Generating Co-occurring Facial Nonmanual Signals in Synthesized American Sign Language

Jerry Schnepp[1], Rosalee Wolfe[2], John C. McDonald[2] and Jorge Toro[3]

[1]*College of Technology, Bowling Green State University, Bowling Green, OH, U.S.A.*

[2]*School of Computing, DePaul University, 243 S. Wabash Ave., Chicago, IL, U.S.A.*

[3]*Department of Computer Science, Worchester Polytechnic Institute, 100 Institute Road, Worcester, MA, U.S.A.*

Keywords: Avatar Technology, Virtual Agents, Facial Animation, Accessibility Technology for People who are Deaf, American Sign Language.

Abstract: Translating between English and American Sign Language (ASL) requires an avatar to display synthesized ASL. Essential to the language are nonmanual signals that appear on the face. In the past, these have posed a difficult challenge for signing avatars. Previous systems were hampered by an inability to portray simultaneously-occurring nonmanual signals on the face. This paper presents a method designed for supporting co-occurring nonmanual signals in ASL. Animations produced by the new system were tested with 40 members of the Deaf community in the United States. Participants identified all of the nonmanual signals even when they co-occurred. Co-occurring question nonmanuals and affect information were distinguishable, which is particularly promising because the two processes move an avatar's brows in a competing manner. This brings the state of the art one step closer to the goal of an automatic English-to-ASL translator.

1 INTRODUCTION

Members of the Deaf community in the United States do not have access to spoken language and prefer American Sign Language (ASL) to English. Further, they do not have effective access to written English because those born deaf have an average reading skill at or below the fourth-grade level (Erting, 1992). ASL is an independent natural language in its own right, and is as different from English as any other spoken language. Because it is a natural language, lexical items change form based on the context of their usage, just as English verbs change form depending on how they are used. For this reason, video-based technology is inadequate for English-to-ASL translation as it lacks the flexibility needed to dynamically modify and combine multiple linguistic elements. A better approach is the synthesis of ASL as 3D animation via a computer-generated signing avatar.

The language of ASL is not limited to the hands, but also encompasses a signer's facial expression, eye gaze, and posture. These parts of the language are called *nonmanual signals*. Section 2 describes facial nonmanual signals, which are essential to forming grammatically correct sentences. Section 3 explores the challenges of portraying multiple nonmanual signals and Section 4 lists related work. Section 5 outlines a new approach; section 6 covers implementation details and section 7 reports on an empirical test of the new approach. Results and discussion appear in sections 8 and 9 respectively.

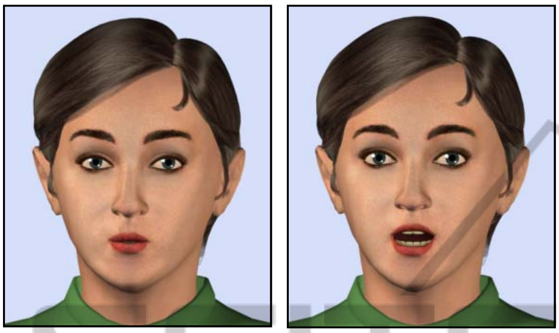
2 FACIAL NONMANUAL SIGNALS

Facial nonmanual signals can appear in every aspect of ASL (Liddell, 2003). Some nonmanual signals operate at the lexical level and are essential to a sign's meaning. Others carry adjectival or adverbial information. For example, the nonmanual OO indicates a small object, while CHA designates a large object. Figure 1 shows pictures of these nonmanuals demonstrated by our signing avatar.

Another set of nonmanual signals operate at the clause or sentence level. For example, raised brows can indicate yes/no questions and lowered brows can indicate WH-type (who, what, when, where, and how) questions. Figure 2 demonstrates the difference

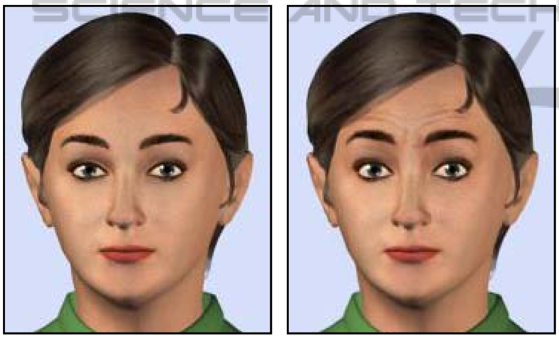
Schnepp J., Wolfe R., McDonald J. and Toro J..
Generating Co-occurring Facial Nonmanual Signals in Synthesized American Sign Language.
DOI: 10.5220/0004217004070416
In *Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications* (GRAPP-2013), pages 407-416
ISBN: 978-989-8565-46-4

between a neutral face and one indicating a yes/no question. With the addition of the Yes/No-nonmanual signal, a simple statement such as “You are home” becomes the question, “Are you home?” In fact, it is not possible to ask a question without the inclusion of either the Yes/No- or WH-question nonmanuals.


Nonmanual OO, indicating a small size

Nonmanual CHA, indicating a large size

Figure 1: Nonmanual signals indicating size.


Statement

Yes/no question

Figure 2: Sentence-level nonmanuals.

