation and the phase type
Gestures	92% (3%)	88% (5%)
Signs	93% (3%)	88% (5%)

Note: The number in ( ) is the estimated population standard deviation across informants.

The level of agreement was the same for gestures and signs. This suggests a possibility that when coders become more experienced with signs, then the discrepancy between gestures and signs in the identification of the phase type in Table 1 will reduce. There are cues in the recording, which the two coders did not consistently pick up but eventually agreed to use.

It is worth noting here that there are a number of factors that could influence the level of inter-coder reliability. A better spatio-temporal resolution of video images and simply spending more time could improve the match between the initial independent codings. An improvement on the descriptive criteria can also contribute to the eventual level of agreement. Also, the knowledge of SLN and of the content of the concurrent speech would contribute to higher match at both initial and eventual match among coders.

Differences between Gestures and Signs. The following differences between signs and gestures could have caused the different inter-coder reliability levels. The data in this subsection are drawn from the gestures and signs in which the two coders eventually agreed on the gross segmentation and the phase type (the last column in Table 4).

Table 5. Frequency of phase types

	preparation	stroke	retraction	partial retraction	hold
Gestures (N=88)	31%	39%	18%	1%	11%
Signs (N=72)	32%	47%	7%	1%	13%

As shown in the frequency of retraction in Table 5, the hand of signers went back to the resting position about half as often as gestures; in other words, their movement phrases were twice as long as those of gestures in terms of the number of phases.

The phase length is different between gestures and signs and also between holds and active phases (i.e. preparation, stroke, partial retraction, and retraction). As in Fig.2, active phases in signs tend to be shorter (the average duration, 6.7 frames; standard deviation 5.4 frames) than the ones in gestures (the average duration, 8.5 frames; the standard deviation 4.9 frames). For holds, the tendency is reversed. As in Fig.3, holds in signs tend to be longer (the average duration, 26.3 frames; the standard deviation 30.4 frames) than the ones in gestures (the average duration, 13.6; the standard deviation 14.4 frames). In general, holds are far longer than the active phases.

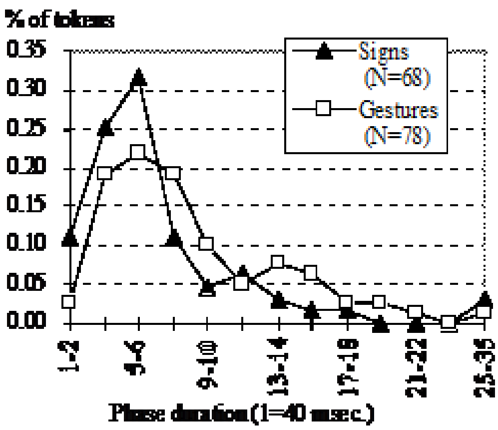

Fig.2. Length of active phases

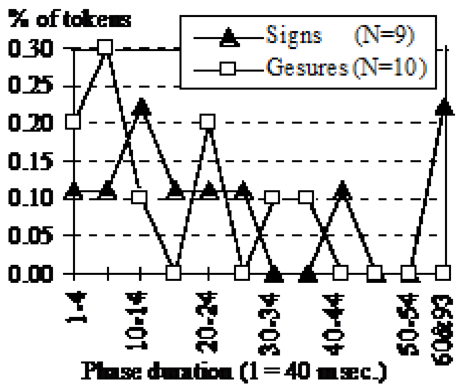

Fig.3. Length of holds

Conclusions

The syntagmatic rules for movement phases, (2), were descriptively adequate for both signs and gestures that were produced during a natural narrative task. The descriptive criteria in Section 4 yielded good inter-coder reliability between two human coders in the analysis of video-recordings of the signs and gestures. The match between the two initial independent codings of the gross-level segmentation and the phase type identification was better for gestures than for signs. This may be due to the frequency of retractions and phases with short duration. For the phases for which the two independent codings matched on the gross segmentation, the exact location of the phase boundaries in the two codings were, in majority of the cases, within two frames. Most of the initial mismatches between the two coders were resolved in the discussion of their coding decisions, and they eventually agreed on about 90% of the phases. This suggests that it is in principle possible to further improve the match at the initial independent coding. This method of analyzing continuous production of gestures and signs, which is based purely on the formal basis, could also be applied to automatic recognition of signs and gestures in order to locate the potentially meaning-bearing phases in the continuous production.

