Prosody and Movement in American Sign Language: A Task-Dynamics Approach

Martha E. Tyrone [1], *Hosung Nam* [1], *Elliot Saltzman* [1, 2], *Gaurav Mathur* [3], *Louis Goldstein* [1, 4]

[1] Haskins Laboratories, New Haven, CT, USA
[2] Department of Physical Therapy and Athletic Training, Boston University, Boston, MA, USA
[3] Department of Linguistics, Gallaudet University, Washington, D.C., USA
[4] Department of Linguistics, University of Southern California, Los Angeles, CA, USA

tyrone@haskins.yale.edu, nam@haskins.yale.edu, esaltz@bu.edu, gaurav.mathur@gallaudet.edu, louisgol@usc.edu

Abstract

This study examines prosody in American Sign Language using the theoretical framework of articulatory phonology, which proposes that the basic units of speech are articulatory gestures. We hypothesize that articulatory gestures are also the structural primitives of sign, and we are investigating what the gestures are and how they are timed. Kinematic data are collected as ASL users produce target signs with movements toward or away from the body, in phrase-initial, medial, or final position. Preliminary data suggest that signs are lengthened at phrase boundaries in a manner consistent with the predictions of a task-dynamic model of prosodically induced slowing.

Index Terms: ASL, signed language, task dynamics, articulatory phonology

1. Introduction

This paper reports on a new study investigating prosody and movement in American Sign Language (ASL). Sign data are analyzed in the framework of Articulatory Phonology, which proposes that the minimal units of speech production are invariant articulatory gestures.

Signed languages are the natural languages used by Deaf communities worldwide, and American Sign Language (ASL) is the language used by the Deaf community in the United States and parts of Canada. Sign prosody has primarily been described in terms of the actions of the non-manual articulators, including head movement, facial expression, and eyegaze. For example, many researchers have observed that signers raise their eyebrows to mark a new discourse topic or a yes-no question [1, 2], lower their eyebrows during a WH-question [3, 4], and produce a head nod at the end of a phrase or for emphasis [5]. In this study, we are examining sign prosody by looking at the timing of the movement of the dominant hand, as an effect of a sign's position in a phrase. Based on previous studies [6-8], we predicted that signs would be lengthened at phrase boundaries, particularly in phrase-final position.

There are four major phonological parameters that can differentiate one ASL sign from another: handshape, location, orientation and movement [9, 10]. Movement is probably the least well understood of the phonological parameters, because its realization has unlimited degrees of freedom and is thus difficult to typologize or to reliably measure. In addition, the realization of the movement for a given sign can vary depending on many factors, such as rate, emphasis, or prosodic context [5, 11, 12]. Grosjean (1979) found that increased signing rate caused greater modifications to movement than to the other phonological parameters [12]. Similarly, Wilbur and Schick (1987) found that an emphasized sign may be produced with greater movement amplitude or speed [13].

In part due to the variability of movement compared to other phonological parameters, researchers have debated whether a sign's movement is encoded independently from the other sublexical units, or whether it is sufficient to describe movement in terms of a change in handshape, location or orientation [14, 15]. Analysis of sign kinematics can be used to determine more precisely how sign movement is affected by prosodic context or by other factors. For this study, we are looking at two simple types of sign movements (movement toward contact with a location on the body, and movement away from contact with a location on the body) across multiple positions in a phrase. We are focusing on articulatory gestures that we hypothesize are organized sequentially.

Past studies of phrase-final lengthening in ASL have focused primarily on the holds that are appended to sign movements at phrase boundaries [5-7, 11]. Indeed, Perlmutter (1993) argues that phrase-final lengthening in ASL is different from lengthening in spoken languages, because ASL lengthens the final hold at the end of the sign rather than the movement itself [7]. (However, some studies have also found lengthening of movements phrase-finally [8].) Grosjean and Lane (1977) asked experienced signers to watch videos of ASL narrations and to hold down a button for the full duration of any pause that they observed [6]. Using this method, the researchers found that the longest pauses occurred at sentence boundaries.