Affect is another type of facial expression which conveys meaning and often occurs in conjunction with signing. Deaf signers use their faces to convey emotions (Weast, 2008). Figure 3 demonstrates how a face can convey affect and a WH-question simultaneously.

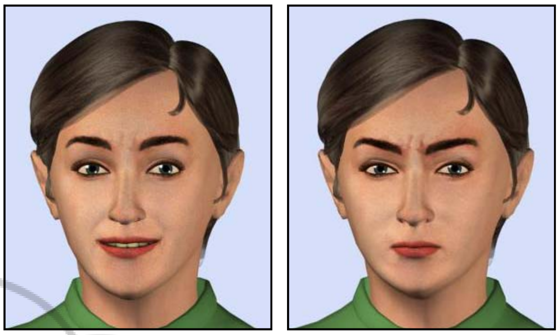
Challenges arise when nonmanual signals co-occur. Multiple nonmanual signals often influence the face simultaneously.

If a cheerful person asks a yes/no question about a small cup of coffee, this will combine happy affect with the Yes/No-question nonmanual and the small nonmanual OO.

The Yes/no question and the happy affect will influence the brows, and the happy affect and small nonmanual OO will influence the lower face.

Further, each signal has its own start time and duration.

The happy affect would continue throughout, with the Yes/No signal appearing well before the small nonmanual.


WH-question combined with happy

WH-question combined with angry

Figure 3: Co-occurrence.

3 SYNTHESIS CHALLENGES

For translation purposes, a video recording of ASL will suffice only when the text is fixed and will never change. However, video lacks the necessary flexibility to create new sentences. Attempting to splice together a sentence from previously-recorded video will result in unacceptably choppy transitions. A more effective alternative is to use an avatar to synthesize sentences. Signing avatars are implemented as 3D computer animation, and facial movements are handled either by morphing between fixed facial poses, or by using a muscle based approach (Parke and Waters, 1996).

ASL synthesis places unique requirements on an animation system which differ from those of the film industry. Since the avatar needs to respond to spontaneous speech, its facial expressions must be highly flexible and dynamic. Compare this to motion picture animation, where expressions are scripted for a scene, and then printed to film.

This brings us to a second and, for this discussion, even more critical requirement. Given the co-occurring nature of nonmanual signals, any signing avatar must take into account multiple simultaneous linguistic processes. Such a system must combine different types of expressions and facilitate the ways in which those expressions will interact.

No currently-available animation system completely fulfils these requirements. For example, the animation technique of simple morphing allows an animator to pre-model a selection of facial poses, and then choose one of these poses for each key frame in the animation. The system then blends between poses to make the face move. This method allows for only one pre-modelled facial pose at a time, which is extremely limiting in sign synthesis. Consider portraying a question involving a happy person asking about a small cup of coffee. This has three simultaneously occurring facial processes: the question nonmanual, the small size nonmanual and the happy affect. If the animator has modelled each of these separately, then a morphing system is forced to choose only one of them and ignore the other two, resulting in a failure to communicate the intended message.

Attempting to mitigate this issue by pre-combining poses lacks flexibility and is labour intensive to the point of impracticality. There are six basic facial poses for emotion (Ekman and Friesen, 1978) and at least fifty-three nonmanual signals which can co-occur (Bridges and Metzger, 1996). Trying to model all combinations would result in hundreds of facial poses. In addition, the timing of these combinations would suffer the same problems of flexibility associated with using video recordings.

Maskable morphing attempts to address the inflexibility problem by subdividing the face into regions such as Eyes, Eyebrows, Eyelids, Mouth, Nose, and allows the animator to choose a distinct pose for each region. This is an improvement but the "choose one only" problem now migrates to individual facial features, and thus it still does not support simultaneous processes that affect the same facial feature. For example, both the nonmanual OO and the emotions of joy and anger influence the mouth.

The technique of muscle-based animation more closely simulates the interconnected properties of facial anatomy by specifying how the movement of bones and muscles affect the skin (Magnenat-Thalmann, Primeau and Thalmann, 1987) (Kalra, Mangili, Magnenat-Thalmann and Thalmann, 1991). If two different expressions use the same muscle, their combined effect will pull on the skin in a natural way. However, managing and coordinating all of these muscle movements have a tendency to become overwhelming.

Timing is the main problem. Co-occurring facial linguistic processes will generally not have the same start and end times. Some processes may be present for a single word, others for a phrase, and others for an entire sentence. Errors in timing can change the meaning of the sentence. For example, both the affect anger and the WH-question nonmanual involve lowering the brows. If the timing is not correct, the WH-question nonmanual can be mistaken for anger (Weast, 2008). Errors in timing can also cause an avatar to seem unnatural and robotic, which can distract from the intended communication. This is analogous to the way that poor speech synthesis is distracting and requires more hearing effort (Warner, Wolff and Hoffman, 2006).

4 RELATED WORK

Several active research efforts around the world have a shared goal of building avatars to portray sign language. Their intended applications include tutoring deaf children, providing better accessibility to government documents and broadcast media, and facilitating transactions with service providers. This section examines their approaches to generating facial nonmanual signals.