References

[1] McNeill, D.: Hand and Mind. University of Chicago Press, Chicago (1992)

[2] Kendon, A.: Some Relationships Between Body Motion and Speech. In: Siegman, A., Pope, B. (eds.): Studies in Dyadic Communication. Pergamon Press, New York (1972) 177-210

[3] Kendon, A.: Gesticulation and speech: Two Aspects of the Process of Utterance. In: Key, M. R. (ed.): The Relation Between Verbal and Nonverbal Communication. Mouton, The Hague (1980) 207-227

[4] Liddel, S.K., Johnson R.E.: American Sign Language -- Compound Formation Processes, Lexicalization, and Phonological Remnants. Natural Language and Linguistic Theory. **4** (1986) 445-513

[5] Sandler, W.: The Spreading Hand Autosegment of American Sign Language. Sign Language Studies **50** (1986) 1-28

[6] Perlmutter, D.M: Sonority and Syllable Structure in American Sign Language. Linguistic Inquiry **23** (1992) 407-442.

[7] Birdwhistell, R. L: Kinesics and Context. University of Pennsylvania Press, Philadelphia (1970)

[8] Calbris,G.: The Semiotics of French Gesture. Indiana University Press, Bloomington (1990)

[9] Kendon, A.: Gestures as Illocutionary and Discourse Structure Markers in Southern Italian Conversation. Journal of Pragmatics (1995) **23** 247-279

[10] Liddel, S.K.: THINK and BELIEVE: Sequentiality in American Sign Language. Language **60**.372-399 (1984)

[11] Liddel, S.K. Structure for Representing Hand Shape and Local Movement at the Phonemic Level. In: Fischer, S., Siple, P. (ed.): Theoretical Issues in Sign Language Research, Vol.1. University of Chicago Press, Chicago (1990) 37-65

[12] Dell, C.: A Primer for Movement Description: Using Effort-Shape and Supplementary Concepts. Dance Notation Bureau, New York (1970)

[13] Kita, S.: The Temporal Relationship between Gesture and Speech: A Study of Japanese-English Bilinguals. Master's thesis, University of Chicago (1990)

[14] de Ruiter, J.: Production of Gesture and Speech. unpublished Ph.D. dissertation, University of Nijmegen (1998)

[15] McNeill, D.: A Straight Path -- To Where? Reply to Butterworth and Hadar. Psychological Review. **96** (1989) 175-179.

[16] Duncan, S.: Grammatical Form and 'Thinking-for-Speaking' in Mandarin Chinese and English: An Analysis Based on Speech-accompanying Gesture. Unpublished Ph.D. dissertation, University of Chicago (1996)

[17] McNeill, D., Levy, E.T., Pedelty, L.L.: Gesture and Speech. In: Hammond, G.R. (ed.): Advances in Psychology: Cerebral Control of Speech and Limb Movements. Elsevier/North Holland Publishers, Amsterdam (1992) 203-256

[18] van der Hulst, H: On the Other Hand. Lingua. **98** (1996) 121-143

[19] van Gijn, I.: Vormkarakteristieken van Gestures en Gebaren [Form Characteristics of Signs and Gestures]. Master's thesis, Leiden University (1997)

[20] van Gijn, I., Kita, S., van der Hulst H.: Symmetry Condition in Sign Language Phonology and Gesture.(in preparation)

[21] Brugman, H., Kita., S.: Impact of Digital Video Technology on Transcription: a Case of Spontaneous Gesture Transcription. KODIKAS/CODE: Ars Semeiotica **18** (1995) 95-112