

AN ACTION-BASED CONCEPT FOR THE PHONETIC ANNOTATION OF SIGN LANGUAGE GESTURES

Bernd J. Kröger¹, Jim Kannampuzha¹, Dominik Bauer¹, Peter Birkholz¹,
Philippe Dreuw², Hermann Ney²

¹Department of Phoniatics, Pedaudiology, and Communication Disorders,
University Hospital Aachen and RWTH Aachen University, Aachen, Germany
bkroeger@ukaachen.de

²Department of Computer Science, Human Language Technology and Pattern Recognition,
RWTH Aachen University, Aachen, Germany

Abstract: Background: Communicative actions are specific movements or gestures which are accomplished by vocal tract articulators (lips, tongue, velum etc.) for speech, by facial articulators (eye brows, eye lids etc.) for co-verbal facial expression and by other bodily articulators (hands, arms etc.) for co-verbal gesturing. While action-based approaches already exist for spoken language processing it is the aim of this paper to adopt action theory for signed language processing (i.e. production and perception of sign language). **Method:** An action-based method for the phonetic annotation of sign language has been developed and a 100 sentence American Sign Language corpus has been analyzed using this method. **Results:** Five basic types of sign actions were identified, all indicating the importance of movement phases even if the goal of a gesture is to reach a specific target (e.g. specific hand shape, orientation, location, and/or direction). **Conclusion:** This study is a starting point for investigating sign language production quantitatively in terms of a unified action theory.

1 Introduction

Specific movements of bodily articulators form the basis of human communication in all modalities (signed as well as spoken language) and in all bodily domains (vocal tract, hand-arm, and facial system). The goal of these *communicative actions* is always *shape forming*, i.e. forming a distinct vocal tract shape (distinct positioning of lips, tongue, velum) for realizing a specific speech sound resulting in a distinct auditory shape, forming a distinct position, and/or orientation, and/or shape of hands relative to the body, or producing distinct hand or finger movements (e.g. hand or finger circling or describing a shape of an object) for realizing a specific co-verbal hand-arm gesture, and/or forming a distinct shape of facial articulators (e.g. eye brows, eye lids, cheeks, mouth) for realizing a specific facial expression. Moreover an action is not just the basic unit for production but also for perception in face-to-face communication (Kröger et al. in press). While the concept of communicative actions has been introduced for the spoken language modality (ibid.), it is the goal of this paper to apply it to the signed language modality.

2 The organization of sign language actions

On the basis of sign language literature (e.g. Liddell and Johnson 1989, Corina and Sandler 1993, Valli and Lucas 2000) a preliminary action-based concept for sign language can be postulated: A communicative action, here a *sign action* or *sign* comprises form and meaning. The *meaning* of a sign is a word-level lexical concept like LOVE, RUN, HE/SHE/IT, YESTERDAY etc. (see e.g. Fant and Bernstein-Fant 2008 for an contemporary ASL lexicon).

The *form* of each sign action can be described as an ensemble of *manual* or *hand sub-actions* (i.e. left hand and right hand sub-actions) controlling specific *articulator systems* (or *articulatory tiers*), e.g. the *strong (or dominant) hand-arm tier* and the *weak hand-arm tier* (Fig. 1), where the right hand is the strong or dominant hand in case of right handed speakers and vice versa. Hand sub-actions are produced sequentially on each articulatory tier in all utterances (e.g. “WOMAN ARRIVE HERE.”). The first hand sub-actions within an utterance starts from a speaker- and situation-specific rest position of the hand-arm-systems (e.g. hands in the fold of the sitting signer) and the last hand sub-action of an utterance is followed by an “end” hand action, which lets return the hands towards that rest position. Hand sub-actions occur without temporal gaps on the strong hand tier but in addition hand sub-actions occur for weak hand tier in the case of symmetric signs (i.e. left and right hand produce the same movements) or in the case of asymmetric signs (i.e. weak hand serves as a reference location for a strong hand sign). Each hand sub-action is composed of *elementary movement actions* (or simply *movement actions*) i.e. *path movement*, *shape movement*, *orientation movement*, or *secondary movement action* (see Fig. 1). Elementary movement actions in comparison to sign actions are not meaning-carrying units. Rather they represent a bundle of *sign features*. Possible feature specifications for the movement sub-actions of the hand-arm systems are listed in Tab. 1 (cf. Lebourque and Gibet 1999).