Very early efforts focused exclusively on the manual aspects of the language only (Lee and Kunii, 1993; Zhao et al., 2000; Grieve-Smith 2002). Some acknowledge the need for nonmanual signals but have not yet implemented them for all facial features (Karpouzis, Caridakis, Fotinea and Efthimiou, 2007). Others have incorporated facial expressions as single morph targets. This has been done using traditional key-frame animation (Huenerfauth, 2011) and motion capture (Gibet, Courty, Duarte and Le Naour, 2011).

The European Union has sponsored several research efforts, starting with VisiCast in 2000, continuing with eSIGN in 2002 and currently, DictaSign (Elliott, Glauert and Kennaway, 2004; Efthimiou et al., 2009). One of the results of these efforts is the Signing Gesture Markup Language (SiGML), an XML-compliant specification for sign-language animation (Elliott et al., 2007). SiGML relies on HamNoSys as the underlying representation for manuals (Hanke, 2004), but introduces a set of facial nonmanual specifications, including head orientation, eye gaze, brows, eyelids, nose, and mouth and its implementation uses the maskable morphing approach for synthesis. However, there is no consensus on how best to specify facial nonmanual signals, particularly for the mouth, and other research groups have either developed their own custom specification (Lombardo, Battaglino, Damiaro and Nunnari, 2011) or are using an earlier annotation system such as Signwriting (Krnoul, 2010). Further, none of these efforts have yet specified an approach to generating co-occurring facial nonmanual signals.

Recent efforts have begun exploring alternatives to morphs and maskable morphs by exploiting the muscle based approach (López-Colino and Colás, 2012). However this work has not addressed portraying co-occurring nonmanual signals.

There is consensus that animating the face is an extremely difficult problem. Consider the sentence, "What size coffee would you like?" signed happily.

In a conventional system based solely on facial features, the brow would need to be lowered to indicate a WH-question, but happiness requires an upward movement of the brows. How much should the brows be raised to indicate this? Raise them too little and the face will not appear happy. Raise them too much and the face is no longer asking a WH-question. This type of manual intervention makes automatic synthesis difficult to the point of impracticality.

Given the challenges, it is not surprising that the previously-published empirical evaluation of synthesized nonmanual signals yielded mixed results. Huenerfauth (2011) reports that only animations containing emotion affected perception at a statistically significant level. Deaf participants did not comprehend any portrayals of nonmanual signals in the synthesized ASL.

5 A NEW APPROACH

Findings from linguistics yield fresh insight into the challenge of representing co-occurrences. ASL linguists have developed a useful strategy for annotating them. Figure 4 demonstrates a sample annotation for the question, "Do you want a small coffee?" The lines indicate the timing and duration of the nonmanual signals. Nonmanual signals co-occur wherever the lines overlap

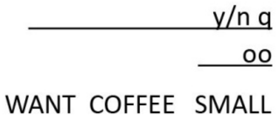


Figure 4: Linguistic annotation for the sentence "Do you want a small coffee?"

Using this notation as a metaphor makes it possible to express timing of co-occurring signals. The key is to view ASL synthesis as linguistic processes, rather than a series of facial poses. Linguistic processes can provide the timing and control for underlying muscle movements. This new approach creates a mapping of linguistic processes to anatomical movements which facilitates the flexibility and subtleties required for timing.

In the new approach, each linguistic process has its own *track* analogous to the timing lines in Figure 4. Each track contains blocks of time-based information. Each block has a label, a start time, an end time, as well as a collection of subordinate geometry blocks as outlined in Table 1.

Geometry blocks describe low-level joint transformations necessary to animate the avatar and can contain animation keys or a static pose. *Linguistic tracks* contain *linguistic blocks* which contain intensity and timing information that controls the geometry blocks. Additionally, linguistic blocks can contain intensity curves that control the onset and intensity of a pose to facilitate the requisite subtlety. The effect of each joint transformation is weighted by the curve values to vary the degree to which each pose is expressed.

Table 1: Representation Structure.

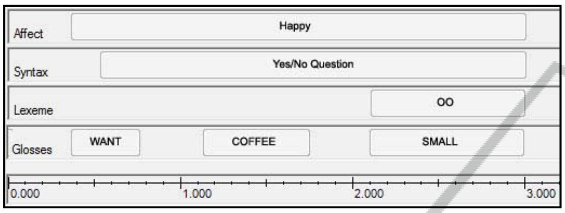
High Level Tracks Linguistic: syntax gloss (manuals) lexical modifier Extralinguistic: affect mouthing	**Lexical Modifier Block** Label Start time and duration Intensity curve Viseme(s) Label Geometry block
Syntax Block Label Start time and duration Intensity curve Geometry block **Gloss Block** Label Start time and duration Geometry block	**Affect Block** Label Start time and duration Intensity curve Geometry block **Mouthing Block** Label Start time End time Curve Viseme(s) Label Geometry block

Overlapping blocks in multiple linguistic tracks simultaneously influence the face. To implement co-occurrence, for each joint or landmark, keys are gathered from the relevant tracks. A matrix **M** is computed by combining the weighted transformations from each track at the current time per (1) and the new transformation **M** is applied to the joint.

$$\mathbf{M} = \prod_i \boldsymbol{w_i} \boldsymbol{M_i} \qquad (1)$$

This representation does not simply store animation data as a collection of keys, but organizes them into linguistic processes. This facilitates a natural mapping to a user interface. Figure 5 shows how the main interface of our ASL synthesizer reflects the annotation system that linguists use to analyse ASL sentences. In the interface, linguistic tracks are labelled on the left, and block labels refer to the linguistic information they contain.

