Cogn Comput
DOI 10.1007/s12559-010-9071-2

Movements and Holds in Fluent Sentence Production of American Sign Language: The Action-Based Approach

Bernd J. Kröger · Peter Birkholz · Jim Kannampuzha · Emily Kaufmann · Irene Mittelberg

Received: 7 April 2010 / Accepted: 1 September 2010

Abstract The importance of bodily movements in the production and perception of communicative actions has been shown for the spoken language modality and accounted for by a theory of communicative actions, Cogn. Process. 2010;11:187–205. In this study, the theory of communicative actions was adapted to the sign language modality; we tested the hypothesis that in the fluent production of short sign language sentences, strong-hand manual sign actions are continuously ongoing without holds, while co-manual oral expression actions (i.e. sign-related actions of the lips, jaw, and tip of the tongue) and co-manual facial expression actions (i.e. actions of the eyebrows, eyelids, etc.), as well as weak-hand actions, show considerable holds. An American Sign Language (ASL) corpus of 100 sentences was analyzed by visually inspecting each frame-to-frame difference (30 frames/s) for separating movement and hold phases for each manual, oral, and facial action. Excluding fingerspelling and signs in sentence-final position, no manual holds were found for the strong hand (0%; the weak hand is not considered), while oral holds occurred in 22% of all oral expression actions and facial holds occurred for all facial expression actions analyzed (100%). These results support the idea that in each language modality, the dominant articulatory system (vocal tract or manual system) determines the timing of actions. In signed languages, in which manual actions are dominant, holds occur mainly in co-manual oral and co-manual facial actions. Conversely, in spoken language, vocal tract actions (i.e. actions of the lips, tongue, jaw, velum, and vocal folds) are dominant; holds occur primarily in co-verbal manual and co-verbal facial actions.

Keywords Sign language · Spoken language · Action theory · Manual actions · Oral actions · Facial actions · Co-verbal actions · Co-manual actions · Movements · Holds

Introduction

Three articulatory systems can be distinguished in signed as well as spoken language in face-to-face communication (Fig. 1): the *manual system* (the left and right hands); the *vocal tract system* (the lips, jaw, tongue, velum, and vocal folds), including its visible components, called the *oral system* (i.e. the lips, jaw, and visible part of the tongue); and the *facial system* (eyebrows, eyelids, corners of the mouth, etc.). The manual system is dominant in the signed language modality, since manual signs convey the bulk of the linguistic information. The vocal tract system (or verbal system, see Table 1) is dominant in the spoken language modality, since spoken words convey the linguistic information in this modality. The facial system is not dominant from a linguistic viewpoint in either of these two modalities, since facial actions do not primarily convey linguistic information. However, facial expression actions are very important in communicating paralinguistic information such as the emotional state of the speaker (see the "facial action coding system" FACS, [6, 10]). While the dominant articulatory system in a language modality is defined as the system which, in the absence of all other systems, is capable of carrying the *linguistic content* of a sentence, all other (i.e. non-dominant) articulatory systems play an

B. J. Kröger (✉) · P. Birkholz · J. Kannampuzha
Department of Phoniatrics, Pedaudiology and Communication Disorders, RWTH Aachen University, Aachen, Germany
e-mail: bkroeger@ukaachen.de

E. Kaufmann · I. Mittelberg
Human Technology Centre,
RWTH Aachen University, Aachen, Germany

Journal : Large 12559 | Dispatch : 23-9-2010 | Pages : 17
Article No. : 9071 | □ LE | □ TYPESET
MS Code : | ☑ CP | ☑ DISK

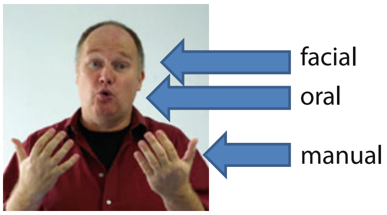

Fig. 1 Articulatory systems (facial, oral, and manual) in the signed language modality. Midsagittal views of the vocal tract system are displayed in Fig. 3 for the spoken language modality

important role for signaling additional information in face-to-face communication (e.g. [13, 16, 17, 29, 30] for co-verbal gesturing; see e.g. [12, 34] for co-manual mouthing). It should be noted that with reference to action theory [23], the term *gesture* is generally replaced by the term *action* in this paper. A detailed definition of actions in each articulatory system will be given in the next paragraph.

A complication in distinguishing the three articulatory systems introduced above results from the fact that the lips, jaw, and front part of the tongue are part of the oral system (and thus of the vocal tract system) as well as the facial system. On the one hand, these articulators are involved in the realization of speech actions in the spoken language modality as well as in the realization of mute (sometimes speech-like) oral expression actions ("mouthings") in the signed language modality (e.g. [12]). On the other hand, the oral articulators are involved in expressing emotional states as well (e.g. corner of the mouth pulling up in "happiness", jaw lowering in "surprise"). Thus, these articulators are controlled with respect to linguistic functions (i.e. the production of speech or speech-like actions) as well as with respect to paralinguistic functions.

Table 1 summarizes the functions of the three articulatory systems (manual, oral, and vocal) in the two language modalities (signed and spoken). For example, facial actions occur as co-verbal in the spoken language modality and as co-manual in the signed language modality. Also, oral actions occur as co-manual actions in the signed language modality, while manual actions occur as co-verbal actions in the spoken language modality. In addition to signaling emotional states (e.g. "happiness", "sadness", and "surprise"), in sign language, facial actions may also convey prosodic information such as "I have no idea" (i.e. a facial expression of uncertainty for communicating "this is a question"; see the sign language example sentence given in the next paragraph) or "that is really true" (i.e. facial expression of confidence for communicating "this is a statement"). Thus, facial actions in both modalities in most cases are temporally coordinated with a phrase or with a whole sentence (i.e. "phrase timing", see Table 1). But a considerable portion of co-verbal manual actions in the spoken language modality as well as co-manual oral actions in the signed language modality are clearly related to specific lexemes, i.e. to spoken words in the spoken language modality (e.g. [19]) or to manual signs in the signed language modality (e.g. [12]) (called "word timing" here; see Table 1). For example, in the spoken language modality, two co-verbal hand actions exhibiting an extended index finger may occur in temporal synchrony with the words "there" in the sentence "You have to go there and not there!" in order to specify the words "there" in more detail by communicating specific directions. Or, in the signed language modality, a co-manual mute labio-dental mouth closing action which signals the (spoken) sound /v/ may temporally co-occur with the manual ASL sign for the word ARRIVE in order to emphasize the lexical meaning of that manual sign (see the American Sign Language sample sentence "WOMAN ARRIVE HERE?", described in detail in the next paragraph).

It is hypothesized in our action-based approach—which can be used as a concept for quantitatively modeling cognitive and sensorimotor aspects of communication in the signed or spoken language modality—that communicative actions occurring in the dominant articulatory system are executed expediently in order to convey information from speaker to listener quickly and efficiently. These actions (a sign in the signed language modality; a word in the spoken language modality) are sequenced without pauses. Thus, the execution of a subsequent action can basically be started when the movement phases of all important movement sub-actions of the currently executed action within a sentence have been accomplished. This leads to the hypothesis that no articulatory holds occur in the

Table 1 Dominant and non-dominant actions and articulatory systems in the spoken and signed language modalities

Modality	Spoken languages	Signed languages
Dominant	Verbal actions = vocal tract actions vocal tract articulatory system	Manual actions = hand-arm actions manual articulatory system
Non-dominant; word timing	Co-verbal manual actions manual articulatory system	Co-manual mute oral actions oral articulatory system
Non-dominant; phrase timing	Co-verbal facial actions facial articulatory system	Co-manual facial actions facial articulatory system

For the terms *narrow* and *broad timing*, see the text dominant articulatory system. In the case of signed language, no articulatory holds occur in the manual system (considering the strong hand only) in the case of fluent signing, while in the case of spoken language, no articulatory holds occur in the vocal tract system. Furthermore, since actions of non-dominant articulatory systems occur in temporal synchrony with specific actions of the dominant articulatory system and since it can be assumed that a sentence comprises more movement actions in the dominant articulatory system than in the non-dominant articulatory system, it can be hypothesized that holds mainly occur for movement actions of non-dominant articulatory systems in each language modality. In other words, the rate of information given by the dominant system is higher than that of the non-dominant systems, but movement actions occurring in all three systems are temporally well correlated, and thus the timing of actions in the dominant system dictate the timing of actions in the non-dominant systems.

The Cognitive and Sensorimotor Concept of Communicative Actions

Communicative actions are the only vehicle humans are able to use to transfer information in face-to-face communication [23]. These actions comprise *meaning* and *form*. The *meaning* of a communicative action is specific linguistic and/or paralinguistic information the speaker intends to transfer to the interlocutor. The *form* of each communicative action is composed of a number of temporally well-coordinated, goal-directed movement subactions, which are called *elementary movement actions* (see the examples given below in this paragraph for different articulatory systems). On the one hand, communicative actions (defined by their meaning) and the discrete descriptions of elementary actions (of which each communicative action is composed) are both *cognitive entities*. On the other hand, the planning, programming, and execution of a communicative action are *sensorimotor processes*. Thus, the concept of action in spoken as well as in signed languages connects *cognitive linguistic and paralinguistic entities* with their *phonetic or sensorimotor realizations* in production. Moreover, the perception and understanding of spoken or signed language units can be exemplified in action theory with respect to this dual nature of actions. On the one hand, the cognitive neural state representing a communicative action must be activated in order to understand that action. On the other hand, communicative actions comprise elementary movement action units which are the basic vehicles in production as well as in (lower level) perception (see [24] for manual actions). Thus, when elementary movement actions are perceived, it is very likely that there is an activation of neurons that represent the appropriate cognitive neural state of the superordinate action, since the association of sensorimotor neural states, which represent elementary movement actions, with cognitive neural states of a superordinate communicative action have already been established for frequently occurring actions during action acquisition (see [23]).