We are examining prosody in the framework of Articulatory Phonology, and hypothesizing that signed languages are going to recruit the same prosodic mechanisms as spoken languages. While the two language modalities use a different set of articulators and perceptual organs, they nonetheless both rely on multivariate production systems, which can organize the linguistic information stream flexibly, via adjustments in prosody. The key idea of Articulatory Phonology is that spoken words are a constellation of vocal tract constriction units, or gestures, which are coordinated with respect to each other [16, 17]. Gestures' relative locations and temporal overlap determine the structure of language output. While constriction gestures are discrete and context-independent, articulator trajectories are continuous and context-dependent. For example, the articulators (tongue tip, tongue body, and jaw) involved for /d/ in /idi/ behave differently from those for /d/ in /ada/. At the same time, the constriction for /d/ can be captured as a discrete and context-independent gestural action unit. Coordination among speech gestures is also affected systematically by prosodic and performance (e.g. rate, precision) contexts. The Task Dynamic model of sensorimotor control and coordination has been used to implement the gestural units of Articulatory Phonology computationally [18, 19].

Gestural activations for an utterance are defined by a *gestural score*, which is input to the interarticulator level of the model. In turn, this level generates a corresponding set of time-varying trajectories for the constriction tract variables and the articulator degrees of freedom of a model vocal tract. Articulator trajectories can then be used to calculate area functions, sound sources, and acoustic output. In gestural scores, each gesture is assigned appropriate values for its intrinsic parameters (stiffness, damping, and target), and its activation function is specified over an interval of time with a magnitude ranging between 0 and 1. During the activation interval of a gesture, “forces” created at the tract variable level are a function of activation magnitudes, and are used to shape coordinated movement of the articulators that result in the attainment of the task-demanded constriction target. Coarticulatory variation in articulator trajectories is accounted for naturally in the model as a result of temporal overlap in the activation intervals of adjacent gestures.

Within this framework, Byrd, Kaun, Narayanan, and Saltzman (2000) and Byrd (2000) described a conceptual approach to boundary-adjacent slowing [20, 21]. They proposed that phrase boundaries are instantiated by a pi-gesture (or prosodic-gesture), which functions to slow all simultaneously active constriction gestures in proportion to the activation level of the pi-gesture. Like articulatory gestures, which have durational properties and are temporally coordinated and can overlap with other gestures, pi-gestures also have durations and overlap with vocal tract constriction gestures.

These studies show that when speech is discretely divided into a pattern of dynamically-controlled actions, each of which achieves a goal defined in an abstract task space, then it is clear that the ensemble of actions exhibits systematic slowing in proximity to prosodic boundaries. In addition, the kinematics of this slowing can be accurately modeled by modulation gestures whose goal is to slow down the (internal) clock used to activate the production of actions. Thus, these slowing events can be viewed as a primary mechanism by which the flow of phonological primitives are organized into informational groupings [22]. Our question in the current study is whether this same mechanism of informational grouping of primitive units also applies in signed languages, irrespective of the differences in the articulators themselves.

2. Methods

2.1. Data Collection and Apparatus

Sign movements are recorded with the Vicon motion capture system. Thirty markers are attached to participants’ sign articulators (7 on each arm, 7 on the head, and 9 on the torso) and tracked by six cameras at a 100Hz sampling rate. Figure 1 illustrates the positions of the markers that are tracked during sign production. The data are analyzed off-line using Matlab software and other applications specifically designed for task dynamic analyses of speech.

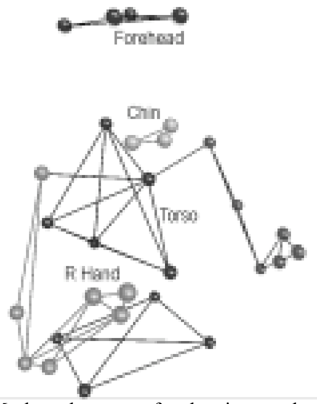

Figure 1: Marker placement for the sign production task. The end effector and locations of interest are labeled.

2.2. Procedure

Participants in the study are native ASL signers from the local Deaf community. During data collection, participants produce ASL phrases that are presented as written English glosses with accompanying illustrations. Participants direct their productions to a Deaf interlocutor. Target signs include a simple movement toward or away from the body, and phrase boundaries are manipulated, so that the target signs occur phrase-initially, finally or medially. Movement trajectories and velocity profiles are compared across different prosodic contexts.