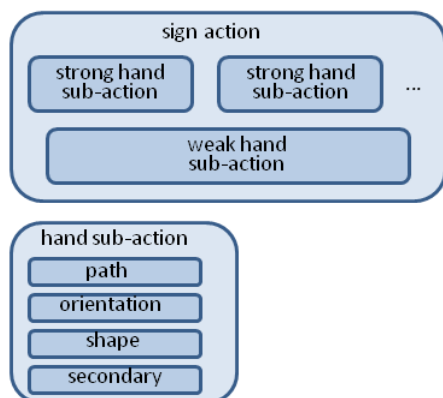


Figure 1. Organization of a sign action. Top: A sign action comprises a strong hand (manual) sub-action (optionally more) and optionally one temporally co-occurring weak hand sub-action. Bottom: Each manual sub-action comprises one or more elementary movement actions (path, orientation, shape, and secondary movement actions).

hand-arm path movement	hand shape movement	palm orientation movement	secondary movement
<u>type:</u> line - arc - circle	<u>type of hand shape:</u> straight - spread - fist - angle - hook/claw/C	<u>movement direction:</u> towards - away - up - down - left - right	<u>articulator:</u> finger (+spec.) - whole hand
<u>direction:</u> up - down left - right forward - backward	<u>active finger:</u> thumb, index, middle, ring, little, combinations	<u>type of movement:</u> rubbing- flicking - bending - hooking - wiggling - rotating - nodding
<u>location:</u> near - far - contact (+spec.)	<u>finger form:</u> straight - hook - round		...
...	<u>finger distance:</u> spread - contact - crossing		
	<u>FA (+ spec.)</u>		
	...		

Table 1. Features and their specifications (not complete) for elementary movement actions for the hand-arm articulator systems, i.e. for hand-arm path movement, hand shape movement, palm orientation movement, and secondary movement actions (see Fig. 1). Specification of contact means touching a cheek, an ear, the nose, the chest etc.; specification of FA means: specific letter of finger alphabet.

The goal of any communicative action is *shape forming* (Kröger et al. in press). In the signed language modality we can separate static and dynamic shape forming. *Static shape forming* is e.g. to attain a specific hand shape, a specific target orientation of the palm, and/or to attain a specific body space target location of the hand. *Dynamic shape forming* occurs in the case of

secondary movements, also called micro, hand internal, or local movements like finger wiggling (Liddell and Johnson 1989, Perlmutter 1992, Corina and Sandler 1993). In the case of secondary movement actions the shape or target is the secondary movement itself. In action theory it is assumed that static shapes are perceived basically during the *movement phase* of elementary actions. Thus a *shape movement action* has the goal of producing a specific hand shape. A *path movement action* can have the goal (i) of reaching a (static) target body location/position with or without body or hand contact or (ii) of forming a specific sign language movement for signaling a movement direction. The *orientation movement action* is in between. In most cases this movement action supports the basic hand shape forming sub-action, i.e. realizes a specific hand orientation, but in the absence of a path movement action the orientation movement action can substitute a specific path movement (e.g. in BREAK DOWN).

Beside manual sign actions described above the head-body articulator tier comprises *facial expression actions* and *oral expression actions* as well as further actions controlling gaze, head, and body movements. Feature specifications for these movement actions are listed in Tab. 2 (cf. Nonhebel et al. 2004). *Facial expression actions* control articulators like eyebrows, forehead, eyelids, cheeks, mouth corners etc. as is described in the facial action coding system (e.g. Cohn et al. 2007). They convey underlying emotions (e.g. ANGRY, SAD), connotations like DOUBT or BELIEF, or they convey sentence level concepts like QUESTION or STATEMENT (see Fig. 2). *Oral expression actions* control speech-like movements of lips, jaw and tongue tip and temporally co-occur with sign production (see mouth-tier in Fig. 2; see also Valli and Lucas 2000 for definition of mouth movements and mouthing in sign language).

head movement	body movement	mouth movement	gaze	facial movement
<u>movement type:</u> nod - shake - tilt	<u>movement of shoulder:</u> up - normal (down)	<u>mouth aperture:</u> open - closed - pressed	<u>direction toward:</u> communication partner - hands (left, right)	<u>eyebrows:</u> raised - lowered
<u>movement direction:</u> left - ahead - right	<u>movement of upper part of body:</u> left- right forward - backward	<u>lip rounding:</u> rounded - unrounded	<u>direction free:</u> left - right - upward - downward	<u>forehead:</u> smooth - furrowed
...	...	<u>visible tongue:</u> not - apical - labial	...	<u>eyelids:</u> wide - narrow - closed - blink
		...		<u>mouth corners:</u> lowered - raised
				...