In the case of communicative actions, the goal of each subordinate elementary movement action is *shape-forming*. This shape can be a specific vocal tract configuration which produces the specific features of a speech sound; a specific shape of parts of the face (e.g. eyebrows, eyelids, and corners of the mouth) which together produce a specific facial expression; a specific hand orientation, position, and hand shape which together produce a specific lexical sign language item; or a specific shape of parts of the oral region which together produce a specific mute oral expression. Shapes accomplished by elementary movement actions are distinct, and the amount of shapes required for a meaningful action is comparable to a bundle of distinctive features which defines that action.

The basic temporal form of each elementary movement action is a *movement phase* followed by a *target phase* (Fig. 2). During the movement phase, all articulators involved in an elementary movement action (e.g. arm, palm of the hand, and fingers in the case of a manual action; jaw, lower and upper lips, and tongue in the case of an oral action) act in a synergetic way in order to reach the final shape or target. From a perceptual viewpoint, it can be argued that final shape or target need not be fully reached in order for the listener to perceive the goal of the elementary movement action satisfactorily (see [23]). Thus, the target phase of an elementary movement action which exhibits the target shape may be relatively short or even absent without any negative impact on the perception of the important distinctive shapes which are necessary for understanding the whole communicative action.

In the case of spoken languages, *elementary vocal tract movement actions* are, for example, bilabial, apical, and dorsal closing actions; glottal and velo-pharyngeal opening or closing actions, and shape-forming vocal tract actions

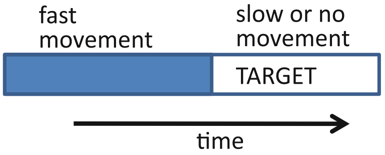

Fig. 2 Temporal organization of an elementary movement action. Its time course can be divided into a movement phase (*blue rectangle*) followed by a target phase (*white rectangle*). The target is nearly reached during the movement phase. Many elementary movement actions do not exhibit a pronounced target phase, since target perception is mainly done during the movement phase

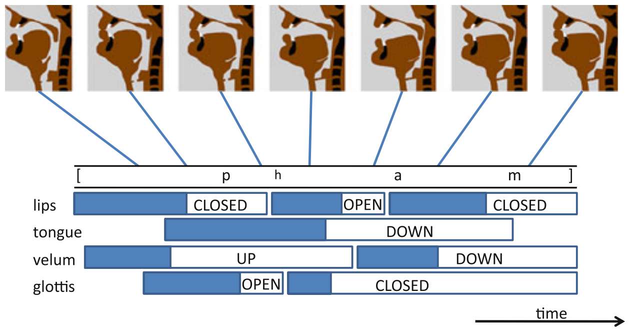

Fig. 3 *Top*: Midsagittal views of the vocal tract for different points in time during the production of the spoken English word "palm". *Middle*: Phonetic transcription of the sound segments of this word. *Bottom*: Action score for the production of this word using an articulatory speech synthesis system [20, 22]. Target phases of elementary movement actions are labeled with words in *capital letters*. Movement phases of elementary movement actions are always indicated as blue rectangles (cf. Fig. 2)

(see Fig. 3). A meaningful *verbal or speech action* (e.g. a word) is composed of a number of specific elementary vocal tract actions which must be well coordinated in time in order to produce that action. As an example, all elementary vocal tract movement actions needed for the production of the English word "palm" are given in Fig. 3. The /p/ is produced by a labial closing action temporally coordinated with a velopharyngeal closing action (i.e. to raise the velum) and a glottal opening action. The /a/ is produced by a tongue-lowering action temporally coordinated with a glottal closing action. The /m/ is produced by a second labial closing action temporally coordinated with a velo-pharyngeal opening action (i.e. to lower the velum) and with a glottal closing action. In order to combine these three sounds properly for the word "palm", the vocalic tongue lowering action must also be hidden behind the first labial closure in order to avoid a diphthong-like sound. Furthermore, /a/ and /m/ profit from phonation resulting from the same glottal closing action. Thus, a *meaningful or lexical verbal action* is a spoken word which comprises a number of temporally well-coordinated labial (i.e. lips), lingual (i.e. tongue), velo-pharyngeal (i.e. velum) and glottal (i.e. glottis) movement actions (*vocal tract movement actions*, see Fig. 3). If a word comprises more than one syllable, a syllable tier must be introduced as an intermediate level for *syllable sub-actions*. It is assumed that the temporal organization of elementary vocal tract movement actions is always done on a syllable level (c.f. [14]). It should be noted, as can be seen in Fig. 3, that during the production of fluent speech, the movement of at least one vocal tract articulator occurs at all points in time; in other words, no *absolute holds* can be found in the vocal tract system during fluent speech production. Thus, in our example, movement phases of elementary movement actions (marked by blue rectangles in Fig. 3) occur at each point in time during the production of this word with the exception of the final part of /m/. But in fluent speech, i.e. when the word "palm" is produced in context, in this final part the next vocalic tongue movement action is normally already active for the preparation of the vowel of the following syllable. It should be noted that the concept of vocal tract movement actions as basic units or "atoms" of speech production was introduced by Browman and Goldstein [4, 5] as a concept connecting phonology (cognitive entities) and phonetics (speech articulation), and this concept was further developed by Saltzman and Byrd [33], Goldstein et al. [14, 15] and adopted by Kröger and Birkholz [20, 21], by Kröger et al. [22], and by Bauer et al. [2] for German.

In the case of signed languages, the *elementary manual movement actions* are path, orientation, shape, and secondary movement actions (as they will be called throughout this paper). *Path* movement elementary actions lead to specific hand movement directions or locations; *orientation* movement elementary actions lead to specific hand orientations; *shape* movement elementary actions lead to specific hand shapes; and *secondary* movement elementary actions like finger wiggling may be overlaid on other elementary movement actions or occur during holds. For basic concepts of direction, location, orientation, shape, and secondary movement in sign language, see Liddell and Johnson [26], Klima and Bellugi [18], Perlmutter [32], Stokoe [38, 40, 42]. The composition of a meaningful or lexical *sign action* made up of elementary manual movement actions is complex. First, each sign action comprises at least one *strong-hand manual sub-action*, but it may comprise a sequence of two or more strong-hand manual sub-actions; in addition, it may comprise a *weak-hand manual sub-action* (for the terms strong and weak hand see e.g. [40]). Each strong-hand (as well as each weak-hand) manual sub-action comprises at least one path, orientation, shape, or secondary movement elementary action or a temporally overlapping combination of two or more of these elementary manual movement actions. For example, in the sample sentence given below (Fig. 4), the sign action for WOMAN comprises a sequence of two strong-hand manual sub-actions. For this sign, a shape and orientation movement elementary action together with a path movement elementary action establishes the first strong-hand manual sub-action (i.e. spread hand shape, palm orientation to the right and contact with the chin at the end of the path movement action), while an orientation movement elementary action establishes the second strong-hand manual sub-action (palm-up orientation; see Fig. 4 and Table 2). In contrast, the sign action for ARRIVE comprises a temporally co-occurring (asymmetrical) strong- and weak-hand manual sub-action. (In symmetrical signs, the weak hand mirrors the strong hand; in asymmetrical signs, the weak hand functions as a base upon which the strong hand acts,

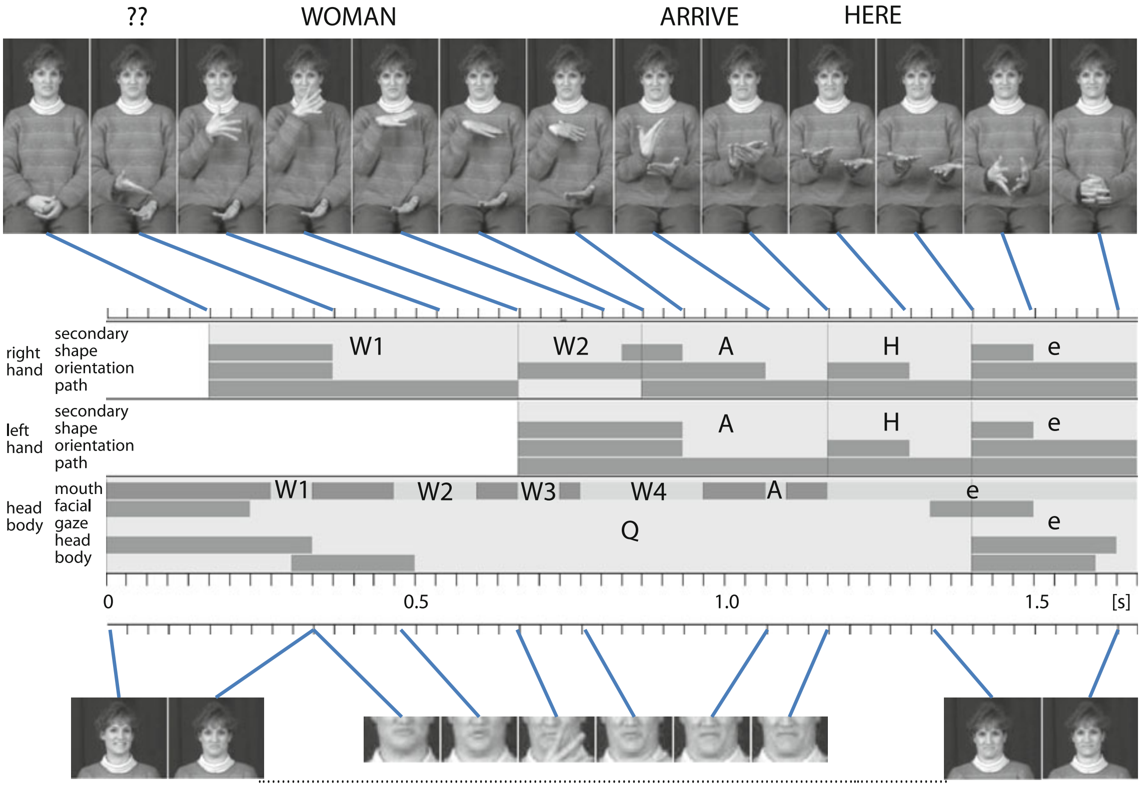

Fig. 4 Score for hand and head–body sub-actions (three main tiers: right hand, left hand, and head–body) including appropriate elementary movement actions for the ASL sentence "WOMAN ARRIVE HERE?" produced by speaker 1 (sentence 109 from our corpus [3]). Movement phases of elementary movement actions are indicated by *dark gray bars*. The *right hand* is the strong hand and *left hand* is the weak hand in the case of this signer

just as the face and body act as bases upon which the strong hand acts.) The strong hand path movement elementary action draws the strong hand down toward the weak hand, resulting in contact with the weak hand. It should be noted that the intermediate level of strong hand manual sub-actions in the sign language modality may be comparable to the syllable sub-action level in the spoken language modality. (See also the concept of syllables in ASL introduced by [32], which is in line with our approach).

