

Computer-based Tracking, Analysis, and Visualization of Linguistically Significant Nonmanual Events in American Sign Language (ASL)

Carol Neidle*, Jingjing Liu**, Bo Liu**, Xi Peng**, Christian Vogler***, Dimitris Metaxas**

*Boston University / **Rutgers University / ***Gallaudet University

Boston University, Linguistics Program, 621 Commonwealth Avenue, Boston, MA 02215
Rutgers University, Computer and Information Sciences, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019
Gallaudet University, Technology Access Program, 800 Florida Ave NE, Washington, DC 20002

E-mail: carol@bu.edu, jl1322@cs.rutgers.edu, lb507@cs.rutgers.edu, px13@cs.rutgers.edu,
christian.vogler@gallaudet.edu, dnm@cs.rutgers.edu

Abstract

Our linguistically annotated American Sign Language (ASL) corpora have formed a basis for research to automate detection by computer of essential linguistic information conveyed through facial expressions and head movements. We have tracked head position and facial deformations, and used computational learning to discern specific grammatical markings. Our ability to detect, identify, and temporally localize the occurrence of such markings in ASL videos has recently been improved by incorporation of (1) new techniques for deformable model-based 3D tracking of head position and facial expressions, which provide significantly better tracking accuracy and recover quickly from temporary loss of track due to occlusion; and (2) a computational learning approach incorporating 2-level Conditional Random Fields (CRFs), suited to the multi-scale spatio-temporal characteristics of the data, which analyses not only low-level appearance characteristics, but also the patterns that enable identification of significant gestural components, such as periodic head movements and raised or lowered eyebrows. Here we summarize our linguistically motivated computational approach and the results for detection and recognition of nonmanual grammatical markings; demonstrate our data visualizations, and discuss the relevance for linguistic research; and describe work underway to enable such visualizations to be produced over large corpora and shared publicly on the Web.

Keywords: American Sign Language (ASL), nonmanual grammatical marking, computer-based sign language recognition

1. Overview

The linguistic annotation that has been carried out over the last 20 years or so by the American Sign Language Linguistic Research Project (ASLLRP) on video data collected from native users of American Sign Language (ASL) has included close attention to facial expressions and head gestures that can convey essential linguistic information. We have annotated, for example, events involving changes in eyebrow configuration, eye aperture, and head position—distinguishing the "onset" and "offset" phases, where relevant, of types of specific events (such as raised or lowered eyebrows, or head nods/shakes). Furthermore, we have labeled the linguistic information signaled by various combinations of these behaviors (topics, negation, multiple types of questions, *if/when* clauses, relative clauses, and so on) (Neidle 2002; Neidle 2007).

Our annotated corpora have formed a basis not only for linguistic research, but also for research to automate sign language detection by computer (e.g., Dreuw et al. 2008; Neidle et al. 2000). The ability to recognize linguistic information conveyed nonmanually is, of course, essential for computer-based sign language recognition and other types of applications (including, but not limited to, automatic translation) that rely upon such recognition. The general approach described here to recognition of nonmanual grammatical markers in ASL would be applicable, as well, to other signed languages.

In our earlier work, we tracked the position of the head and deformations of the face, and we used computer learning, based on the annotations of human transcribers from high-quality video images of native ASL signers, to develop the ability to discern and differentiate markings of topics, conditional clauses, negation, *wh*-questions, and *yes-no* questions, and we achieved fairly good success (Liu et al. 2013; Metaxas et al. 2012; Michael et al. 2011).

Our ability to detect, identify, and temporally localize the occurrence of nonmanual grammatical markings in ASL videos has recently been improved by incorporation of two principal innovations: (1) Newly developed techniques for deformable model-based 3D tracking, from a single video track, of head position and facial expressions (Liu et al. in press); and (2) A computational learning approach incorporating 2-level Conditional Random Fields (CRFs (Lafferty, McCallum, and Pereira 2001)) that is suited to the multi-scale spatio-temporal characteristics of the data (Liu et al. 2014). The computational analyses also enable us to produce visualizations showing the positions, over time, of the major articulators.

In Section 2, we summarize our current, linguistically motivated, computational approach, and the overall success rates now achieved for detection and discrimination of nonmanual grammatical markers. Section 3 addresses the computer-generated visualizations that we are now able to produce and their potential

value for linguistic research. In Section 4, we briefly describe work now underway to enable such visualizations to be produced over large corpora and shared publicly on the Web—as an extension of the interface described in (Neidle and Vogler 2012).

2. Computational Approach

Our current approach is summarized here. For further details about the methods and results, see Liu et al. (2014).