Table 2. Features and their specifications (not complete) for elementary movement actions for the head-body articulator system, i.e. for head, body, and mouth (or oral) movement, for gaze, and for facial movement actions (see Fig. 1).

3 Phonetic annotation of sign language actions

A 200-sentences-ASL-corpus (Boston 2000), i.e. a corpus of 201 sentences produced by 3 different deaf signers (2 female, 1 male, 28-42 years old; data acquisition by video, frontal view, 30 frames/s) has been analyzed using our action-based phonetic annotation tool for sign language (cf. Fig. 2). The beginning of a hand sub-action (small dark gray vertical lines in Fig. 2) is defined by the beginning of the shape movement or - if the shape has not changed - by the beginning of the path movement action for this hand sub-action. The end of a basic hand sub-action coincides with the beginning of the next movement action for that hand or is defined by and absolute hold for that hand (i.e. no more arm, hand, and finger movement). The transcription of each movement action for a sample utterance is listed in Tab. 3. It should

be noted that sign actions can be composed of two or more sequential strong hand sub-actions (see the strong-hand sub-actions W1 and W2 on the right hand tier for the sign action WOMAN in Fig. 2). After estimating the beginning and end of all hand sub-actions the exact beginning and end of the *movement phases* of all movement action is determined and marked by dark gray bars. The resulting light gray intervals indicate that the appropriate articulator is in its target state for that movement action (*target phase* or *steady-state phase* of a movement action).



Figure 2. Score for hand and head-body sub-actions (three main tiers: right hand, left hand, head body) including appropriate elementary movement actions for the ASL-utterance “WOMAN ARRIVE HERE?” produced by speaker 1 (utterance 109 from Boston 2000). Movement phases of movement actions are indicated by dark gray bars. Right hand is the strong hand and left hand is the weak hand in the case of this speaker.

We use the convention to write sub-actions of different tiers in brackets in brackets $\{\dots\}_{h,o,f}$, where h, o, f indicate the tier of the sub-action, i.e. hand, oral, or facial tier. Thus $\{W1\}_h$ and $\{W2\}_h$ indicate two strong hand sub-actions for realizing the sign WOMAN (cf. Tab. 3). $\{A\}_h$ indicates a strong and weak hand sub-action for the sign ARRIVE. It can be seen that (i) the hand shape of the weak hand is already adjusted during $\{W2\}_h$. This indicates *co-production of elementary movement actions*, leading to a more efficient and faster realization of signs. $\{H\}_h$ indicates the temporally synchronous hand sub-actions for HERE and $\{e\}_h$ indicates the “end”- or “to-rest”-movement action towards the hands rest-position. Within the head-body sub-action tier $\{W1\}_o$ to $\{W4\}_o$ indicate oral movement actions realizing the mouth shapes of speech sounds of WOMAN (cf. Tab. 3). The same holds for the oral movement action $\{A\}_o$ for ARRIVE. $\{Q\}_f$ indicates facial (as well as head and shoulder) movement actions signaling the phrase level concept QUESTION (cf. Tab. 3). This action is hold during the whole utterance. Its “end”- or “to-rest”-action is labeled as $\{e\}_f$.

4 Results: Types of sign language actions in ASL

From a preliminary analysis of 100 randomly chosen sentences from our 200-sentence-ASL-corpus (Boston 2000) five different types of sign actions based on differences in parallel and serial ordering of shape, orientation, path, and secondary movement actions of the strong hand can be identified.

meaning of sign action	hand, facial, or oral sub-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	labiodental 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 facial	Q	oral	closure (also labeled as W1)	
			facial	mouth corners down; inner brows down and tightened; outer brows up	
			gaze	towards communication partner	
			head	slightly down	
			body	shoulders up	
Ending: “end” or “to-rest”	right hand	e	all	towards rest position	
	left hand	e	all	towards rest position	
	oral	e	oral	closure	
	head-body	e	all	towards neural	

Table 3. Listing of all features for each elementary movement action occurring within the ASL-utterance “WOMAN ARRIVE HERE?” as produced by speaker 1 (utterance 109 from Boston 2000). The abbreviations for hand and head-body sub-actions are also indicated in Fig. 2. Orientation means palm orientation. Letters specified as features for mouth movement actions in this table indicate letters of the appropriate concept word. No sub-specification of oral and facial articulators occurs in this table.