An example of a complete American Sign Language sentence, "WOMAN ARRIVE HERE?", along with a temporal specification of all emerging sign actions, manual sub-actions, and elementary movement actions, is shown in Fig. 4; all sign actions, manual sub-actions and elementary movement actions including their target or shape features are labeled in Table 2. In addition to the three (manual) sign actions WOMAN, ARRIVE, and HERE and the oral expression actions (or mouthings) "W1, …, W4, A", a facial expression action occurs which represents the prosodic category of this sentence, i.e. QUESTION: the speaker's facial expression (corners of the mouth down, outer eyebrows up, etc.) is "doubt" during the production of the entire sentence. The duration of each action for the right and left hands, the facial system and the oral system (here labeled as "mouth") as well as for gaze and for the body system is shown in Fig. 4. The movement phases of all elementary movement actions are indicated by dark gray bars. It can be seen from this example that the facial expression action "Q" starts before the first manual sign actions start and ends when the last manual sign actions of this sentence ends—while all oral actions ("W1, …, W4, A") occurring on the mouth tier temporally co-occur with the manual sign actions WOMAN and ARRIVE—or, to be more precise, they temporally co-occur with the strong-hand manual sub-actions ("W1, W2, A") occurring on the right-hand tier.

In the case of both the signed and the spoken language modalities, *facial expression actions* occur. Their purpose can be, for example, to communicate the emotional state of the speaker (see above). *Elementary facial movement actions* are also called *facial action units* (AU's, see [6, 7], and the complete system of facial action units is called *Facial Action Coding System FACS* [9] and [10]). As an example, a facial expression for the emotional state "happiness" is shown in Fig. 5. The arrows in the figure illustrate that this *facial expression action* with the meaning "happiness" results from the activation of at least two elementary facial movement actions, i.e. cheek raising including passive lid compressing (AU6, [6]) and lip corner

Table 2 Listing of target features for each elementary movement action occurring within the ASL sentence "WOMAN ARRIVE HERE?", produced by speaker 1 (sentence_109 from our corpus [3])

Meaning of sign action	Manual sub-action, oral or facial expression action		Elementary movement action	Feature
WOMAN	Right (strong) hand	W1	Shape	Spread (all fingers)
			Orientation	Right
			Path	Up, contact (thumb) at chin
		W2	Shape	Spread (little finger and thumb)
			Orientation	Up
			Path	Down, contact (thumb) at chest
	Oral	W1	Oral	Closure (for "w")
		W2	Oral	Open, rounded (for "o")
		W3	Oral	Bilabial closure (for "m")
		W4	Oral	Open
ARRIVE	Right (strong) hand	A	Shape	Straight, spread thumb
			Orientation	Down
			Path	Down, contact with left hand
	Left (weak) hand	A	Shape	Straight, spread thumb
			Orientation	Down
			Path	Up, contact with right hand
	Oral	A	Oral	Labio-dental closure (for "v")
HERE	Right (strong) hand	H	Shape	Spread
			Orientation	Down
			Path	Right
	Left (weak) hand	H	Shape	Spread
			Orientation	Down
			Path	Left
QUESTION	Oral	Q	Oral	Closure (also labeled as W1)
	Facial		Facial	Mouth corners down; inner brows down and tightened; outer brows up
	Gaze		Gaze	Toward communication partner
	Head-body		Head	Slightly down
			Body	Shoulders up
Ending	Right hand	E	All	Toward rest position
	Left hand	E	All	Toward rest position
	Oral	E	Mouth	Closure
	Facial, gaze, head-body	E	Facial	Toward neural

The abbreviations for the manual sub-actions and for the oral and facial expression actions are also indicated in Fig. 4. *Orientation* means palm orientation. Letters specified as target features for oral expression actions in this table indicate letters of words representing the appropriate sign action

pulling (AU12, [6]). Little is known about the temporal relations between the elementary facial movement actions which make up a superordinated facial expression action (i.e. a complete facial expression like "happiness"). It is assumed that the movement phases of elementary facial movement actions occur more or less synchronously in time and that these elementary movement actions show hold phases for a considerable time interval (Fig. 5). However, some studies indicate that facial dynamics and thus the movement phases of elementary facial movement actions and their timing are different for different facial expressions and that this facial dynamics is a very important factor in the perception of facial expressions [1, 6, 35–37, 39].

In the case of the signed language modality, *oral expression actions* (i.e. mouthings, see [12, 34]) temporally co-occur with manual movement actions (see the sample sentence above). In some cases, oral expression actions can be interpreted as speech-like expressions, and in other cases they function in a different way, i.e. adding linguistic information to the manual sign action. It is important to state that, similarly to facial expression actions, oral

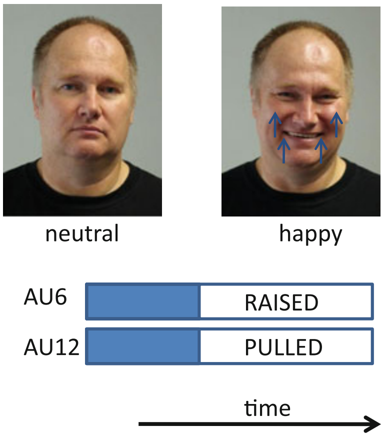

Fig. 5 *Top*: Examples for facial expressions of the emotional states "neutral" and "happy" (from [41]). *Arrows* indicate the activated facial regions for two elementary movement actions: cheek raiser AU6 and lips corner puller AU12 (see text). *Bottom*: Hypothetical time course of movement and target phase for both elementary movement actions

expression actions are composed of labial, jaw, and apical (i.e. tongue tip) elementary movement actions.

In the case of all facial and all oral expressions, the composition of the superordinate meaningful facial or oral expression action is directly based on elementary movement actions, without an intermediate level of syllabic sub-actions as is the case for speech or sign actions. In addition, it is assumed that all elementary movement actions that make up a facial or oral expression action exhibit more or less temporally synchronous movement phases. Thus, in this study, i.e. in the action-based analysis of our ASL corpus, movement phases are determined for facial and oral expression actions without a detailed temporal analysis of the subordinate elementary movement actions.

Thus, the concept of communicative actions made up of specific temporally well-coordinated elementary movement actions can be applied in a similar way to all three *articulatory systems* important in face-to-face communication, i.e. the manual system, the vocal tract (including oral) system, and the facial system. It is our hypothesis that with respect to the language *modality* (i.e. signed or spoken) the "temporal tightness of actions" can vary from system to system. Thus, in the signed language modality, a high level of temporal tightness of actions is assumed for the manual system, and thus the temporal organization of sign actions dominates the temporal organization of co-manual oral and facial actions. Conversely, in the spoken language modality, a high level of temporal tightness of actions is assumed for the vocal tract system, and here the temporal organization of speech actions dominates the temporal organization of co-verbal manual and facial actions. The detailed hypotheses we will test in this study are derived from a preliminary inspection of an American Sign Language (ASL) corpus. The hypotheses are the following: (1) All lexical signs occurring in fluent sign language production can be mapped onto sign actions. (2) Each sign action (i.e. each manual sign action occurring in the signed language modality) is made up of one or more temporally sequent strong-hand sub-action plus up to one temporally co-occurring weak-hand sub-action; and these sub-actions comprise elementary manual movement actions (path, shape, orientation, and secondary movement actions). (3) Sign actions do not show *absolute holds* (i.e. complete hold of the strong hand for more than 100 ms) in short non-emphatic sentences (short statements or questions), while the co-occurring or co-manual mute oral expression actions as well as the co-occurring or co-manual facial expression actions (and in the case of asymmetrical signs, the co-occurring weak-hand sub-actions) show considerable absolute holds (i.e. complete holds of the oral, facial, or weak-hand manual system for more than 100 ms).

It should be noted that the weak hand in asymmetrical signs functions as a spatial base for strong hand movements, and as such it shows holds (called *buoys*; see e.g. [28]). But articulatory timing is always dominated by the strong hand, in asymmetrical signs as well as in single strong-hand signs or in symmetrical signs. Thus, it is sufficient to analyze the strong-hand movements in order to test hypothesis three (stated above). Moreover, it is important to state that hypothesis three only holds for short sentences (as occur in our ASL corpus). The sentences occurring in our ASL corpus are (neutral) statements or short questions. In face-to-face communication, strong hand holds may occur which signal e.g. the end of a sentence ("That's all I want to say now, so please take a turn and answer"); a strong emphasis on a lexical item; or the beginning and end of a contrastive phrase. Thus, sentence-final sign actions are excluded from analysis even in our short sentence corpus. But these more discourse-related features are beyond the scope of this paper.

Methodology

A 201-sentence ASL corpus recorded at the National Center for Sign Languages and Gesture Resources at Boston University [3] which comprises 201 videos of short sign language sentences produced by 3 different signers (2 female "speakers", 32 and 38 years old, 1 male "speaker", 42 years old, all born deaf and native ASL signers) was chosen for visual analysis. One hundred randomly chosen sentences of this corpus were used for the analysis (our main data set). A further forty sentences of the ASL corpus were used for annotation training purposes (our training data set). Frontal views of the speakers were used, and the video frame rate was 30 frames per second.