To create a sentence, a linguist or artist types the glosses (English equivalences) for a sentence. The synthesizer automatically creates an initial draft of the animation and displays the linguistic blocks in the interface. Based on the animation's appearance, a linguist can shift or time-stretch any block and edit its internals by using its context-sensitive menu. After making desired adjustments, the linguist can rebuild the sentence, view the updated animation and repeat the process as necessary. We continually mine the editing data because it provides insights for improving the initial step of automatic synthesis.



Affect: Happy
Syntax: Yes/No Question
Lexeme: OO
Glosses: WANT, COFFEE, SMALL

Figure 5: Screen shot of ASL synthesizer interface, for the sentence, "Do you want a small coffee?".

Thus, the interface is not presenting the animation data as adjustments to a virtual anatomy. Rather, the interface allows researchers to focus on the linguistic aspects of the language instead of the geometric details of the animation. They can describe sentences in the familiar terms of linguistic processes such as "The Yes/No-question nonmanual begins halfway through the first sign and finishes at the end of the sentence."

This approach helps manage the complexity of ASL synthesis. It is useful for linguists, because the animation-specific technology is abstracted. What is presented to the linguist is an interface of linguistic constructs, instead of numerical animation data. The complexity of 3D facial animation is hidden; although it is available through the context-sensitive menus should a researcher want to access it.

6 IMPLEMENTATION DETAILS

The synthesis program contains a library of facial poses. To speed the creation of the initial library, we set up an "Expression Builder" which has a user interface similar to (Miranda et al., 2012). Figure 6 shows the Expression Builder interface on the left. On the upper part of the face, the interface controls correspond to landmarks that are constrained to move along the surface of a virtual skull.

For the mouth, we began with the landmarks specified in the MPEG-4 standard (Pandžić, and Forchheimer, 2003), but artists found that working with them directly was only partially satisfactory. Artists found it time-consuming to create the necessary nonmanual signals because jaw movements interfered with them, and the resulting animations were not deemed satisfactory by our Deaf informants.

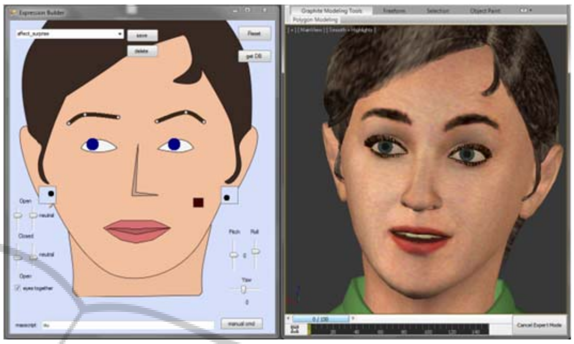


Figure 6: The Expression Builder Interface.

We addressed the problem by creating an oral sphincter that simulates the inward and downward motion at the corners of the mouth which occurs as the jaw drops. This technique automatically integrated the jaw movement with the mouth, and artists found the task of creating the nonmanual signals much easier.

For forehead wrinkling, we needed to incorporate the effects of both raising and furrowing the brows. We created two textures, one depicting horizontal lines caused by raised brows, and one depicting furrowed brows. Sometimes these effects can occur simultaneously and even asymmetrically. To support these possibilities, we created visibility masks that are generated dynamically in real-time based on the position of the landmarks in the brows. When the landmarks are in neutral position, the masks are transparent and the textures are invisible. Raising a landmark causes the horizontal texture to become visible near it. Similarly when a landmark is lowered or moved inward, the furrow texture becomes visible in the vicinity of the landmark. Figure 7 contains schematics of the textures and a rendered example demonstrating simultaneous and asymmetric brow configurations.

The interfaces for the Expression Builder and for the ASL Synthesizer were developed in Microsoft Visual Studio, and currently utilize a commercially-available animation package as a geometry engine. Further details of the implementation can be found in (Schnepp, 2012).

7 EVALUATION

Sign synthesis has an analogue to speech synthesis: the correct phonemes must be created as a precursor to attempting to synthesize entire paragraphs. Thus we focused our evaluation exclusively on short phrases and simple sentences. If we can ascertain that simple language constructs are understandable and acceptable to Deaf viewers, we can then use them as a basis for building more complex constructs in future efforts.

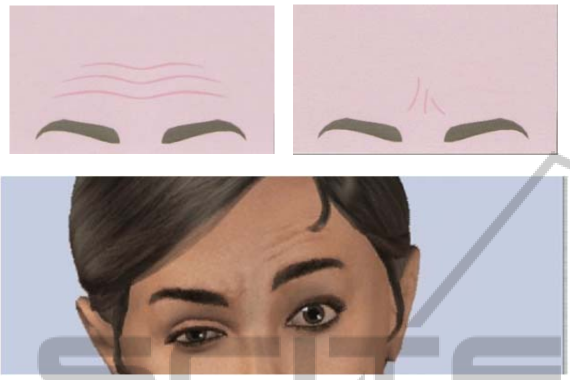


Figure 7: Sketches of the raised and furrowed textures and a rendered image of model with the textures.