2.1. New tracking methods

Precise analysis of facial expressions, requiring the capture of spatio-temporal characteristics of facial events, has long been a challenging problem in computer vision. Most previous methods have been developed in controlled laboratory environments, with near-frontal faces and hardly any occlusions. For obvious reasons, these methods cannot be applied directly to ASL videos more generally, since large head movements and partial occlusions frequently occur while a subject is signing. Large and varied head movements would result in serious feature distortions of facial events. To address this problem, we take a 3D approach whereby facial expressions can be represented in a pose-invariant way.

We use a 3D deformable model-based face tracker that twines facial point localization and head pose estimation in a unique 3D shape model. Our two-stage cascaded 3D deformable shape face model localizes facial landmarks, allowing large head pose variations (Yu et al. 2013). For deformation, the first step uses mean-shift local search with a constrained local model (CLM) to achieve the global optimum. The second step uses component-wise deformable models to refine the subtle shape variation. From a single video track, we obtain 2D image coordinates of 66 facial landmarks, the corresponding 3D face shape, as well as 3 head rotations (i.e., pitch, yaw, and roll). Then feature extraction, representations, and comparisons are carried out in 3D space.

Our face tracker is capable of tracking facial expressions in the presence of large head rotations (over 30 degrees) and occlusions of the face by the hands that may occur during signing. The use of the 3D face model eliminates the alignment procedure required in 2D approaches (e.g., Active Shape Models (ASM) (Ari, Uyar, and Akarun 2008) and Active Appearance Models (AAM) (Forster et al. 2012)), which often leads to errors in head pose and expression features, restricting use of such 2D approaches to videos with small head pose variations.

See Yu et al. (2013) for comparisons of the tracking accuracy of our current 3D face tracker with that of some state-of-the-art 2D techniques, including those we have used in our previous work on recognition of nonmanual markers in ASL. In all cases, the 3D method reduces the error rate by at least 50%. When tested on three public datasets (LFW (Huang et al. 2007), LFPW

(Belhumeur et al. 2011), and AFW (Zhu and Ramanan 2012)), the multiple-ASM tracker (the best of the 2D trackers) and our current 3D tracker had mean average pixel errors for the facial landmark image locations as shown in Table 1.

Dataset tested	Multiple-ASM (2D) tracker	Our 3D tracker
LFW	8.53	3.64
LFPW	17.33	7.37
AFW	20.33	9.13

Table 1. Multiple-ASM (2D) tracker vs. our 3D tracker: comparison of mean pixel error rate when tested on three public corpora

We cannot provide definitive validation of the tracking for this ASL dataset, since ground truth of the locations of the facial landmarks is not available. However, the tracking appears to be working well (based on human observations) except in 12 extreme cases out of 161, where it fails: 10 video clips had severe occlusions (in which 60% of the face is occluded for over 15 frames), and 2 had large head rotations (over 60 degrees). In these cases, however, because we are using a model-based tracker, we know that the tracking has failed (because of abrupt shape changes to the model). We, therefore, are able to reinitialize the tracker, as compared with 2D methods, where this is not possible. Thus, our face tracker provides a timely tracking failure alarm and recovers quickly from temporary loss of track, thereby resulting in significantly better tracking accuracy.

2.2. Computational learning approach

Whereas previous approaches to detection of linguistic information expressed nonmanually have generally focused on low-level appearance-based features found in individual video frames (e.g., Grossman and Kegl 2006; Michael, Neidle, and Metaxas 2010; Nguyen and Ranganath 2008; Piater, Hoyoux, and Du 2010; Rodomagoulakis et al. 2011), temporal patterning over domains of variable length is also extremely important. For example, periodic head movements (nods and shakes) are an important component in the expression of many types of linguistic information. However, evaluation of a head nod or head shake requires consideration of a pattern that occurs over a time period that can vary considerably in length. Thus, we need an approach that is well suited to the multi-scale spatio-temporal characteristics of the data, one that combines low-level appearance-based features and high-level features that involve recognition of particular types of gestures—such as events involving raised or lowered eyebrows, head nods, or head shakes—and linguistically motivated evaluation of their specific characteristics and temporal phases.

We use a computational learning approach that incorporates 2-level Conditional Random Fields (CRFs

(Lafferty, et al. 2001)). At the first level of the CRF, we attend to the low-level features, based on facial geometry and appearance as well as head pose, obtained through accurate 3D deformable model-based tracking. At the second level, we learn to recognize some of the major component events that are typically found as part of the nonmanual expressions that convey specific types of grammatical information, such as raised/lowered eyebrows and head nods/shakes. Furthermore, we partition these events into their temporal phases, so that we can, for example, separate out the anticipatory movements (as the articulators get into position) from the linguistically significant region of the event; see Figure 1. We also identify the relevant characteristics of the various types of events. For example, for periodic head movements, variations in frequency and amplitude can correlate with different types of grammatical markings. Negation typically involves a side-to-side head shake; however, this head shake differs in appearance from the slight rapid head shake that is sometimes found over at least part of a wh-question; see Figure 2. We then use this multi-scale, spatio-temporal combination of low- and high-level features, in combination with the linguistically annotated corpus, to learn to detect specific linguistically important markers and to determine the temporal extent of those markings (Liu, et al. 2014; Liu, et al. 2013). Our current overall framework is shown schematically in Figure 3.