The term “speaker” is used here for both the signed and the spoken language modalities. In addition to the video data, the lexical sign items were already given for each sentence in the order in which they appear in each sentence; e.g. WOMAN, ARRIVE, HERE for the sample sentence exemplified above. A sign language annotation tool was developed which makes possible the annotation of: (1) the temporal location of strong- and weak-hand manual sub-actions for each sign action; the temporal location of facial and oral expression actions; (2) the temporal location of movement phases for all elementary movement actions occurring in the manual system (right and left hand); (3) one (overall) movement phase for each oral or facial expression action; and (4) movement phases for gaze, head, and other body movement actions. The tool makes possible a frame-by-frame inspection of each video image and the annotation of a time stamp for the beginning and end of each manual sub-action for the right and left hand and for each facial and oral expression action (small vertical lines in Fig. 4) as well as for the movement phase of each elementary movement action (horizontal gray bars for secondary, shape, orientation, path, mouth (i.e. oral), facial, gaze, head, and body movements in Fig. 4). The manual system is divided into the right-hand and left-hand sub-systems, where the right hand is the strong or dominant hand and the left hand is the weak or non-dominant hand for all three speakers in our corpus. The head–body system comprises the oral articulatory system, which is labeled “mouth” in our annotation tool (Fig. 4) and the facial system. The annotation of the beginning and end of the movement phase of manual, oral, and facial elementary actions was done through frame-by-frame inspection of the video images.

For the manual system, path movements are defined as movements of the arm which lead to translation movements of the palm; orientation movements are defined as additional rotational movements of the palm resulting from lower arm rotations and wrist movements; shape movements are defined as movements of the fingers leading to specific hand shapes (e.g. spread fingers, straight or flat, compact or fist, curved or c-shaped, etc.); and secondary movements are defined here as well as in sign language phonology as additional overlaid movements such as finger wiggling or finger circling (e.g. [32]).

In the case of manual actions, the annotation procedure was done in two steps. In step one, a gross time interval was annotated for left- and right-hand manual sub-actions. In step two, timestamps for the beginning and end of each movement phase of each elementary movement action were annotated. Annotation was done independently by two annotators. During an initial training phase, the two annotators exchanged and discussed their annotation results in order to establish a common approach for annotation (annotation of 20 sentences from the training data set). After this training phase, one annotator transcribed the complete main data set (100 sentences). These data were used later for the analyses. Of these 100 sentences, speaker 1 (female, 32 years old) uttered 35 sentences, speaker 2 (female, 38 years old) uttered 29 sentences, and speaker 3 (male, 42 years old) uttered 36 sentences. Mean sentence duration was 2.66 s (between min = 1.37 s and max = 4.87 s). Later, the second annotator analyzed 20 sentences which were a subset of the main data set. This annotation done by the second annotator was performed in order to evaluate the degree of congruence of annotations. The degree of congruence of annotations came out to be acceptable (see Table 3): among these 20 sentences, 65 identical sign actions were annotated by both persons. Moreover, in total, 173 elementary movement actions were annotated consistently by both persons, i.e. 62 path movement, 65 shape movement, 38 orientation movement, and 8 secondary movement elementary actions. All elementary movement actions were detected by both annotators, and no additional elementary movement action was noted by either of the annotators; no elementary movement action remained undetected by either of the annotators. In terms of the time stamp, the annotation of the beginning and end of the movement phases of these elementary movement actions (173 items) was identical between the two annotators in 102 cases (59.0%). In the cases in which there were disparities between the annotators, the difference was a single frame (i.e. 33.3 ms) in 61 cases (35.2%); there was a difference of two or more frames in only 10

Table 3 Number of elementary movement actions in each articulatory system (manual, oral, and facial), annotated by two persons (see text) (1) without any temporal divergence (2) with a temporal divergence of one video frame (temporal interval of 33.3 ms), and (3) with a temporal discrepancy of two or more video frames (temporal interval equal to or greater than 66.6 ms) concerning the timestamp for the beginning and/or end of the movement phase of each annotated elementary movement action

System	Elementary movement actions, annotated in total	No temporal divergence	Temporal divergence = one frame (33.3 ms)	Temporal divergence ≥ two frames (66.6 ms)
Manual	173 (100%)	102 (59.0%)	61 (35.2%)	10 (5.8%)
Oral	68 (100%)	18 (26.5%)	38 (55.9%)	12 (17.6%)
Facial	52 (100%)	12 (23.1%)	32 (61.5%)	8 (15.4%)

cases (5.8%). Sign actions as well as oral expression actions occurring in a sentence-initial or sentence-final position were excluded from our analysis, since we assume that these actions may be produced with a different temporal behavior, e.g. more slowly compared to actions produced in fluent utterance context. In addition, fingerspelling was excluded from our sign action analysis due to the overabundance of hand shape movements in comparison with other (normal) sign actions.

In the case of oral expression actions, the annotation convention was to annotate the sequence of lip movement actions with increasing and decreasing distances between the upper and lower lips, representing vowel- or consonant-like speech sounds. Thus, for the oral articulatory system, the primary annotations were the opening and closing actions of the mouth; however, a further annotation of oral actions was done by inspecting whether lips were wide open or narrow and whether lips were rounded or spread in the case of opening actions; or whether a constriction was formed in a bilabial, labio-dental or apical (i.e. tongue tip) way in the case of closing actions. For each oral expression action, it was sufficient to annotate only one timestamp each for the beginning and end of the movement phase of all elementary movement actions (lips, jaw, tongue). As was done for the manual actions, the two annotators developed a similar way of annotating points in time for the beginning and end of movement phases of oral actions on the basis of 20 sentences of the training data set. Subsequently, one person annotated all 100 sentences of the main data set, while the second person annotated 20 of these 100 sentences in order to evaluate the degree of congruence of the annotations. The degree of congruence was acceptable (see Table 3): among these 20 sentences, 68 oral expression actions were annotated by both persons. All oral expression actions occurring in these 20 sentences were detected by both annotators, and no additional oral expression action was notated by either of the annotators. No temporal difference in the annotation of the beginning and end or the movement phase of oral actions occurred for 18 oral expression actions (26.5%). In the cases in which there were disparities between the annotators, difference was one frame (i.e. 33.3 ms) in 38 cases (55.9%) and two or more frames in only 12 cases (17.6%). All oral expression actions which temporally co-occur with excluded sign actions (sentence initial, sentence final, and fingerspelling) were excluded from our analysis.

In the case of facial expression actions, the annotation convention was to concentrate mainly on the eye region, i.e. on eyebrow and eyelid movements. In addition, expressive changes with respect to the cheek and mouth region were analyzed if they were visually dominant. The annotation of facial expression actions was done using the FACS technique [7, 10]: i.e. by inspecting whether inner or outer eyebrows were raised or lowered; whether upper eyelids were raised (eyes wide open) or eyelids were partly closed; whether cheeks or upper lips were raised; whether the corners of the mouth were pulled up or down; and/or whether lips were stretched or pressed. As for the manual sub-actions and oral expression actions described above, the annotators developed a similar technique of annotating the beginning and end of movement phases of facial actions separately on the basis of all 40 sentences of the training data set. Subsequently, one person annotated all 100 sentences of the main data set, while the second person annotated 40 of these 100 sentences in order to evaluate the degree of congruence between annotators. The degree of congruence was acceptable (see Table 3): among these 40 sentences, 52 facial expression actions were annotated by both persons. All facial expression actions occurring within each sentence were detected by both annotators, and no additional facial expression action was notated by either of the annotators; no facial expression action remained undetected by either of the annotators, and no facial expression action was added by either of the annotators. There was no temporal difference between the annotators in the annotation of the beginning and end or the movement phase of facial actions for 12 facial expression actions (23.1%). In the cases in which there were disparities between the annotators, the temporal difference in annotation of the beginning and/or end of the movement phase of facial expression actions was one frame (i.e. 33.3 ms) 32 cases (61.5%) and two or more frames in 8 cases (15.4%). Sentence-initial facial actions were included in our analysis if these facial expression actions were held for longer than the initial manual sign action.

Results

Suitability of the Theory of Communicative Actions for ASL

We were able to apply the theory of communicative actions introduced in chapter 2 of this paper in a straightforward way to our corpus of ASL data. One major result of this study is that we were able to account for all movements of all articulators in all tiers (manual, oral, facial) as movement phases of elementary movement actions and assign each elementary movement action to a superordinate manual sign, oral expression, or facial expression action. For the manual system, we were able to map all lexical signs occurring in the sentences produced by the speakers onto sign actions where each sign action is made up of one or more strong-hand manual sub-actions and up to one weak-hand sub-action and where these sub-actions comprise elementary movement actions as described above.

Table 4 Number of strong-hand manual sub-actions, oral expression actions, and facial expression actions produced by all three speakers in our main data set (100 sentences)

System	Speaker 1 (35 sentences)			Speaker 2 (29 sentences)			Speaker 3 (36 sentences)			All speakers (100 sentences)		
	Total	No holds	Holds	Total	No holds	Holds	Total	No holds	Holds	Total	No holds	Holds
Manual	176	176 (100%)	0 (0%)	180	180 (100%)	0 (0%)	164	164 (100%)	0 (0%)	520	520 (100%)	0 (100%)
Oral	149	107 (72%)	42 (28%)	148	113 (76%)	35 (24%)	183	155 (85%)	28 (15%)	480	375 (78%)	105 (22%)
Facial	60	0 (0%)	60 (100%)	53	0 (0%)	53 (100%)	64	0 (0%)	64 (100%)	177	0 (0%)	177 (100%)

A distinction is made with respect to whether an action or sub-action exhibits an absolute hold or not

While manual movement actions are assigned to sign actions via the intermediate state of strong- and weak-hand manual sub-actions, oral and facial elementary movement actions can be assigned directly to a superordinate oral or facial expression action.