(i) The *single path type* comprises a single path movement action for example in order to signal the sign specific direction or location. Its hand shape movement action mostly starts at the same time but is completed in the first half of the duration of the path movement action. This *sign-specific hand shape* is held during the rest of the path movement action (e.g. in HE-SHE-IT, YOU, SIT, LOVE). (ii) The *serial paths type* comprises two or more sequent path movement actions. Generally the first path movement action is a so-called epenthetic movement in terms of movement-hold phonology (cf. Valli and Lucas 2000, p. 42ff) for reaching the starting hand position for this sign. This path movement action is overlaid by the sign specific shape movement action in the same way as described for the single path type actions. The sign-specific hand shape is held during the second sub-action and its movement action comprises the *sign-specific movement* (e.g. in SEE, YESTERDAY, FRIEND, BUY). (iii) The *shape-secondary type* comprises a path movement action for reaching the *sign-specific hand position* and this movement action is temporally overlaid by a shape movement action for generating the sign-specific hand shape. If the hand shape is reached, it is followed by a secondary movement action (*hold-secondary type* e.g. in COLOR, MOTHER, DIRTY, GERMANY). In some sign actions of the shape-secondary type, the secondary movement is temporally overlaid with a sign-specific path movement action (*movement-secondary type* e.g. in PLEASANT, FINGERSPELL, FLY). (iv) The *path-shape synchronous type* comprises an epenthetic path movement action overlaid by a specific hand shape movement action in order

to generate the sign-specific initial hand shape and to achieve the starting hand position. Then the sign-specific (second) path movement action takes place which is temporally totally synchronous with a sign-specific (second) hand shape movement action (e.g. in LIKE, LEAVE, TAKE, ASK). (v) The *shape-tap* type comprises a path movement action overlaid by a shape movement action at its beginning. After shape forming a tap or transient touch with the weak hand or with a body location occurs during the rest of the path movement action (e.g. in NEW and in fast or reduced articulated variants of single path or serial path types).

The frequency distribution of these five types of sign actions is given in Fig 3. It can be shown that these five types are comparable to different sequential types of signs as defined in movement-hold phonology (e.g. M-, H-, MH-, HMM-signs; M = movement, H = hold; cf. Perlmutter 1992, Valli and Lucas 2000) if in addition epenthetic movements (cp. Valli and Lucas 2000, p. 42) are allowed to form the initial parts of the sign action. It should be noted that the most frequent type (i.e. single path type) comprises only one strong hand sub-action while all other types comprise two sequent strong hand sub-actions.

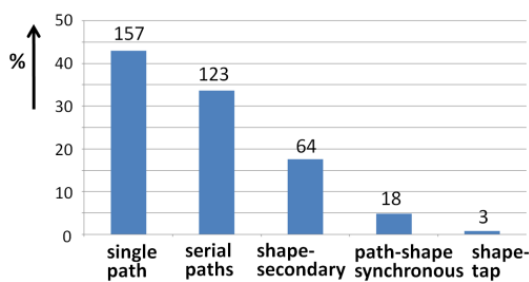


Figure 3. Percentage and absolute number of different types of sign actions for all analyzed signs occurring within 100 randomly chosen sentences from our ASL corpus (Boston 2000). The total number of analyzed signs is 365 (=100%). The values above bars indicate the absolute number of sign actions for each type.

A first quantitative analysis of the four most frequent types of sign actions (i.e. excluding the shape-tap type) reveals that the temporal duration of shape movements (i.e. movement phases of shape movement actions) is approximately 150 ms in all types, while the duration of one single path movements is approximately 250 ms while the duration of two sequent path movements (in the serial path type) is approximately 400 to 500 ms for different signs and different speakers (Fig. 4). Peak velocity of path movements after shape forming is around 1 m/s for all types of signs except the shape-secondary type, since secondary movements produced without path movements occurred in our analysis. These constant durations and peak movement velocities for different signs and different speakers may help to perceive and understand the continuum of ongoing hand-arm movements in sign language production.