2.3. Recognition of NMMs

The new tracking and computational learning methods described above provide substantial improvements over previous methods in identification, discrimination, and temporal localization of nonmanual grammatical markers in ASL. Compared with a baseline method using only low-level features, the use of the 2-level CRF improved recognition accuracy by 20%.

Currently we validate our system on recognition of 5 major types of NMMs in 85 utterance-length videos collected at Boston University by C. Neidle and her research group. The recognition results were evaluated on a test set that contained 55 instances of topic/focus marking, 16 conditional/*when* clauses, 35 negations, 7 wh-questions, and 5 yes/no questions. As shown in the confusion matrix in Table 2, about 90% of those NMMs were correctly detected and identified; 4% were not picked up; and 6% were detected but misidentified (and all those examples involved confusion between conditional/*when* clauses and either topics (5 cases) or a yes/no question (1 case); these markings are very similar in appearance, all including raised eyebrows). In addition, there were 3 instances of false positives, where NMMs were detected that had not been identified as such in the annotations.

For details about improvements in temporal accuracy resulting from the use of the new methods, and for comparisons with the success rates for NMM recognition obtained from previous methods, see Liu et al. (2014).

		Results of detection					
		Wh	Neg	Top	Y/N	C/W	NM
From annotations	Wh	6	0	0	0	0	1
	Neg	0	34	0	0	0	1
	Top	0	0	46	0	6	3
	Y/N	0	0	0	5	0	0
	C/W	0	0	0	1	15	0
	NM	1	0	1	1	0	-

Table 2. Confusion matrix of NMM recognition: **Wh** (Wh-question), **Neg** (Negation), **Top** (Topic/Focus), **Y/N** (Yes/no question), **C/W** (Conditional/*when* clause), **NM** (no marker)

3. Computer-generated Visualizations

3.1. Visualizations that can now be produced

Figure 4 shows graphs for two example sentences illustrating degrees of eyebrow height and eye aperture, as well as 3D head position. The purple lines in the bottom graphs represent the temporal extent of manual signs, for which English-based glosses are also displayed. The 5 types of NMMs that we are currently detecting are also displayed in the visualizations that are produced from the computational analysis. Although still images are illustrated in this figure, these are actually videos that can be advanced frame by frame, with the video alignment indicator marking the current frame in the graphs.

3.2. Potential value for linguistic research

The nonmanual channel plays a vital role in the expression of various kinds of linguistic and paralinguistic information. Although this has received a fair amount of attention in the linguistic literature since about the 1970's (Baker 1976; Baker 1979; Baker and Cokely 1980; Baker and Padden 1978; Liddell 1978; Liddell 1980; Neidle, et al. 2000; Sandler 2010; Wilbur 2000, among many others), precise analysis over large data samples has been limited by the unavailability of appropriate tools.

The need to quantify observations has been felt. This has led to various approaches involving painstaking techniques for measurement and annotation by humans. For example, Grossman and Kegl (2006) used SignStream® to record impressionistically-assigned numerical values for degrees of eyebrow height; West (2008) used a “Screen Calipers tool” to measure pixels, by hand, in order to determine eyebrow height; the 3500 measurements for this study of 270 sentences took about 170 hours.

The possibility of producing computer-generated measurements of nonmanual components of sign language in temporal relation to the production of manual signs, for substantial data sets, opens up exciting possibilities for types of linguistic research on signed languages that have never before been possible, as well