For the manual articulatory system, we were able to identify 365 sign actions comprising 520 strong-hand manual sub-actions (many sign actions are composed of two temporally sequential strong-hand manual sub-actions) in our main data set (i.e. 100 ASL sentences). All temporal intervals representing the sequence of strong-hand manual sub-actions occur with no gaps in all sentences (see the sample sentence given in Fig. 4). Each sentence is finalized by a manual "end" or "to-rest" movement action (not included in our analysis). In the case of the weak hand, "to-rest" movement actions also occur during the course of a sentence if the weak hand is not used for sign production during a longer part of that sentence. Moreover, we were able to annotate the beginning and end of all movement phases of all elementary manual movement actions occurring for the right and left hands, and we were able to allocate all elementary movement actions to the previously annotated strong- or weak-hand manual sub-actions and thus to superordinate sign actions.

Concerning the oral system, we were able to annotate 480 oral expression actions in total from the main data set (Table 4), and we were able to assign each of these oral expression actions to manual sign actions on the basis of temporal co-occurrence. Concerning the facial system, we were able to annotate 177 facial expression actions in total from the main data set (Table 4; one to three facial expression actions per sentence).

Different Types of Sign Actions in ASL

In the case of the manual system, five basic types of sign actions, comprising different numbers of strong-hand manual sub-actions and different numbers of elementary movement actions within a strong-hand manual sub-action, were identified on the basis of our corpus analysis. The five types of sign actions are single path, serial path, shape-secondary, path-shape synchronous, and shape-tap. In the case of a *single path type* sign action (e.g. in sign actions such as I, YOU), only one strong-hand manual sub-action occurs. The movement phase of the path movement elementary action covers the duration of the entire sign action, while the movement phase of the shape movement elementary action is completed in the first half of the sign action (Fig. 6a). An orientation movement elementary action is optional and can occur in temporal synchrony with the shape movement elementary action. In the case of a *serial path type* sign action (e.g. in sign actions such as WORK, SEE), two subsequent strong-hand manual sub-actions occur (Fig. 6b). The first or initial strong-hand manual sub-action is organized like the strong-hand manual sub-action that occurs in the single path type. This initial strong-hand manual sub-action basically moves the strong hand to its sign-initial position and adjusts the sign-specific hand shape; then the path or orientation movement elementary action of the second strong-hand manual sub-action is executed. This elementary movement action of the second strong-hand manual sub-action covers the sign-specific movement. This type of sign action mainly represents signs of the "HMH" type or "HM" type in the movement hold phonology paradigm [26, 40], and here the first strong-hand sub-action of the sign action can be interpreted as an epenthetic movement which sets up the hand shape and hand position of the first "H" (hold) of the sign. The *shape-secondary type* sign action (e.g. in sign actions such as GERMANY, PLEASANT) also comprises two subsequent strong-hand manual sub-actions (Fig. 6c). The initial strong-hand manual sub-action is comparable to that of the serial path type, i.e. an epenthetic movement action which sets up the initial hand shape and hand position for the sign. The subsequent strong-hand manual sub-action mainly comprises a secondary movement (elementary) action (e.g. finger wiggling, see [32]). In many cases, no path movement occurs during the second strong-hand manual sub-action (e.g. GERMANY). Only in few cases did a path movement elementary action occur in addition to the secondary movement (elementary) action during this second strong-hand sub-action (e.g. PLEASANT). The *path-shape synchronous type* sign action (e.g. in sign actions such as LEAVE, TAKE) also comprises two

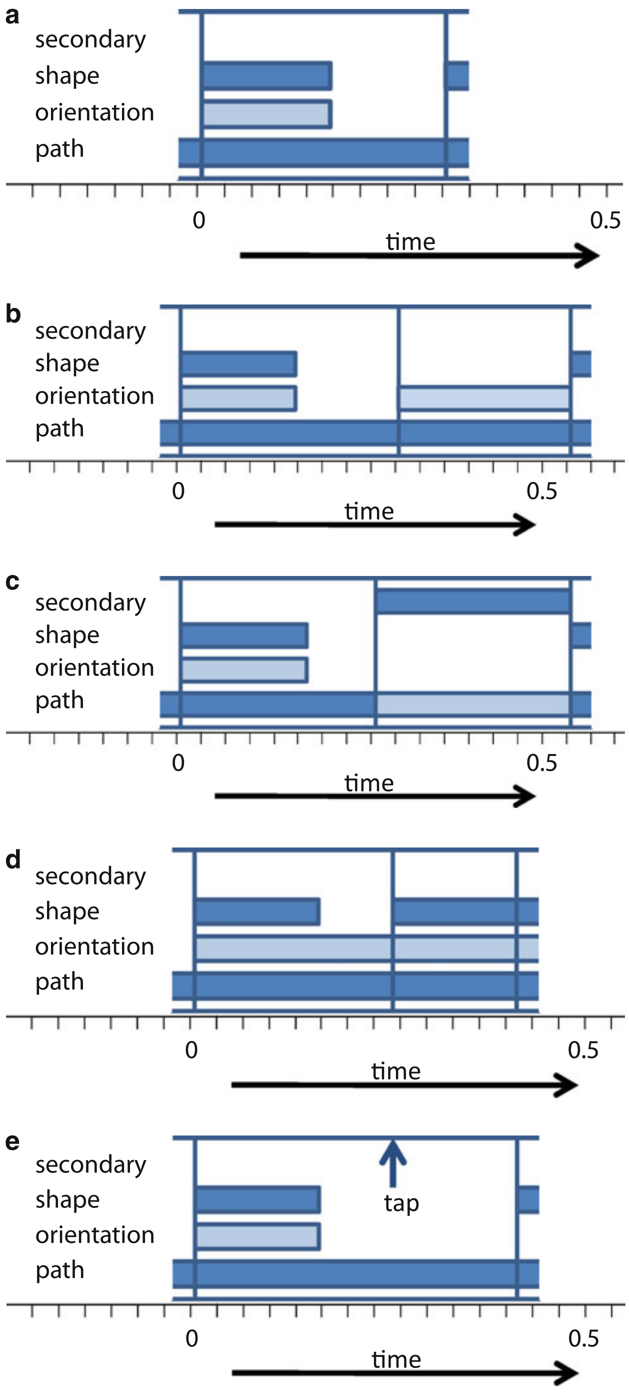

Fig. 6 Temporal organization of elementary movement actions for the strong hand path, orientation, shape, and secondary movement tier for five different types of sign actions, i.e. **a** single path, **b** serial path, **c** shape-secondary, **d** path-shape synchronous, and **e** shape-tap type. *Solid blue bars* indicate the movement phases of elementary movement actions. *Solid light blue bars* indicate optional elementary movement actions

sequential strong-hand manual sub-actions. The initial strong-hand manual sub-action is organized as described for the serial path type, i.e. an epenthetic movement action that sets up the initial hand shape and hand position for the sign. Within the second strong-hand manual sub-action, a temporally synchronous movement phase occurs for a shape movement and for a path movement elementary action which forms the sign-specific change in hand shape as well as the sign-specific hand movement which covers the entire duration of the second strong-hand manual sub-action (Fig. 6d). The *shape-tap type* sign action (e.g. in sign actions such as NEW) mainly occurs as a reduced form of a serial path type sign action. The temporal organization of the strong-hand manual sub-actions of this sign action (Fig. 6e) is mainly like that of a single path type, but here the path movement elementary action includes a tap, i.e. a cursory touch of a body part (e.g. the cheeks, chin, or weak hand).

The analysis of the main data set (100 sentences) and three speakers revealed that single path types are most frequent, followed by serial path, shape-secondary and path-shape synchronous types (Fig. 7). The shape-tap type occurred in our corpus only three times in the 365 sign action items we analyzed.

A further quantitative analysis was done for 15 tokens of each of the four most frequent types of sign actions (i.e. single path, serial path, shape-secondary and shape-path synchronous, covering 99% of all sign actions in our main data set; see Fig. 7). Five tokens were chosen randomly per speaker and per type of sign. The quantitative analysis was performed for (1) total duration of the sign action; (2) duration of the movement phase of the shape movement elementary action (i.e. shape movement duration); and (3) movement peak velocity (i.e. peak velocity of the path movement elementary action within the second or sign-specific strong-hand manual sub-action in the case of serial path, path shape synchronous, and shape-secondary types; or within the first or only strong-hand manual sub-action in the case of the single path type). The movement velocity was calculated from the trajectory of strong hand positions over time. These trajectories were derived by an automatic

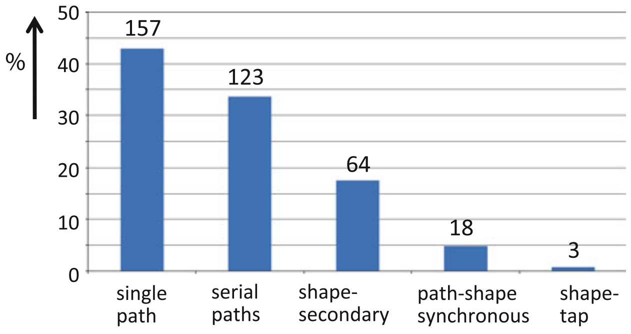

Fig. 7 Frequency of occurrence of different types of sign actions based on the analysis of the main data set (100 ASL sentences). In total, 365 sign actions (=100%) were analyzed

tracking algorithm applied to the video data and adjusted for the palm of the strong hand [8]. The raw velocity data were Gaussian-smoothed using a time window of 8 samples (i.e. 266 ms). The analysis revealed that the duration of the movement phases of all shape movement elementary actions is constant at approximately 150 ms—with the exception of the path-shape synchronous movement type, where it is approximately 200 ms (Fig. 8). The total duration of a sign action is approximately 250 ms for the single path type; the total duration of a sign action is approximately 450–550 ms in the case of the serial path and shape-secondary type, i.e. approximately double the length, presumably due to the fact that these types comprise two subsequent strong-hand manual sub-actions. Thus, it can be concluded that path movement elementary actions—which determine the length of most strong-hand manual sub-actions—are of a constant length, i.e. approximately 250 ms. The total duration of a sign action of the path-shape synchronous type is slightly shorter than that of the serial path or shape-secondary type, presumably because the second strong-hand manual sub-action in the case of this type is a temporally synchronous shape *and* path movement elementary action. This second strong-hand manual sub-action is slightly shorter than the second strong-hand sub-action in a serial path or shape-secondary type of sign action (i.e. approximately 200 ms, see Fig. 8), presumably because a compromise must be struck here between the duration of the shape and the path movement action. These results indicate a specific duration for groups of elementary movement actions which make up a sign action across different speakers and different sign meanings. Moreover, movement peak velocities estimated for the sign-specific path movement indicate a constant peak velocity of approximately 1 m per second. In the case of the shape-secondary type, the movement peak velocity is approximately zero, since no elementary path movement actions occur; only stationary secondary movements (elementary) actions occurred in the tokens analyzed here.