The target signs are located at the forehead, chin, and torso. Some example target signs are: STRAIGHT, WILLING, DISAPPOINTED and SICK (Figures 2-5). The target signs are embedded in carrier phrases and occur either phrase-initially, phrase-finally or phrase-medially. For example, the target sign DISAPPOINTED is embedded in each of the following three ASL utterances:

KNOW DAD DISAPPOINTED. || NOT PROUD.
KNOW DAD DISAPPOINTED NOT. || PROUD.
KNOW DAD? || DISAPPOINTED NOT. || PROUD.

These constructions make use of the flexible word order of ASL, which allows the sign NOT to modify the sign that either precedes or follows it.

Figure 2: STRAIGHT

Figure 3: WILLING

Figure 4: DISAPPOINTED

Figure 5: SICK

2.3. Analysis

For this study, we are modeling simple sign gestures, such as bringing the hand to the torso and moving it away from that location, as ASL signers do when producing the sign WILLING. These sign gestures can be described in terms of the Euclidian distance between the end effector and the target location for a given sign. For example, in bringing the hand to the torso at the beginning of the sign, the task variable is the distance between the hand and the torso. Task variable motions are due to corresponding motion of the system's model articulators: shoulder flexion and extension, shoulder adduction and abduction, shoulder rotation, elbow flexion and extension, elbow supination and pronation, wrist flexion and extension, wrist adduction and abduction, and metacarpophalangeal flexion and extension. The task dynamic model has not been fundamentally altered for its application to sign language data. This model employs gestural input with appropriate dynamic parameters to generate the trajectories of task variables and model articulator variables. As with speech, each sign gesture is defined in terms of an activation magnitude, time interval, stiffness, damping, and target. In addition, because of the set of signs that we have chosen to examine, for the time being, we are retaining the existing terminology from Articulatory Phonology and referring to an arm movement toward the body as a "constriction" and a movement away from the body as a "release".

The kinematic sign data are segmented to identify the beginnings and ends of individual signs, so that their trajectories and velocity profiles can be compared across utterances with different coarticulatory and prosodic contexts. Because the target signs are all simple movements toward or away from the body, the beginning of a sign can be identified from reversals in movement direction.

3. Results

The data presented here are from one native ASL signer. To visualize the sign gestures, we have used the application mview, to display the movements necessary to produce a gesture. Mview is a multi-channel visualization application for dynamic movements that was developed by Mark Tiede at Haskins Laboratories. Figure 6 shows time funcations of the hypothesized gestural task variable (hand to torso distance) for individual tokens of the signer's production of the target sign WILLING in the following contexts:

KNOW NIECE? || WILLING NOT. || STUBBORN.
KNOW NIECE WILLING NOT. || STUBBORN.
KNOW NIECE WILLING. || NOT STUBBORN.

We used a labeling procedure for delimiting gestural extents, which identifies the following gestural landmarks on the signal(s) using velocity criteria computed with central differencing: gestural onset (time when velocity towards constriction exceeds a local velocity threshold), offset of the peak velocity preceding maximum constriction, plateau onset (time the velocity falls below that threshold again), minimum velocity point (assumed to be maximum constriction), plateau offset (time when velocity away from constriction exceeds a local velocity threshold), and gestural offset (time when velocity away from constriction falls blow threshold again). The durations are defined as follows: constriction is the time from the gesture onset to the plateau onset; plateau is the time from the plateau onset to the plateau offset, and release is the time from the plateau offset to the gesture offset. This gestural analysis suggests that the release phase (when the hand is moving away from the torso) is proportionally elongated in the phrase-final condition. The dotted vertical line in the lowest pane demarcates the long post-release hold frequently observed for signs in phrase-final position. In addition to these phrase-final effects, the plateau (when the hand is nearly stationary at the torso) is longer relative to the release in the phrase-initial condition.