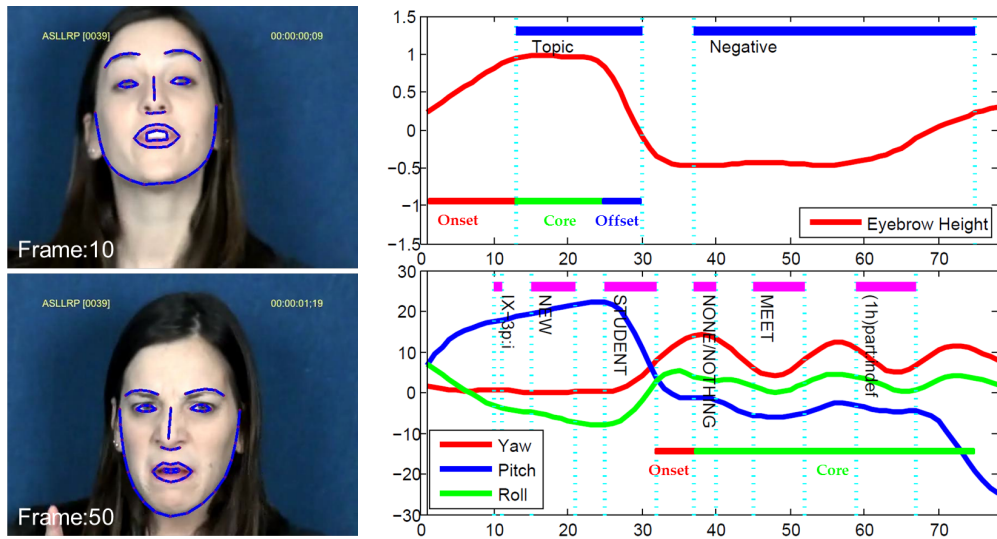


Figure 1: Detection of high-level (linguistically motivated) events—such as periodic head movements (here: head shake) and eyebrow gestures (here: raised eyebrows)—and partitioning of events into temporal phases to enable identification of the portion(s) that are of linguistic significance.

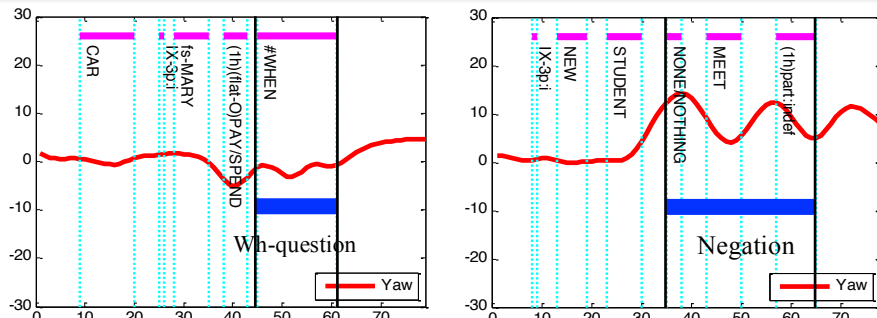


Figure 2: Analysis of the temporal patterns and properties of such detected events. For example, the head shake that occurs with negation is quite different (with respect to amplitude, velocity, peak value) from the slight rapid head shake that is sometimes found within wh-questions.