Movement and Hold Phases in Manual, Oral, and Facial Actions of ASL

A further major result of this study is the finding that sign actions performed with the strong hand do not show absolute holds (i.e. no complete hold of the strong hand for more than 100 ms; weak-hand holds were not analyzed here), while the co-occurring or co-manual mute oral expression actions as well as the co-occurring or co-manual facial expression actions show considerable absolute holds (i.e. complete hold of the oral or facial system for more than 100 ms). Within the main data set, strong-hand manual actions show no holds (0%, i.e. 0 of 520 actions); all facial expression actions show holds (100%, i.e. 177 of 177 actions); and oral expression actions are in between these two extremes with around 22% holds (i.e. 105 of 480 actions) over all three speakers (Table 4). The number of strong-hand manual sub-actions, facial expression actions, and oral expression actions per sentence is given as a histogram in Fig. 9. It can be seen that—compared with the total number of strong-hand manual, facial expression, and oral expression actions (i.e. 520, 177, 480)—the number of facial expression actions per sentence (around 2) is considerably lower than the number of oral expression and manual strong-hand actions per sentence (around 5 for

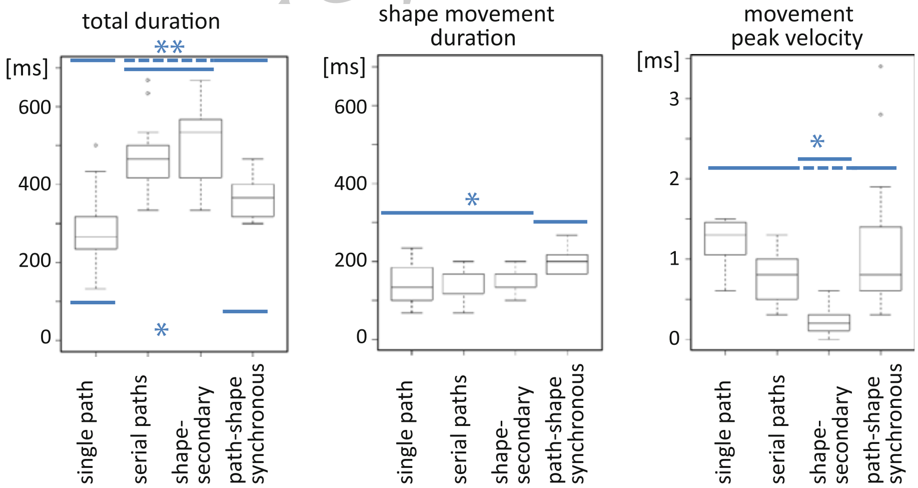

Fig. 8 Box plots of total duration of a sign action, of duration of the movement phase of the shape movement elementary action and movement peak velocity values for the four most frequent types of sign actions. If no median is plotted, the median in all cases coincides with the upper end of the quartile range. Movement peak velocity is estimated by using hand position tracking data [8]. Significant differences in values for different types of sign actions are indicated by horizontal lines (* $p < 0.05$; ** $p < 0.\ 01$ from post hoc Tukey-HSD-test following one factor multivariate ANOVA)

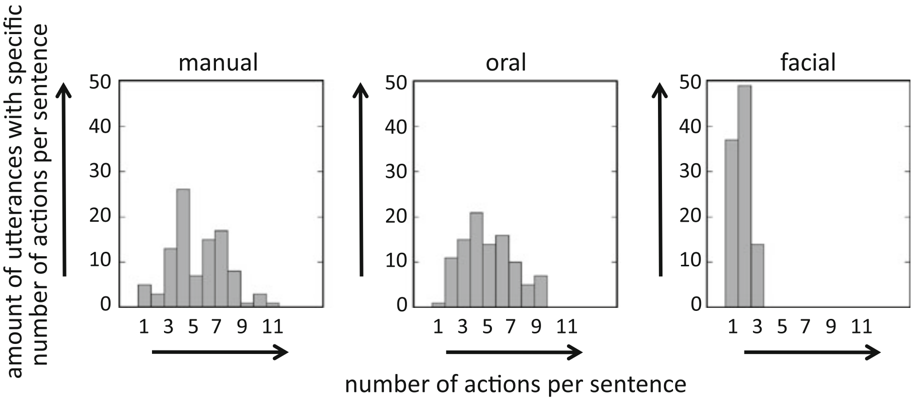

Fig. 9 Histogram of the distribution of number of sentences exhibiting specific numbers of strong-hand manual sub-actions, oral expression actions, and facial expression actions (in total for all three speakers)

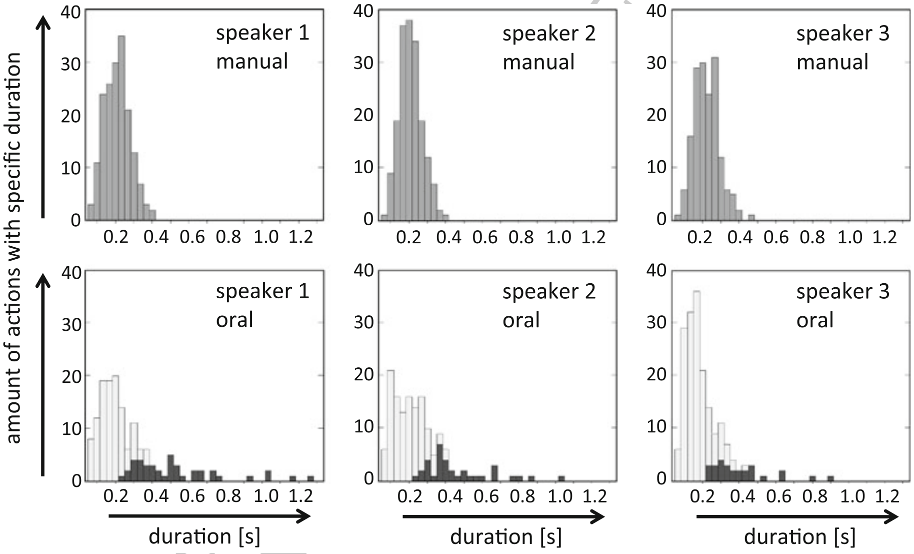

Fig. 10 Histogram of the distribution of number of strong-hand manual sub-actions (*top row*, *gray bars*) and oral expression actions (*bottom row*) with respect to the duration of the actions and sub-actions, respectively. Oral expression actions are divided with respect to actions exhibiting no holds (*light gray bars*) and those exhibiting holds (*dark gray bars*). In the case of strong-hand manual sub-actions, no holds occur (see also Table 4). The time difference of duration per bar is 33.3 ms

each). The finding that facial expression actions are related to the phrase or sentence, while oral expression actions are directly related to sign actions (i.e. strong-hand manual sub-actions), will be discussed below.

An important observation arising from the annotation procedure was that there is little variation in total duration for strong-hand manual sub-actions, while there is a great deal of variance in the total duration of oral expression actions. In order to test the hypothesis that the total duration of strong-hand manual sub-actions is relatively invariant while that of oral expression actions is variable, we created histograms which plot the number of strong-hand manual sub-actions and oral expression actions against the total duration of these actions (Fig. 10). It can be seen that the mean duration of strong-hand manual sub-actions is approximately 250 ms, with a total range of approximately ± 150 ms (i.e. maximum of 400 ms) for all speakers. However, in the case of oral expression actions, some actions are dramatically longer—up to 1300 ms for speaker 1, up to 1100 ms for speaker 2 and up to 900 ms for speaker 3. It will be posited below that these oral expression actions with extremely long durations especially exhibit absolute holds (indicated by the dark gray bars in Fig. 10). In addition, it can be seen from Fig. 10 that there are more oral expression actions than strong-hand manual sub-actions (light gray bars) with a total duration of less than 200 ms.

The high number of oral expression actions with a total duration below 200 ms found in our analysis can be explained by the fact that the movement phase of elementary movement actions is relatively short in the case of oral expression actions in comparison with the movement phase, e.g., of path movement elementary actions within strong-hand manual sub-actions. This leads to a shorter duration for oral expression actions in comparison with strong-hand manual sub-actions, if both actions are no-hold actions (i.e. they exhibit either short hold phases or no hold phases within their subordinate elementary movement actions). The finding of a high variance in the total duration of oral expression actions in contrast to low variance in total duration of strong-hand manual sub-actions can be explained by the fact that strong-hand manual sub-actions are the dominant timing unit in ASL production and thus are of approximately constant duration (see upper row of

Journal : Large 12559 | Dispatch : 23-9-2010 | Pages : 17
Article No. : 9071 | □ LE | □ TYPESET
MS Code : | ☑ CP | ☑ DISK