We wanted to evaluate whether affect would still be perceptible even when there were other, simultaneously-occurring nonmanual signals that could potentially interfere. For first part of our study, we created two pairs of sentences. Each pair consisted of one sentence with happy affect and one sentence with angry affect. The first pair combined the WH-question nonmanual with each of these emotions. The second pair combined the CHA (large) nonmanual with the same two emotions.

We also wanted to assess the perceptibility of grammatical nonmanual signals in isolation from emotion, and to focus on evaluating the effect of nonmanual markers on the perception of size. We created a phrase that contained a manual sign that indicated a medium size, but then synthesized three variations -- one with an OO (small) nonmanual signal, one with a neutral face, and one with the CHA (large) nonmanual signal. Other than the nonmanual, the animations were identical. We could then ask participants to tell us the size of the object in the animation.

7.1 Test Considerations

We evaluated for clarity in three ways. The first method was a coarse measure, which was to simply ask participants to repeat what they saw in the test animation. This has the potential to uncover major problems. For example, if the animation displays a question but the participant responds by signing a statement, then the question nonmanual was not perceived. The second method was to ask questions about the content conveyed through nonmanual signals. For example, if an animation involved a cup of coffee we could ask about the cup's size. The third and final method was to ask the participant to rate the animation's clarity.

To address acceptability, we asked participants to "Tell us what we can do to improve the animation." From the responses, we gained both quantitative and qualitative data. A high number of negative responses would indicate a lower level of acceptability. The open-ended question also elicited suggestions for improvement, which are an invaluable resource for refinements.

When testing with members of the Deaf community, the same considerations need to be taken into account as when testing in a foreign language. Thus everything -- the informed consent, the instructions, the questionnaires, the test instruments -- must be in ASL. To avoid possible bias due to geographic location, we wanted a significant portion of the participants to come from regions other than our local area. To facilitate this we used SignQUOTE, a remote testing software package designed specifically for Deaf communities. A previous study found no significant variations between the responses elicited in face-to-face testing and responses elicited via remote testing with SignQUOTE (Schnepp et al., 2011).

7.2 Procedure

Twenty people participated in a face-to-face setting at Deaf Nation Expo in Palatine Illinois, while another twenty were recruited through Deaf community websites and tested remotely using SignQUOTE. All participants self-identified as members of the Deaf community and stated that ASL is their preferred language. In total, 40 people participated.

Participants viewed animations of synthesized ASL utterances (see http://asl.cs.depaul.edu/co-occuring) and answered questions pertaining to sentence content and clarity. Each participant viewed individual animations one at a time and was given the option to review the animation as many times as desired. When the participant was ready to proceed, the facilitator asked four questions:

1. The first question asked the participant to sign the animation.
2. The second question asked the participant to judge some feature of the animation as shown in Table 2. Participants indicated their responses on a five-point Likert scale.

3. The third question asked the participant to use a five-point Likert scale to rate the clarity of the animation.
4. Finally, the last question asked the participant to offer suggestions to improve the animation.

Table 2: Test animations.

Test animation	English Translation	Feature rated by participant
WH + Happy	How many books do you want? (Happy)	Emotion
WH + Angry	How many books do you want? (Angry)	Emotion
CHA + Happy	A large coffee. (Happy)	Emotion
CHA + Angry	A large coffee. (Angry)	Emotion
OO + Medium sign	A regular coffee (small nonmanual)	Size
No nonmanual + Medium sign	A regular coffee (no nonmanual)	Size
CHA + Medium sign	A regular coffee (large nonmanual)	Size

The face-to-face environment consisted of a table with a flat-panel monitor in front of the participant. On either side of the participant sat a Deaf facilitator, and a hearing note taker. Across from the participant sat a certified ASL interpreter who voiced all of the participant's responses.

The remote testing sessions followed an identical structure while automating the roles of facilitator and note taker. Namely, instructions were presented by pre-recorded video of a signing (human) facilitator and answers were recorded using a clickable interface for scale elements and webcam recordings for open-ended answers.

Once all remote testing sessions were complete, a certified ASL interpreter viewed the collection of webcam video recordings and voiced the participant's responses. The audio of the interpretations was recorded and transcribed as text for analysis.

8 RESULTS

In response to the first question, every participant repeated the utterance correctly for each animation. This included all of the processes that occur on the face.

For the second question, we used the Mann-Whitney statistic to analyze the responses to the paired sets of sentences. For the pair combining the WH-question nonmanual with happiness and anger, the Mann-Whitney test showed a significant difference (z = -6.1, p = 1.06 x 10^{-9} < .0001). The second pair combining the CHA nonmanual with happiness and anger yielded similar results (z = -6.83, p = 8.66 x 10^{-12} < .0001). Figure 8 shows the distribution of the participants' ratings for the first pair of sentences and Figure 9 shows the ratings for the second pair.