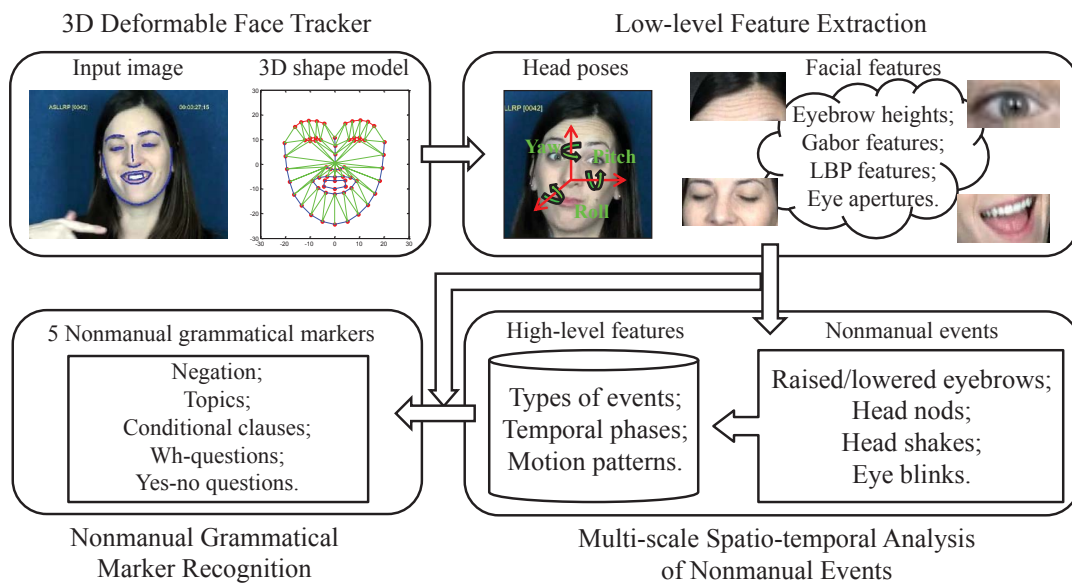


Figure 3. Overview of our current approach

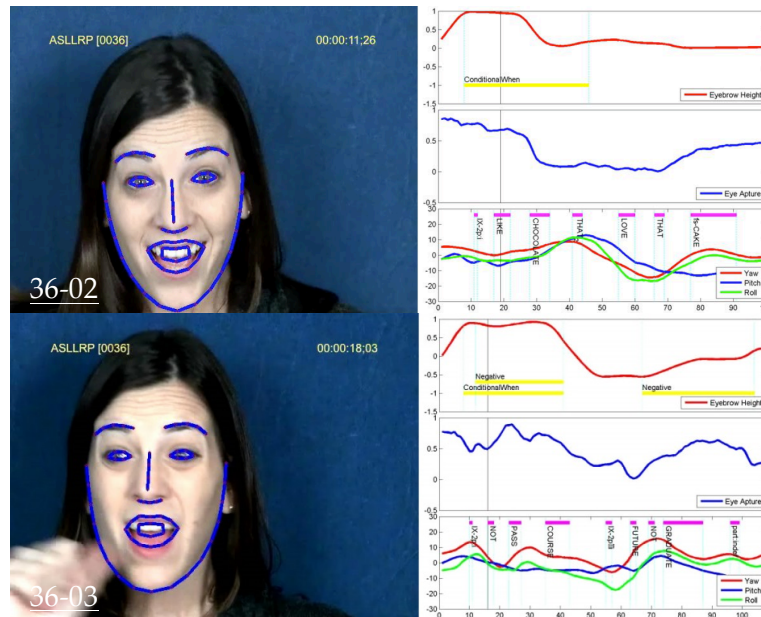


Figure 4. Visualizations of tracking data: eyebrow height (top -- in red); eye aperture (middle -- in blue); head position in three dimensions (bottom -- yaw in red, pitch in blue, roll in green)

as for cross-modality comparisons of various kinds (e.g., comparing the changes in eyebrow height in signed languages with intonation contours in spoken languages). These methods would be similarly applicable to the analysis of facial expressions and head gestures in spoken language, and thus to various types of comparisons between modalities with respect to the use of the non-manual channel. These results can also be applied to improving the linguistic realism of signing avatars, for which the unnaturalness of nonmanual expressions has been a serious issue (Kacorri, Lu, and Huenerfauth 2013).

4. Sharing Computer-generated Analyses and Visualizations

The ASLLRP corpora are shared publicly through a web-based Database Access Interface (DAI) described by Neidle and Vogler (2012). This interface allows easy searching and download of the corpora by gloss, sign type, classifier, and part of speech. Utterance and sign videos in the corpus can be viewed online in real time. The DAI is currently being extended to allow searching for utterances by nonmanual grammatical markers and nonmanual features, and to display the graphs of the computer analysis in the results list. The user can then drill down into each individual result and play back a full video of the computer analysis with the associated graphs.

Figures 5-7 illustrate a representative use case for the extended functionality: A researcher is interested in the kinematics of raised eyebrow movements in ASL, which are an important component of quite a few different NMMs. Starting with the retrieval of examples of topic markers, she selects the “topic/focus” option in the search form (Figure 5). Because eyebrows are the feature

of interest, she elects to display thumbnails of the eyebrow graphs in the search results list (Figure 6; other display options are eye aperture and 3D head pose). Together with the rough glosses in this list, the graphs allow the researcher to see at a glance where in the utterances the topic markers occur, and if they exhibit the typical eyebrow movement pattern. The thumbnail with the dual occurrence of topic markers catches her attention, and she would like to investigate this utterance in more detail. She clicks on the graph to bring up a full-resolution video showing the graphs and tracking of the facial markers in detail, frame by frame (Figure 7). She can subsequently repeat the process for the other grammatical constructions of interest and see at a glance whether the eyebrow movement patterns are similar to the ones seen for topics, or whether they differ.

Sharing the data via the web-based DAI, rather than merely making the annotations and video files available for download, offers several compelling advantages. First, it makes the data accessible to a much wider audience, including those who have no expertise in using linguistic annotation software, and it works out of the box in a web browser, which everyone has installed, as opposed to requiring the installation of special-purpose software. Second, the DAI has been designed for efficient search and retrieval over large corpora, and correlating linguistic phenomena across different annotation files and videos is much quicker and easier than it is with standalone software. Third, because of the nature of the web, referencing a specific linguistic phenomenon (e.g., a topic marker seen in a specific utterance) is as simple as sharing a link with a collaborator or student, which allows them to bring up the utterance in question with a single click; bringing up the same utterance in annotation software, in contrast, takes many more steps.