Fig. 10 and general discussion below), while the timing of oral expression actions must be related to that of (superordinate) sign actions. We were able to distinguish three different kinds of relations between oral expression actions and sign actions: (1) only one single oral expression action (e.g. a C-like mouth closing or a V-like mouth opening oral expression action) is related to one sign action; (2) a subsequent unit of two oral expression actions (e.g. a CV-like mouth closing-opening sequence) is related to a sign action; and (3) no oral expression action is related to a sign action. An example of a mouth closing oral expression action is given in Fig. 4 and Table 2; see the oral expression action which is related to the sign action ARRIVE. Two examples of a sequence of mouth closing-opening oral expression actions is given in Fig. 4 and Table 2; see the oral expression actions which are related to the sign action WOMAN. Here, each sequence of two oral expression actions is related to one of two subsequent strong-hand manual sub-actions W1 and W2. Since sign actions can be composed of one, two, or even more temporally sequential strong-hand manual sub-actions, at least three cases of timing relations between oral expression actions and strong-hand manual sub-actions can be differentiated: (1) a sequence of two oral expression actions is timed with respect to one strong-hand sub-action. In this case, it is most likely that the first oral expression action is very short and that its elementary movement sub-actions exhibit no hold phases. (2) One oral expression action is timed with respect to one strong-hand manual sub-action, and the next oral expression action is timed with respect to the next strong-hand manual sub-action. (3) One oral expression action is temporally related to a (first) strong-hand manual sub-action, but one or more subsequent strong-hand manual sub-actions follow in time without being related to the next oral expression action. In this case, it is most likely that the oral expression action is relatively long and that it exhibits hold phases (see lower row in Fig. 10). In Table 5, all oral expression actions exhibiting holds, as well as their relation to strong-hand manual sub-actions, are shown. In total, we found 75 oral expression actions that exhibit holds and which are related to one strong-hand manual sub-action with no subsequent strong-hand manual sub-actions occurring during the hold phase of the same oral expression action ($n = 1$), while there were 30 oral expression actions that were related to one strong-hand manual sub-action with two of more subsequent strong-hand manual sub-actions occurring during the hold phase of the same oral expression action ($n = 2, 3, 4$). These 30 oral expression actions comprising cases $n = 2, 3$, and 4 (see Table 5; 19 oral expression actions for speaker 1, 7 oral expression actions for speaker 2, and 4 oral expression actions for speaker 3) exhibit long durations from approximately 500 ms up to approximately 1400 ms (see Fig. 10). A closer inspection of the data reveals that the duration of oral expression actions is approximately 500–750 ms for $n = 2$, 750–1000 ms for $n = 3$, and more than 1000 ms for $n = 4$ and higher. This finding is in line with the finding that the duration of strong-hand manual sub-actions is approximately 250 ms.

Table 5 Number of oral expression actions which exhibit holds (105 in total), sorted with respect to the number of related strong-hand manual sub-actions ($n = 1, 2, 3, 4$) over which these oral expression actions are held

	Speaker 1		Speaker 2		Speaker 3		Total	
	Oral	Manual	Oral	Manual	Oral	Manual	Oral	Manual
$n = 1$	23	0	28	0	24	0	75	0
$n = 2$	14	14	6	6	2	2	22	22
$n = 3$	3	6	1	2	2	4	6	12
$n = 4$	2	6	0	0	0	0	2	6
Total	42	26	35	8	28	6	105	40

In addition, the number of strong-hand manual sub-actions which are not associated with an oral expression action are given (second column for each speaker). In total, the number of strong-hand manual sub-actions (520) exceeds the number of oral expression actions (480 including hold and non-hold oral expression actions) by 40

Furthermore, the between-individual comparison of total numbers of strong-hand manual sub-actions and oral expression actions indicates that speakers 1 and 2 use oral expression actions less often than speaker 3 (speaker 1: 176 manual to 149 oral actions; speaker 2: 180 manual to 148 oral actions; speaker 3: 164 manual to 183 oral actions; see Table 4). This finding is also reflected in the between-individual differences concerning the distribution of the duration of oral expression actions for these speakers. Here, speaker 3 exhibits a high number of oral expression actions which last less than 250 ms, while speakers 1 and 2 exhibit more oral expression actions with long durations including holds.

Another important observation arising from the annotation procedure was that—in contrast to (manual) sign actions—oral expression actions exhibit *noticeable absolute hold phases*, i.e. perceived holds for the complete oral articulatory system (lips, jaw, and tip of the tongue if visible), while sign actions do not exhibit this kind of noticeable absolute hold phase for the strong hand. The procedure for detecting noticeable absolute holds was as follows: two persons viewed the videos of all 100 sentences in normal play mode (no slow motion or single-frame inspection was allowed in this case). These two viewers were asked to identify oral absolute holds. All holds identified by both viewers were labeled as noticeable absolute holds; they are indicated by dark gray bars in Fig. 10. A subsequent single-frame inspection of these noticeable absolute holds indicated that these oral expression actions exhibit an absolute hold phase equal to or longer than 100 ms (i.e. three video frames without any visible change for the appropriate articulatory system). From Fig. 10, we can conclude that oral expression actions exhibiting noticeable absolute holds are longer than 200 ms in total duration (i.e. duration of movement *and* hold phases), and that these oral expression actions can last up to 1300 ms. It can be hypothesized that these holds occur due to the fact that manual sign actions are ongoing during oral expression actions and thus that the beginning of the execution of the subsequent oral expression action must be delayed until the manual articulatory system is ready to execute the next sign action.

A further perception experiment performed by the same two viewers was performed. The task was to detect noticeable absolute hold phases in the strong-hand system. In this task, no absolute holds were found if signs in sentence-initial and sentence-final position and fingerspelling were excluded. Even if these sign actions were included, our analysis of the main data set would exhibit only 9 sign actions with absolute holds in sentence-final position and only 3 sign actions with absolute holds in sentence-initial positions. All 3 of the sentence-initial holds were instances of fingerspelling, and a fourth (i.e. only one) instance of fingerspelling exhibiting a noticeable hold occurred in a non-initial and non-final sentence position. And even when fingerspelling is included, only 0.3% of all sign actions exhibit holds if sentence-initial signs are excluded; this figure rises to 1.1% of all sign actions if sentence-initial signs are included.

The visual inspection of facial expression actions as done by an annotator using our annotation tool for detecting facial holds was performed in the same way as described above for the oral system. Our results, listed in Table 4, clearly indicate that all facial expression actions exhibit noticeable hold phases. The total durations of all facial expression actions are displayed in Fig. 11, and they indicate that these actions—like the oral expression actions which exhibit hold phases—show a rather long total duration and a high level of variation in duration. The total duration of these facial expression actions is between approximately 400 and 2800 ms for speakers 2 and 3 and between approximately 300 and 2500 ms for speaker 1. Thus, the histogram of durations of facial expression actions (Fig. 11) is comparable to the histogram of oral expression actions exhibiting hold phases (the dark gray bars in Fig. 10).

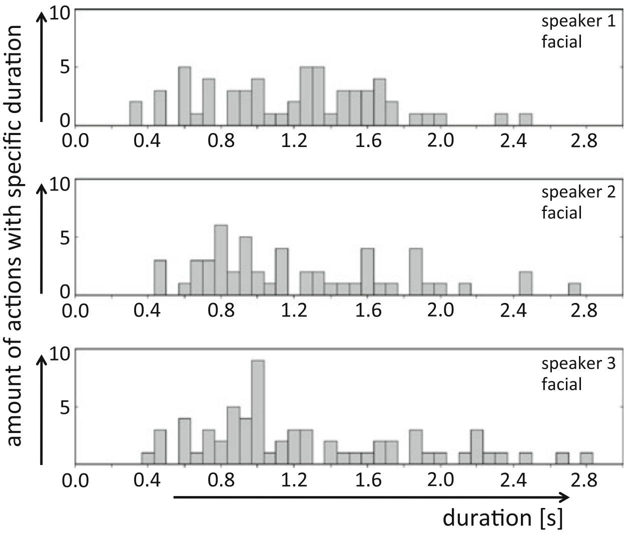

Fig. 11 Histogram of the distribution of number of facial expression actions with respect to duration of the actions and sub-actions, respectively. In this case, all actions exhibit hold phases (see also Table 4). The time difference of duration per bar is 66.6 ms

Discussion

Since most description systems for sign languages are symbolic (for ASL see e.g. the Stokoe notation, cf. [38, 40, 42]), one of the primary goals of this study was to establish an annotation system which specifically takes into account the *quantitative temporal features* of signed language production and which is compatible with the cognitive and sensorimotor concept of action theory as has already been introduced for the spoken language modality [23]. This goal has been reached, and moreover we were able to verify our hypothesis that sign actions are composed of strong- and weak-hand manual sub-actions and that these sub-actions are composed of elementary movement actions. Thus, elementary movement actions can be assumed to be the "atoms" of sign language production and perception in a manner parallel to the way vocal tract actions are assumed to be the "atoms" of spoken language production and perception [4, 25, 5, 14, 23, 25]. Five major types of sign actions were identified on the basis of our corpus analysis and described in detail. These types are comparable to various types of signs introduced in movement and hold phonology (e.g. M-, HM-, and HMH-type signs; see [26, 32]). As noted above, we did not focus on weak-hand sub-actions in our analysis of holds, although they were annotated in our study as well; however, we found that even weak-hand sub-actions are composed of elementary movement actions. Furthermore, it is well known that signs are organized as either single strong-hand signs, symmetrical signs, or asymmetrical signs. In the case of the hold analysis as done here, it was sufficient to analyze strong-hand manual sub-actions, since in the case of symmetrical sub-actions, both hands act in a comparable way; and in the case of asymmetrical signs, the strong hand is the main acting hand, while the weak hand serves as a spatial basis for strong hand movements. Thus, the analysis of strong-hand holds is sufficient in order to identify absolute manual holds, since a strong-hand hold is a precondition for an absolute manual hold.

Despite the fact that we did not find absolute holds for the strong hand in fluent sentence production of ASL, our findings are still compatible with the paradigm of movement and hold phonology (e.g. [26, 40]). "No absolute hold" means that the strong hand is in fact moving with respect to at least *one* of all temporally co-occurring elementary movement actions, i.e. path, shape, orientation, or secondary. Thus, hand shapes are mainly constant during the second half of shape movement elementary actions, while path and/or orientation movements are still ongoing (see discussion of different types of sign actions and see Fig. 6). Even at the end of a path movement elementary action, which is the beginning of a further path or orientation movement elementary action of the following strong-hand manual sub-action, there is a change in movement direction which is often accompanied by a zero crossing of movement velocity. Thus, even in the case of the absence of explicit holds—e.g. for a sequence of path movement elementary actions exhibiting no target (or hold) phase—humans perceive the "hold" of the sign in the terminology of movement and hold phonology, since humans are capable of extracting a goal or target during the movement phase of a goal- or target-directed elementary movement action as a cognitive entity [23].