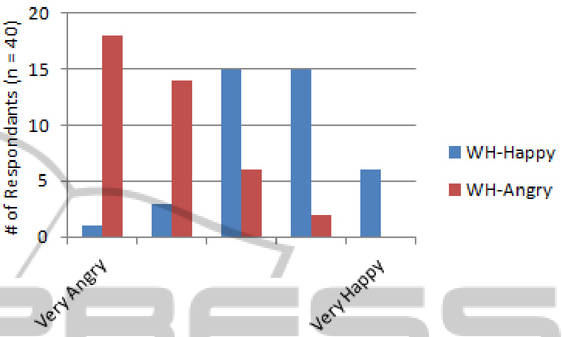


Figure 8: Perception of emotion in the presence of a WH-question nonmanual signal.

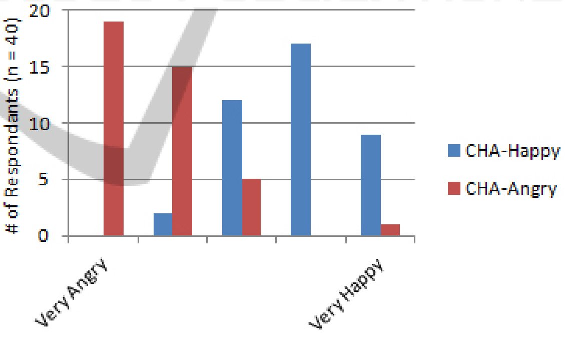


Figure 9: Perception of emotion in the presence of a CHA nonmanual signal.

Figure 10 and Figure 11 show the results of perceived size in the three animations that differed only in the portrayed nonmanual signal. Figure 10 displays the responses to the animation showing the OO (small) nonmanual compared to the animation with a neutral face. The Mann-Whitney statistic (z = -3.75, p < .000179) indicates a significant difference. Figure 11 shows the responses for the neutral face versus the CHA (large) nonmanual. As with the first case, the differences in the responses are significant (z = -3.51, p < .000452).

Figure 12 shows the participant's ratings of clarity. In each case, the majority of participants rated the animation as either 'clear' or 'very clear'.

Table 3 shows a tabulation of the responses to the open-ended question, "What can we do to improve the animation?" The categories are a) suggestions for improvement, b) no comment or positive comments ("she looks fine"). Representative suggestions for improvement were "You always need an expression", "When there's no expression, I'm not really sure," and "She shouldn't be so crabby."

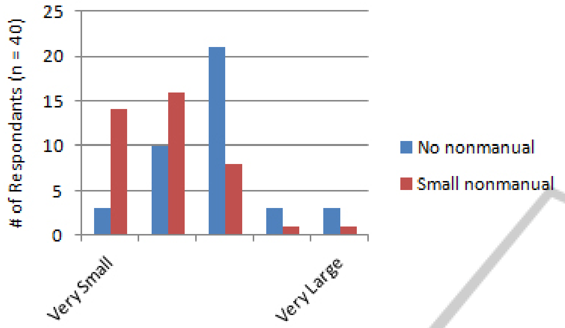


Figure 10: Perception of size (small vs. no nonmanual).

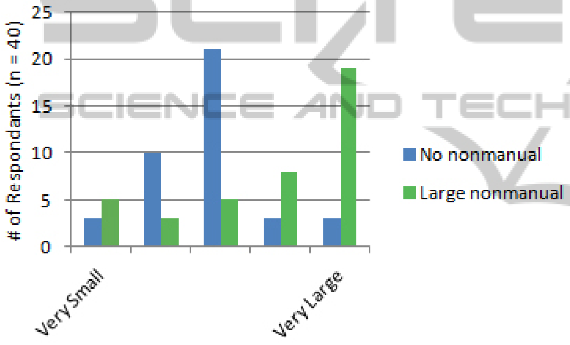


Figure 11: Perception of size (large vs. no nonmanual).

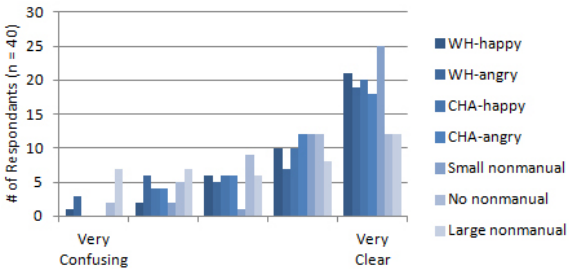


Figure 12: Clarity Results.

Table 3: Responses to open-ended questions.

ASL Animation	Suggestions for improvement	"Fine" or no comment
WH + Happy	23	17
WH + Angry	32	8
CHA + Happy	16	24
CHA + Angry	23	17
OO + Medium sign	15	25
Neutral + Medium sign	21	19
CHA + Medium sign	22	18

9 DISCUSSION

The first concern was whether an individual nonmanual signal produced by this system actually conveys its intended meaning. When participants viewed two animations identical except for a size nonmanual, they perceived the size of the object according to the nonmanual signal. The Mann-Whitney scores demonstrate a significant difference in perception when the size nonmanual occurs.