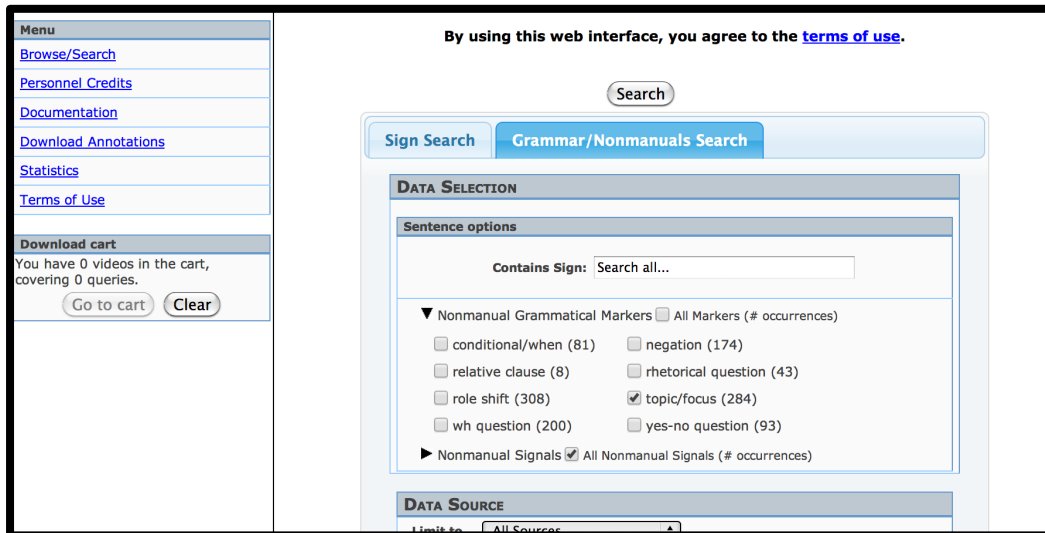


Figure 5. Search interface for nonmanual events or grammatical markers

File Name-Utterance	Utterance Video	Eyebrow Graph	Full Gloss	Rough Gloss
<input type="checkbox"/> rachel51.xml-2			Show ... topic/focus: topic main gloss: IX-3p:i CAR BREAK-DOWN "frustration"	
<input type="checkbox"/> rachel51.xml-16			Show ... topic/focus: topic main gloss: CAR BREAK-DOWN	
<input type="checkbox"/> rachel52.xml-8			Show ... topic/focus: topic main gloss: LIBRARY (lh)HAVE MANY MAGAZINE	
<input type="checkbox"/> rachel52.xml-9			Show ... topic/focus: topic main gloss: fs-BOB IX-loc:i WORK BICYCLE SELL+ IX-3p:i	
<input type="checkbox"/> rachel52.xml-14			Show ... topic/focus: topic main gloss: MOTHER+ NAME fs-JANET #NO+ NAME fs-JANE	
<input type="checkbox"/> rachel52.xml-15			Show ... topic/focus: topic main gloss: BOY+ WANT NEW SHOE	
<input type="checkbox"/> rachel52.xml-17			Show ... topic/focus: topic topic main gloss: FATHER IX-3p:i FUNNY MOTHER IX-3p:j SERIOUS	
<input type="checkbox"/> rachel52.xml-18			Show ... topic/focus: topic main gloss: BOY+ LIE++ BLAME+ #DOG IX-loc:i	

Figure 6. Illustration of Search Results

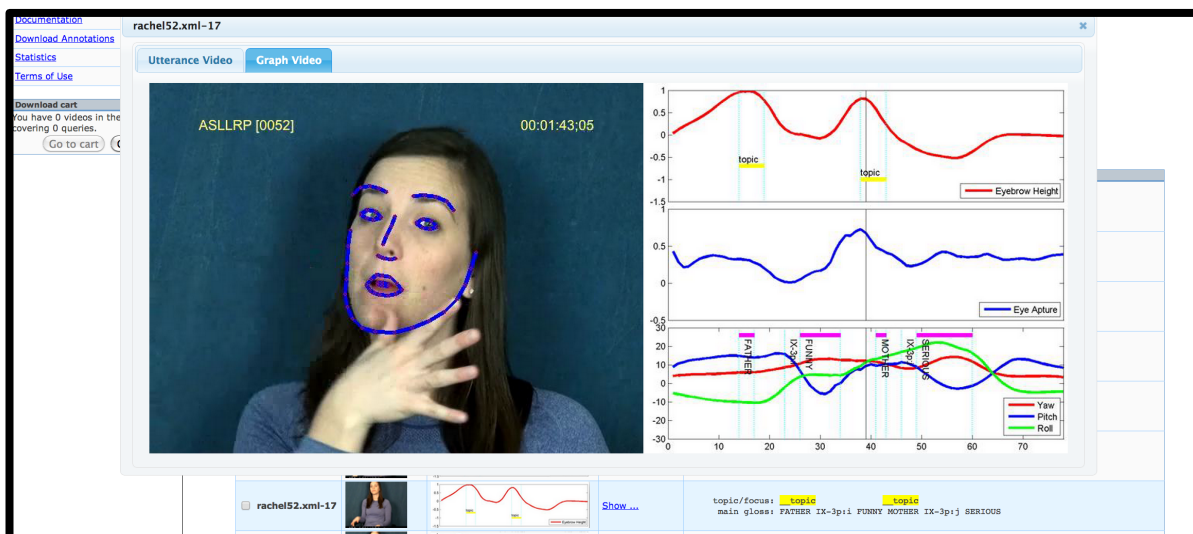


Figure 7. Video playback screen: the alignment indicator in the graph shows the position of the current video frame