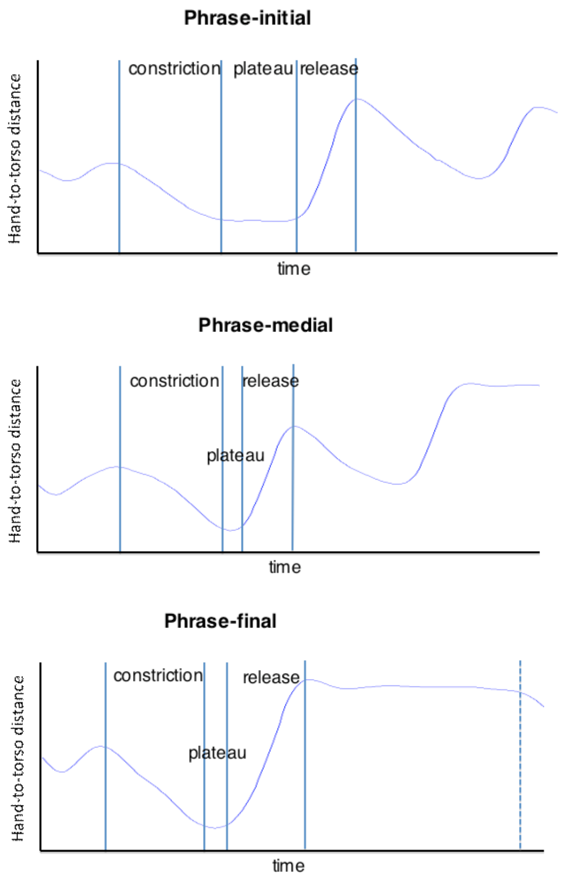

Figure 6: The hand-to-torso distance as a task variable for sample productions of the ASL sign WILLING, in phrase-initial, phrase-medial, and phrase-final position.

	Constriction	Plateau	Release
initial	2653.572	760.563	1429.829
medial	2578.08	471.697	1368.666
final	2554.532	556.137	1966.014

Table 1: Mean durations (msec) of the separate gesture phases for 10 tokens of the sign WILLING in each phrase boundary condition.

Table 1 shows the mean duration values (in milliseconds) of the constriction, plateau, and release phases for the signer's productions of the sign WILLING. The data shown here are for 10 tokens from each phrase boundary condition. The constriction and plateau phases are longest in the phrase-initial condition, and the release is longest in the phrase-final condition. The plateau and release are the shortest when the sign occurs in the phrase-medial condition. A two-way ANOVA reveals a significant interaction between the phrase boundary condition and the duration of the separate gesture phases ($F=2.83$, $p=.03$).

4. Discussion

Research on ASL prosody in general has focused mostly on the actions of the non-manual articulators. Moreover, research specifically on phrase-final lengthening in ASL has highlighted the insertion of pauses at phrase boundaries. Our preliminary findings suggest that phrase-final lengthening in ASL is not solely an epenthesis of a pause onto a sign movement, but rather a slowing of the sign's entire release gesture, analogous to the type of boundary adjacent slowing that occurs in speech. Like other studies, we also identified a cessation of the hand's movement in phrase-final position, but in addition to this we found that the sign movement itself was slowed in proximity to a boundary. On a related point, we are somewhat hesitant to refer to these elongated holds as "pauses", given that the articulators hold the final position of the sign rather than returning to a rest position. The question of what constitutes a pause in sign as opposed to speech deserves further consideration.

In the final stage of this study, the gestural information from the model and from the human sign data will be input to an animation software program. We will recruit native ASL users, who did not participate in the production experiment to judge sign intelligibility and naturalness. The animations will be used to test the whether the introduction of a pi-gesture could lead signers to interpret a sentence according to a different syntactic structure. In the future, the sign animations could also be used to test the relative importance of manual and nonmanual gestures in influencing the perception of prosodic boundaries, and perhaps help to address the role of the non-manual articulators in ASL prosody and syntax [2-4, 23].

In future studies, we hope to conduct more elaborate testing of the naturalness of sign prosody produced according to the task-dynamic model. Such perceptual studies would provide an important line of evidence for the validity of our model of sign language structure. By using synthesized data to test the model, we can precisely manipulate parameters in the model and thereby carry out a principled analysis of the prosodic structure that we have hypothesized.