The second concern was whether affect would be perceived even in the presence of co-occurring nonmanual signals that could interfere with its portrayal. The Mann-Whitney statistics demonstrate that participants were able to perceive affect even in the presence of co-occurring nonmanuals. Further, when asked to repeat animations that involved both affect and another nonmanual, participants consistently signed all nonmanuals present in the animations, including size and question nonmanuals. Participants correctly identified both the emotional state of the avatar and the meaning of the co-occurring nonmanual signals.

The third concern was whether the WH-question nonmanual would be distinguishable from affect. Both anger and the WH-nonmanual lower and furrow the brows. Improperly produced, a WH-nonmanual can be mistaken for anger and vice-versa. But this did not happen in the study. Participants repeating animations depicting a question always signed back the proper form of the question. The Mann-Whitney statistics also demonstrate that they easily perceived the intended emotion. This last case is particularly interesting because happy affect and the WH-question nonmanual move the brows in opposite directions. Still, participants could discern both the emotional state of the avatar, and that the sentence being signed was a WH-question.

When viewing animations produced by our approach, participants accurately repeated each sentence with all included nonmanuals 100% of the time. This is interesting because there were no manual (hand) indications that a sentence was a question; the only indication was on the face. Still, participants all signed the questions accurately, including the intended nonmanuals.

These results are in contrast to (Huenerfauth, 2011), whose animations elicited a significant effect only when portraying affect: no linguistic nonmanual had a significant effect. Further, his approach was only capable of portraying one process on the face at a time. Our approach can express simultaneous processes, and the study data show that each of the simultaneous processes is recognizable.

Thus, the results for this new method promise a significant advance in portraying ASL.

Finally, in every case a majority of participants rated the animation as either ‘clear’ or ‘very clear’. Clarity ratings tended to be highest when nonmanual signals and manual signs reinforced each other. Although the one animation lacking an appropriate nonmanual signal was deemed relatively understandable, participants were in consensus that the animations were clearer when a nonmanual signal was present. To quote one participant, “You always need an expression.”

10 CONCLUSIONS

The use of linguistic abstractions as a basis for creating animations of ASL is a promising technique for portraying nonmanuals that are recognizable to members of the Deaf community. While this approach undoubtedly requires extension and revision, it is a step toward the automatic generation of ASL. In addition to being an essential component of an automatic English-to-ASL translator, an avatar signing correct ASL would be a valuable resource for interpreter training. When interpreting students learn ASL, recognition skills lag far behind production skills (Rudser, 1988). Software incorporating a signing avatar capable of correct ASL would provide a valuable resource for practicing recognition. Another application would be to support Deaf bilingual, bi-cultural (“bi-bi”) educational settings where ASL is used in preference to manually-signed English (Hermans, Ormel, Knoors and Verhoeven, 2008).

The underlying representation itself can support collaboration between ASL linguists and avatar researchers for exploring linguistic theories due to the direct analogue to linguistic annotation. Researchers can quickly make animations to test and refine hypotheses.

The number of suggestions for improvement from the study indicates there is more work to be done before the animations reach full acceptability. We are analyzing the qualitative feedback to determine next steps

Going forward, we plan to develop and evaluate additional nonmanual signals and follow up with more rigorous testing. The current study only tested the co-occurrence of two simultaneous signals. Three or more co-occurring signals often combine in signed discourse, and the system should be tested as to its scalability in terms of the number of co-occurring signals.

ACKNOWLEDGEMENTS

We would like to thank Brianne DeKing for her help and expertise. She provided invaluable insight into the language of ASL. Brent Shiver was essential to the testing sessions and to the development and testing of SignQUOTE. We would also like to thank Dr. Christian Vogler for his mentoring as we wrote this paper.

REFERENCES

Bridges, B. and Metzger, M., 1996. *Deaf Tend Your: Non-Manual Signals in American Sign Language.* Silver Spring, MD: Calliope Press.

Efthimiou, E., Fotinea, S.-E., Vogler, C., Hanke, T., Glauert, J., Bowden, R., Braffort, A., Collet, C. Maragos, P. and Segouat, J., 2009. Sign language recognition, generation and modeling: A research effort with applications in deaf communication. In: UAHCI '09, *Proceedings of the 5th International Conference on Universal Access in Human-Computer Interaction. Addressing Diversity*. San Diego, California, 19-24 July 2009. Berlin, Germany: Springer-Verlag.

Ekman, P. and Friesen, W., 1978. *Facial Action Coding System.* Palo Alto, CA: Consulting Psychologist Press.

Elliott, R., Glauert, J. and Kennaway, J., 2004. A framework for non-manual gestures in a synthetic signing system. In: CWUAAT 04, *Proceedings of the Second Cambridge Workshop on Universal Access and Assistive Technology*. Cambridge, UK, 22-24 March 2004.

Elliott, R., Glauert, J., Kennaway, J., Marshall, I. and Safar, E., 2007. Linguistic modelling and language-processing technologies for avatar-based sign language presentation. *Universal Access in the Information Society*, 6(4) pp.375-391.

Erting, E., 1992. Why can’t Sam read? *Sign Language Studies.* 75(2), pp. 97-112.