5. Conclusions

We have summarized here new methods for computer-based detection of nonmanual grammatical markers in ASL, reported in greater detail in Liu et al. (2014). Such methods could readily be applied, as well, to the analysis of other signed languages, as well as to the production of nonmanual expressions used in conjunction with spoken languages.

These methods rely on computational learning, using a 2-level CRF that incorporates both low-level features and linguistically motivated higher-level features associated with types of head motion and eyebrow events that occur over varying spatio-temporal scales. The extraction of both the low- and high-level features benefits from a new 3D deformable face tracker, which achieves greater accuracy in tracking facial landmarks and head position than has been possible with even the best 2D approaches.

Visualizations of the results of the computational analyses, which can be run on large corpora, can also be generated. We plan to make these publicly available in conjunction with our web-accessible corpora. The availability of such materials offers great potential for use in linguistic research on the nonmanual components of ASL.

6. Acknowledgments

The research reported here was partially funded by grants from the National Science Foundation (CNS-1059281, IIS-1064965, IIS-0964597, IIS-1065013, CNS-0964385, EIA-9528985, and CNS-1059218). We thank Iryna Zhuravlova, SignStream® programmer, and gratefully acknowledge the participation of the many signers and students who have worked on related projects over many years. Thanks especially to Rachel Benedict, Braden Painter, Jonathan McMillan, Jessica Scott, Indya Oliver, Corbin Kuntze, Tory Sampson, Emma Preston, Donna Riggle, Amelia Wisniewski-Barker, and Joan Nash for their invaluable contributions.

7. References

- Ari, İ., Uyar, A. and Akarun, L. (2008). Facial Feature Tracking and Expression Recognition for Sign Language. Paper presented at the 23rd International Symposium on Computer and Information Sciences, 2008.
- Baker, C. (1976) What's not on the other hand in American Sign Language. In Mufwene, C. and Steever, S., (eds.) *Papers from the 12th Regional Meeting of the Chicago Linguistic Society*, Chicago: University of Chicago Press. pp. 24-32.
- Baker, C. (1979) *Nonmanual components of the sign language signal*, Paper presented at the NATO Advanced Study Institute, Copenhagen.
- Baker, C. and Cokely, D. (1980) *American Sign Language: A Teacher's Resource Text on Grammar and Culture*, Silver Spring, MD: T.J. Publishers.
- Baker, C. and Padden, C. A. (1978) Focusing on the nonmanual components of American Sign Language. In Siple, P., (ed.) *Understanding language through sign language research*, New York: Academic Press. pp. 27-57.
- Belhumeur, P., Jacobs, D., Kriegman, D. and Kumar, N. (2011) Localization parts on faces using a consensus of exemplars. *CVPR*.
- Dreuw, P., Neidle, C., Athitsos, V., Sclaroff, S. and Ney, H. (2008). Benchmark Databases for Video-Based Automatic Sign Language Recognition. *Proceedings of the 6th international Conference on Language Resources and Evaluation, LREC 2008, Marrakech, Morocco*.
- Forster, J., Schmidt, C., Hoyoux, T. and Koller, O. (2012). RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus. *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey*.
- Grossman, R. B. and Kegl, J. (2006) To Capture a Face: A Novel Technique for the Analysis and Quantification of Facial Expressions in American Sign Language. *Sign Language Studies*, 6(3), pp. 273-305.
- Huang, G., Ramesh, M., Berg, T. and Miller, E. (2007) *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*. Technical Report 07-49. University of Massachusetts, Amherst, October, 2007.
- Kacorri, H., Lu, P. and Huenerfauth, M. (2013) Evaluating Facial Expressions in American Sign Language Animations for Accessible Online Information. *Universal Access in Human-Computer Interaction. Design Methods, Tools, and Interaction Techniques for eInclusion, Lecture Notes in Computer Science Volume 8009*, pp. 510-519.
- Lafferty, J., McCallum, A. and Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Paper presented at the Intl. Conference on Machine Learning.
- Liddell, S. K. (1978) Nonmanual signals and relative clauses in American Sign Language. In Siple, P., (ed.) *Understanding language through sign language research*, New York: Academic Press. pp. 59-100.
- Liddell, S. K. (1980) *American Sign Language Syntax*, The Hague: Mouton.