5. Conclusions

Our initial findings suggest that the mechanism of phrase-final lengthening in signed languages may be more similar to that of spoken languages than previously realized. With respect to these data, Articulatory Phonology provides a mechanism for explaining both the elongation of the sign movement and the hold that follows it via a single control mechanism. More generally, it provides a cohesive framework for describing invariant properties of speech at one level, and variation (such as what occurs at prosodic boundaries) at another level.

This study contributes to the broader research goals of elucidating the minimal units of sign language structure and developing physiologically-based measures of sign production. In addition to further elucidating the structure of signed language, this type of research is necessary for future development of sign synthesis and recognition systems that are designed appropriately for their intended users.

6. Acknowledgements

This research is funded by National Institutes of Health grant DC009466.

7. References

[1] Klima, E.S. and Bellugi, U. The Signs of Language, Harvard, 1979.
[2] Weast, T.P. Questions in American Sign Language: A Quantitative Analysis of Raised and Lowered Eyebrows. PhD Thesis, University of Texas at Arlington, 2008.
[3] Liddell, S.K. "Non-manual signs and relative clauses in American Sign Language", In P. Siple [Ed] Understanding Language through Sign Language Research, 59-90, Academic, 1978.
[4] Liddell, S.K. American Sign Language Syntax, Mouton, 1980.
[5] Wilbur, R.B., "Effects of varying rate of signing on ASL manual signs and nonmanual markers", Language and Speech 52(2/3): 245-285, 2009.
[6] Grosjean, F. and Lane, H. "Pauses and syntax in American Sign Language", Cognition 5: 101-117, 1977.
[7] Perlmutter, D., "Sonority and syllable structure in American Sign Language", In G.R. Coulter [Ed], Current Issues in ASL Phonology, 227-261, Academic, 1993.
[8] Coulter, G.R., "Phrase-level prosody in ASL: Final lengthening and phrasal contours", In G.R. Coulter [Ed], Current Issues in ASL Phonology, 263-272, Academic, 1993.
[9] Stokoe, W.C., Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. Linstok, 1960.
[10] Battison, R. Lexical Borrowing in American Sign Language, Linstok, 1978.
[11] Wilbur, R.B. "Stress in ASL: Empirical evidence and linguistic issues", Language and Speech 42(2-3): 229-250, 1999.
[12] Grosjean, F. "A study of timing in a manual and a spoken language: American Sign Language and English", Journal of Psycholinguistic Research 8(4): 379–405, 1979.
[13] Wilbur, R.B. and Schick, B.S. "The effects of linguistic stress on ASL signs", Language and Speech 30: 301-323, 1987.
[14] Uyechi, L., The Geometry of Visual Phonology. CSLI Publications, 1996.
[15] Hulst, H.v.d., "On the other hand", Lingua 13: 43-73, 1996.
[16] Browman, C. P. and Goldstein, L., "Articulatory phonology: An overview", Phonetica 49: 155-180, 1992.
[17] Goldstein, L. and Fowler, C., "Articulatory phonology: A phonology for public use", in N.O. Schiller and A.S. Meyer [Eds], Phonetics and Phonology in Language Comprehension and Production, 159-207, Mouton de Gruyter, 2003.
[18] Saltzman, E. and Munhall, K., "A dynamical approach to gestural patterning in speech production", Ecological Psychology 1(4): 333-382, 1989.
[19] Saltzman, E.L. and Kelso, J.A.S., "Skilled actions: A task-dynamic approach", Psychological Review 94(1):84-106, 1987.
[20] Byrd, D., Kaun, A., Narayanan, S. and Saltzman, E., "Phrasal signatures in articulation", in M.B. Broe and J.B. Pierrehumbert [Eds] Papers in Laboratory Phonology V, 70-87, Cambridge, 2000.
[21] Byrd, D., "Articulatory vowel lengthening and coordination at phrasal junctures", Phonetica 57(1): 3-16, 2000.
[22] Byrd, D. and Saltzman, E. "The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening," Journal of Phonetics 31: 149-180, 2003.
[23] Sandler, W. and Lillo-Martin, D. Sign Language and Linguistic Universals, Cambridge, 2006.