One further finding of our study can be called the *temporal form principle* of sign actions. Our quantitative analysis of the duration of movement phases of hand shape movement elementary actions and path movement elementary actions indicated a constant duration of approximately 250 ms for path movements. This is thus also the duration of strong-hand manual sub-actions for most action types. We also found a constant duration of approximately 150 ms for the movement phase of their shape movement elementary actions. Only in the case of path-shape synchronous action types (i.e. approximately 4% of all sign actions) was the duration different—approximately 200 ms for the (important) second strong-hand manual sub-action. This gives sign actions a specific temporal structure, and this temporal structure may facilitate sign perception, e.g. the perceptual segmentation of continuously ongoing hand and arm movements into sub-action and elementary movement action units out of which the distinctive features of sign actions is extracted. But because the data pool of 60 different signs used for this quantitative analysis is rather small, these conclusions are preliminary. In further studies, this hypothesis of a temporal form principle should be tested on the basis of a larger data corpus.

One major hypothesis supported by this study is that holds mainly occur in non-dominant articulatory systems in each language modality. In this study, we were able to demonstrate that in the signed language modality, no absolute holds occur for the strong hand in the manual articulatory system, while considerable holds occur in the oral system (22% of all sign actions) and facial system (100% of all sign actions). This lack of absolute holds in manual actions (again, if sentence-initial and sentence-final signs and fingerspelling are excluded; with absolute holds in only 1.1% of all non-sentence-final signs if fingerspelling is included) supports the assumption that because of the dominance of the manual articulatory system in the case of the signed language modality, articulators of the dominant system are in continuous motion in order to perform actions and transfer information quickly.

Note that one of the central hypotheses of action theory is that target shapes (i.e. "holds") of movement actions can be identified from the *movement* phase of elementary movement actions. Thus, no hold phases are needed, at least in the case of path and orientation movement elementary actions. This can be interpreted as a temporal optimization of the transfer of linguistic information in face-to-face communication. In the signed language modality, the manual articulatory system dominates the temporal pattern of communicative actions, so holds mainly occur in the non-dominant co-manual articulatory systems (i.e. the oral and facial systems). The aim of our next study is to test this hypothesis for the spoken language modality. Here, it can be assumed that the elementary vocal tract movement actions dominate the temporal pattern of communicative actions, while the non-dominant co-verbal manual and facial actions may exhibit noticeable holds (a well-known type of co-verbal manual holds are "post stroke holds" [17]). Future studies could compare the findings based on the present corpus with discourse data in order to gain further insight into the opportunities such a quantitative approach can offer regarding the correlation of the temporal dimensions of holds, cognitive pressures, and communicative strategies (cf. [11, 27, 31]).

Acknowledgments This work was supported in part by the German Research Council DFG grant Kr 1439/13-1 and grant Kr 1439/15-1. We thank the two anonymous reviewers for their helpful comments on earlier versions of this paper, and we thank Cigdem Capaat for labeling the ASL corpus.

References

1. Ambadar Z, Schooler J, Cohn JF. Deciphering the enigmatic face: the importance of facial dynamics to interpreting subtle facial expressions. Psychol Sci. 2005;16:403–10.
2. Bauer D, Kannampuzha J, Kröger BJ. Articulatory speech re-synthesis: profiting from natural acoustic speech data. In: Esposito A, Vich R, editors. Cross-modal analysis of speech, gestures, gaze and facial expressions, LNAI 5641. Berlin: Springer; 2009. p. 344–55.

3. Boston-200-Sentences-ASL-Corpus of the National Center for Sign Languages and Gesture Resources at Boston University. 2000. http://www.bu.edu/asllrp/cslgr/.
4. Browman C, Goldstein L. Articulatory gestures as phonological units. Phonology. 1989;6:201–51.
5. Browman C, Goldstein L. Articulatory phonology: an overview. Phonetica. 1992;49:155–80.
6. Cohn JF. Foundations of human computing: facial expression and emotion. In: Huang TS, Nijholt A, Pantic M, Pentland A, editors. Artifical intelligence for human computing (LNAI 4451). Berlin: Springer; 2007. p. 1–16.
7. Cohn JF, Ambadar Z, Ekman P. Observer-based measurement of facial expression with the facial action coding system. In: Coan JA, Allen JJB, editors. Handbook of emotion elicitation and assessment. New York: Oxford University Press; 2007. p. 203–21.
8. Dreuw P, Rybach D, Deselaers T, Zahedi M, Ney H. Speech Recognition Techniques for a Sign Language Recognition System. 2007. Proceedings of Interspeech 2007 (Antwerp, Belgium). pp. 2513–2516.
9. Ekman P, Friesen WV. Measuring facial movement. Environ Psychol Nonverbal Behavior. 1976;1:56–75.
10. Ekman P, Friesen WV. Facial action coding system. Palo Alto, CA: Consulting Psychologists Press; 1978.
11. Emmorey K. Language, cognition, and the brain: insights from sign language research. Lawrence Erlbaum Associates; 2002.
12. Fontana S. Mouth actions as gesture in sign language. Gesture. 2008;8:104–23.
13. Goldin-Meadow S. Hearing gesture. Cambridge, London: Belknap & Harvard University Press; 2003.
14. Goldstein L, Byrd D, Saltzman E. The role of vocal tract action units in understanding the evolution of phonology. In: Arbib MA, editor. Action to language via the mirror neuron system. Cambridge: Cambridge University Press; 2006. p. 215–49.
15. Goldstein L, Pouplier M, Chen L, Saltzman L, Byrd D. Dynamic action units slip in speech production errors. Cognition. 2007;103:386–412.
16. Kendon A. Language and gesture: unity or duality? In: McNeill D, editor. Language and gesture. Cambridge: Cambridge University Press; 2000. p. 47–63.
17. Kendon A. Gesture: visible action as utterance. New York: Cambridge University Press; 2004.
18. Klima E, Bellugi U. The signs of language. Cambridge, MA: Harvard University Press; 1979.
19. Kopp S, Wachsmuth I. Synthesizing multimodal utterances for conversational agents. J Comput Animat Virtual Worlds. 2004;15:39–51.
20. Kröger BJ, Birkholz P. A gesture-based concept for speech movement control in articulatory speech synthesis. In: Esposito A, Faundez-Zanuy M, Keller E, Marinaro M, editors. Verbal and nonverbal communication behaviours, LNAI 4775. Berlin: Springer; 2007. p. 174–89.
21. Kröger BJ, Birkholz P. Articulatory Synthesis of Speech and Singing: State of the Art and Suggestions for Future Research. In: Esposito A, Hussain A, Marinaro M, editors. Multimodal signals: cognitive and algorithmic issues. LNAI 5398. Berlin: Springer; 2009. p. 306–19.
22. Kröger BJ, Kannampuzha J, Neuschaefer-Rube C. Towards a neurocomputational model of speech production and perception. Speech Commun. 2009;51:793–809.
23. Kröger BJ, Kopp S, Lowit A. A model for production, perception, and acquisition of actions in face-to-face communication. Cogn Process. 2010;11:187–205.
24. Lausberg H, Sloetjes H. Coding gestural behavior with the NEUROGES-ELAN system. Behav Res Meth. 2009;41(3):841–9.
25. Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition. 1985;21:1–36.
26. Liddell SK, Johnson RE. American sign language: the phonological base. Sign Lang Stud. 1989;64:195–277.
27. Liddell SK, Metzger M. Gesture in sign language discourse. J Pragmat. 1998;30:657–97.
28. Liddell SK. Grammar, gesture and meaning in American sign language. New York: Cambridge University Press; 2003.
29. McNeill D. Hand and mind: what gestures reveal about thought. Chicago: University of Chicago Press; 1992.
30. McNeill D. Gesture and thought. Chicago: University of Chicago Press; 2005.
31. McNeill D, Quek F, McCullough K-E, Duncan SD, Furuyama N, Bryll R, Ansari R. Catchments, prosody and discourse. Gesture. 2001;1(1):9–33.
32. Perlmutter DM. Sonority and syllable structure in American sign language. Linguist Inquiry. 1992;23:407–42.
33. Saltzman E, Byrd D. Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. Hum Mov Sci. 2000;19: 499–526.
34. Sandler W. Symbolic symbolization by hand and mouth in sign language. Semiotica. 2009;174:241–75.
35. Schmidt KL, Ambadar Z, Cohn JF, Reed LI. Movement differences between deliberate and spontaneous facial expressions: zygomaticus major action in smiling. J Nonverbal Behav. 2006;30:37–52.
36. Schmidt KL, Bhattacharya S, Denlinger R. Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises. J Nonverbal Behav. 2009;33:35–45.
37. Schmidt KL, Cohn JF, Tian Y. Signal characteristics of spontaneous facial expressions: automatic movement in solitary and social smiles. Biol Psychol. 2003;65:49–66.
38. Stokoe WC (1960) Sign language structure: An outline of the visual communication systems of the American Deaf, Studies in Linguistics Occasional Paper 8, University of Buffalo.
39. Tian YL, Kanade T, Cohn JF. Facial expression analysis. In: Li SZ, Jain AK, editors. Handbook of face recognition. New York: Springer; 2005. p. 247–75.
40. Valli C, Lucas C. Linguistics of American sign language. An Introduction. Washington: Gallaudet University Press; 2000.
41. Vanger P, Hoenlinger R, Haken H. Computer aided generation of prototypical facial expressions of emotion. Methods of Psychological Research Online. 1998. Vol. 3, No. 1. http://www.dgps.de/fachgruppen/methoden/mpr-online.
42. Wilcox S, Morford JP. Empirical methods in signed language research. In: Gonzalez-Marquez M, Mittelberg I, Coulson S, Spivey MJ, editors. Methods in cognitive linguistics. Amsterdam/Philadelphia: John Benjamins; 2007. p. 171–200.

Journal : Large 12559 Dispatch : 23-9-2010 Pages : 17
Article No. : 9071 □ LE □ TYPESET
MS Code : ☑ CP ☑ DISK