- Liu, B., Liu, J., Yu, X., Metaxas, D. and Neidle, C. (2014). 3D Face Tracking and Multi-scale, Spatio-temporal Analysis of Linguistically Significant Facial Expressions and Head Positions in ASL. Paper presented at the Language Resources and Evaluation Conference, LREC 2014, Rejkjavik, Iceland.
- Liu, J., Liu, B., Zhang, S., Yang, F., Yang, P., Metaxas, D. N. and Neidle, C. (2013) Non-manual Grammatical Marker Recognition based on Multi-scale Spatial Temporal Analysis of Head Pose and Face. In *Proceedings of the 10th IEEE International Conference on Automatic Face and Gesture Recognition, Shanghai, China, April 25, 2013*.
- Liu, J., Liu, B., Zhang, S., Yang, F., Yang, P., Metaxas, D. N. and Neidle, C. (in press) Non-manual Grammatical Marker Recognition based on Multi-scale Spatio-temporal Analysis of Head Pose and Facial Expressions. *Image and Vision Computing Journal, The Best of Face and Gesture 2013 (Special issue)*.
- Metaxas, D., Liu, B., Yang, F., Yang, P., Michael, N. and Neidle, C. (2012) Recognition of Nonmanual Markers in American Sign Language (ASL) Using Non-Parametric Adaptive 2D-3D Face Tracking. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, the Language Resources and Evaluation Conference, LREC 2012, Istanbul, Turkey*.
- Michael, N., Neidle, C. and Metaxas, D. (2010). Computer-based recognition of facial expressions in ASL: from face tracking to linguistic interpretation. *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, the Language Resources and Evaluation Conference, LREC 2010, Malta*.
- Michael, N., Yang, P., Liu, Q., Metaxas, D. and Neidle, C. (2011) A Framework for the Recognition of Nonmanual Markers in Segmented Sequences of American Sign Language. *BMVC*.
- Neidle, C. (2002) *SignStream™ Annotation: Conventions used for the American Sign Language Linguistic Research Project*. Boston, MA: American Sign Language Linguistic Research Project Report No. 11, Boston University.
- Neidle, C. (2007) *SignStream™ Annotation: Addendum to Conventions used for the American Sign Language Linguistic Research Project*. Boston, MA: American Sign Language Linguistic Research Project Report No. 13, Boston University.
- Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B. and Lee, R. G. (2000) *The Syntax of American Sign Language: Functional Categories and Hierarchical Structure*, Cambridge, MA: MIT Press.
- Neidle, C. and Vogler, C. (2012) A New Web Interface to Facilitate Access to Corpora: Development of the ASLLRP Data Access Interface (DAI) In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, the Language Resources and Evaluation Conference, LREC 2012, Istanbul, Turkey*.
- Nguyen, T. D. and Ranganath, S. (2008). Tracking facial features under occlusions and recognizing facial expressions in sign language. *Proceedings of the IEEE Conference on Automatic Face & Gesture Recognition*.
- Piater, J., Hoyoux, T. and Du, W. (2010). Video Analysis for Continuous Sign Language Recognition. *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, the Language Resources and Evaluation Conference, LREC 2010, Malta*.
- Rodomagoulakis, I., Theodorakis, S., Pitsikalis, V. and Maragos, P. (2011). Experiments on Global and Local Active Appearance Models for Analysis of Sign Language Facial Expressions. *Proceedings of the 9th International Gesture Workshop (GW 2011): Gesture in Embodied Communication and Human-Computer Interaction, May 2011, Athens, Greece*.
- Sandler, W. (2010) Prosody and Syntax in Sign Languages. *Transactions of the Philological Society, 108(3)*, pp. 298-328.
- Weast, T. P. (2008) *Questions in American Sign Language: A quantitative analysis of raised and lowered eyebrows*. Unpublished Doctoral dissertation, University of Texas, Arlington.
- Wilbur, R. B. (2000) Phonological and prosodic layering of nonmanuals in American Sign Language. In Lane, H. and Emmorey, K., (eds.) *The signs of language revisited: Festschrift for Ursula Bellugi and Edward Klima*, Hillsdale, NJ: Lawrence Erlbaum, pp. 213-241.
- Yu, X., Huang, J., Zhang, S., Wang, Y. and Metaxas, D. N. (2013) Pose-free Facial Landmark Fitting via Optimized Part Mixtures and Cascaded Deformable Shape Model. *ICCV*.
- Zhu, X. and Ramanan, D. (2012) Face detection, pose estimation and landmark localization in the wild. *CVPR*.