Gibet, S., Courty, N., Duarte, K. and Le Naour, T., 2011.The signcom system for data- driven animation of interactive virtual signers: Methodology and evaluation. *ACM Transactions on interactive intelligent systems*. 1 (1), 1-26.

Grieve-Smith, A., 2002. SignSynth: A sign language synthesis application using Web3D and Perl. In: I Wachsmuth and T Sowa, eds. *Gesture and Sign Language in Human-Computer Interaction.* Lecture Notes in Computer Science. 2298/2002 Berlin, Germany: Springer-Verlag. pp. 37-53.

Hanke, T., 2004. HamNoSys -- Representing sign language data in language resources and language processing contexts. In: LREC 2004: *Fourth International Conference on Language Resources and Evaluation Representation and Processing of Sign Languages Workshop.* Lisbon, Portugal, 24–30 May

2004. Paris: European Language Resources Association.

Hermans, D., Ormel, E., Knoors, H. and Verhoeven, L., 2008. The relationship between the reading and signing skills of deaf children in bilingual education programs. *Journal of Deaf Studies and Deaf Education,* 13(4), pp. 518-530.

Huenerfauth, M., Lu, P. and Rosenberg, A., 2011. Evaluating importance of facial expression in American Sign Language and pidgin signed English animations. In: ASSETS'11: *Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*. Dundee, UK, 22 – 24 October 2011. New York, NY: ACM.

Kalra P., Mangili A., Magnenat-Thalmann N. and Thalmann D., 1991. SMILE: A Multilayered Facial Animation System, In: IFIP WG 5.10*: Proceedings of International Federation of Information Processing.*Tokyo, Japan, 8-12 April 1991. Berlin, Germany: Springer.

Karpouzis, K., Caridakis, G., Fotinea, S.-E. and Efthimiou, E., 2007. Educational resources and implementation of a Greek sign language synthesis architecture. *Computers & Education.* 49(1), pp. 54-74.

Krnoul, Z., 2010. Correlation analysis of facial features and sign gestures. In: ICSP 2010: *Proceedings of the 2010 IEEE 10th International Conference on Signal Processing.* Beijing, China, 24–28 October 2010. Washington, DC: IEEE.

Lee, J. and Kunii, T., 1993. Computer animated visual translation from natural language to sign language. *The Journal of visualization and computer animation.* 4(2), pp. 63-68.

Liddell, S., 2003. *Grammar, Gesture, And Meaning in American Sign Language.* Cambridge, UK: Cambridge University Press.

Lombardo, V., Battaglino, C., Damiario, R. and Nunnari, F., 2011. A avatar–based interface for Italian Sign Language. In: CISIS 2011: *Proceedings of the 2011 International Conference on Complex, Intelligent, and Software Intensive Systems.* Seoul, Korea, 30 June – 2 July 2011. Washington, DC: IEEE.

López-Colino, F. and Colás, J., 2012. Spanish Sign Language synthesis system. *Journal of Visual Languages and Computing.* 23(3), pp. 121-136.

Magnenat-Thalmann, N., Primeau, E. and Thalmann, D., 1987. Abstract Muscle Action Procedures for Human Face Animation. *The Visual Computer*. 3(5), pp. 290-297.

Miranda, J.C., Alvarez, X., Orvalho, J., Gutierrez, D., Sousa, A. and Orvalho, V., 2012. Sketch express: A sketching interface for facial animation. *Computers & Graphics*, 36(6) , pp. 585-595.

Pandžić, I. and Forchheimer, R., 2003. *MPEG-4 Facial Animation: The Standard, Implementation And Applications*. Hoboken, NJ: Wiley.

Parke, F. and Waters, K., 1996. *Computer Facial Animation.* Wellesley, MA: A.K. Peters.

Rudser, S., 1988. Sign language instruction and its implications for the Deaf. In: M. Strong, ed. 1988. *Language Learning and Deafness.* New York: Cambridge University Press, pp. 99-112.

Schnepp, J., Wolfe, R., Shiver, B., McDonald, J. and Toro, J., 2011. SignQUOTE: A remote testing facility for eliciting signed qualitative feedback. In SLTAT 2011: *Proceedings of the Second International Workshop on Sign Language Translation and Avatar Technology.* Dundee, UK, 23 October 2011. Dundee: University of Dundee.

Schnepp, J. 2012. A representation of selected nonmanual signals in American Sign Language. Ph.D. DePaul University.

Weast, T., 2008*. Questions in American Sign Language: A quantitative analysis of raised and lowered eyebrows.* Ph.D. The University of Texas, Arlington.

Werner, S., Wolff, M. and Hoffman, R., 2006. Pronunciation variant selection for spontaneous speech synthesis: Listening effort as a quality parameter. In: IEEE ICASSP, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. Toulouse, France, 14-19 May 2006. Washington, DC: IEEE.

Zhao, L., Kipper, K., Schuler, W., Vogler, C. and Palmer, M., 2000. A machine translation system from English to American Sign Language. In: J.S. White, ed. 2000. *Envisioning Machine Translation in the Information Age.* Lecture Notes in Computer Science. 1934/2000 Berlin, Germany: Springer-Verlag. pp. 191-193.