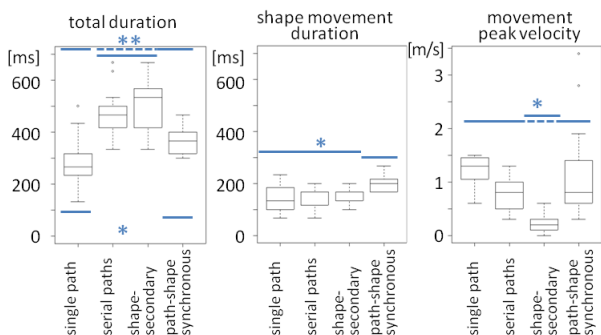


Figure 4. Box plots for total duration, shape movement duration, and movement peak velocity for 15 tokens of single path, serial paths, shape-secondary, and path-shape synchronous types of sign actions. If no median is plotted, the median in all cases coincides with the upper end of the quartile range. Movement peak velocity is estimated by using hand position tracking data (Dreuw et al. 2007). Significant differences are estimated by ANOVA followed by post hoc Tukey HSD test (* $p < 0.05$; ** $p < 0.01$).

A further main outcome of this study is that orientation movement actions seem not to be as primary as path, shape, and secondary movement actions. Our analysis of 365 sign actions reveals that all orientation movement actions can be seen as subordinate to path movement actions in 38% or to shape movement actions in 62% of all sign actions.

Furthermore it is important to notice that we did find no absolute holds (i.e. complete hold of both hands) in our 100 sentences corpus if sentence final sign actions are excluded. This

distinguishes sign language gesturing from co-verbal manual gesturing. The later exhibits many and long hold phases in order to stay synchronous with the acoustic speech signal.

5 Concluding remarks

While phonetics-based approaches to sign language description are rare, our phonetic annotation concept delivers a basis for a quantitative phonetic analysis of sign language production. This concept is not limited to American Sign Language. Moreover our action-based annotation concept can be taken as a starting point for computer-based synthesis or recognition of signed languages since elementary movement actions can be assumed as basic units or “atoms” in production and perception of signed languages. Further studies are planned in order to describe quantitatively coarticulation and reduction phenomena occurring in signed languages.

Acknowledgments. This work was supported in part by the German Research Council DFG grant Kr 1439/13-1 and grant Kr 1439/15-1.

Literature

- [1] Boston-200-Sentences-ASL-Corpus of the National Center for Sign Languages and Gesture Resources at Boston University (2000), <http://www.bu.edu/asllrp/csllgr/>
- [2] Cohn JF, Ambadar Z, Ekman P: Observer-based measurement of facial expression with the facial action coding system. In: Coan JA, Allen JJB (eds.) Handbook of Emotion Elicitation and Assessment (Oxford University Press US, New York, 2007) pp. 203-221
- [3] Corina D, Sandler W: On the nature of phonological structure in sign language. *Phonology* 10 (1993) 165-207
- [4] Dreuw P, Rybach D, Deselaers T, Zahedi M, Ney H: Speech Recognition Techniques for a Sign Language Recognition System. Proceedings of Interspeech 2007 (Antwerp, Belgium) pp. 2513-2516
- [5] Fant L, Bernstein Fant B: The American Sign Language Phrase Book (New York, McGraw Hill, 3rd edition, 2008)
- [6] Kröger BJ, Kopp S, Lowit A: A model for production, perception, and acquisition of actions in face-to-face communication. *Cognitive Processing* (in press). DOI: 10.1007/s10339-009-0351-2
- [7] Liddell SK, Johnson RE: American Sign Language: the phonological base, *Sign Language Studies* 64 (1989) 195-277
- [8] Lebourque T, Gibet S: A complete system for the specification and the generation of sign language gestures. In: Braffort A, Gherbi R, Gibet S, Richardson J, Teil D (eds.) *Gesture-Based Communication in Human-Computer Interaction*, LNCS 1739 (1999, Springer, Berlin) pp. 227-238
- [9] Nonhebel A, Crasborn O, van der Kooij E: Sign language transcription conventions for the ECHO Project. Technical report (Radboud University Nijmegen, 2004) http://www.let.kun.nl/sign-lang/echo/docs/ECHO_transcr_conv.pdf
- [10] Perlmutter DM: Sonority and syllable structure in American Sign Language. *Linguistic Inquiry* 23 (1992) 407-442
- [11] Valli C, Lucas C: *Linguistics of American Sign Language. An Introduction* (Gallaudet University Press, Washington, 2000)