Sign Language Subtitling

Nicoletta Adamo-Villani ∗
Purdue University

Gerardo Beni •
University of California Riverside

Abstract

The object of this paper is the development of a new method of sign language subtitling for motion pictures aimed at deaf children who cannot read English yet and can communicate only via signs. The method is based on the recently introduced concept of "semantroid" (an avatar limited to head and hands) and on the implementation of a new scrolling technique which allows for concurrent display of 4 subtitling windows at the bottom of the screen. To maximize the readability of the semantroid's face and hand configurations, we have created a new 3D model and we have shaded it with a combination of 3D and toon shaders. In addition, we have followed recent results in color perception to optimize the visibility of those body parts directly involved in the signing motion. The signing semantroid, optimized for maximum readability, can be scaled to fit in a very small area, and thus it is possible to display four captioning windows simultaneously. The concurrent display of several progressive animated signed sentences allows for review of the information, a feature not present in any sign language subtitling method presented so far. As an example of application, we have used the subtitling scrolling semantroid in a video of a chemistry experiment and we have tested its effectiveness with a group of hearing and non-hearing signers.

1 Introduction

According to the National Association of the Deaf, there are about 45,000 deaf school age (K-12) children in the United States [NSF report 1999]. Deaf children (who don't know how to read yet) don't have access to visual information (e.g. TV, DVDs, interactive media, etc.) with linguistic explanations. Linguistic explanation is given as speech for hearing children or as subtitles for non-hearing reading children [Captioning web 2004]. Considered that reading comprehension is significantly delayed in deaf Children (the median reading comprehension of seventeen and eighteen year old Deaf is at a fourth grade level [Holt, Traxler et al. 1997]), many young deaf children (grades K-3) are deprived of the opportunities for independent learning provided by visual media.

One solution to the problem is to use sign language as subtitles. The problem of "sign language subtitling" is complex if we consider the need to communicate in a restricted space and the methods available so far to represent signed communication.

Sign Language can be communicated by a human signer, a 2D or 3D avatar, or by SignWriting [Signwriting 1974]. Human signers and avatars are not easily scalable; in fact the readability of face and hands configurations is lost when the size of the image is drastically reduced. SignWriting can easily fit in a restricted space but it is a static, symbolic representation of the signs which requires learning and abstraction.

This paper focuses on the development of a new method of sign language subtitling. The method makes use of: (1) the recently developed concept of "semantroid" [Adamo-Villani and Beni 2005], and (2) a new scrolling technique that allows for the simultaneous display of 4 animated signed sentences at the bottom of the screen. In section 2 we describe: (a) the advantages of using the semantroid versus using a human signer or full-body avatar, (b) the semantroid model, (c) the difficulties with traditional subtitling methods, and (d) the new scrolling technique. In section 3 we provide an example of application of the method as a subtitling tool for an educational chemistry video. Conclusive remarks and future work are in section 4.

2 Method

The method we propose is based on the use of 'semantroids' [Adamo-Villani and Beni 2004]. The semantroid (from 'semantic' and 'android') is a reduced avatar (limited to head and hands) which maximizes the semantic content conveyed while minimizing the perceptual effort to perceive it. The concept has been quantified by the notion of 'semantic intensity' [Adamo-Villani and Beni 2005].

2.1 Advantages of the semantroid method

There are several advantages to using a semantroid versus using a human signer or a full-body avatar: (a) a semantroid fits in much smaller space for the same meaning expressed by either a human signer or an avatar; (b) it can be *toon-shaded* [Osmond and Thacker 2004] to improve semantic intensity and hence readability; (c) it represents naturally the signs and thus requires no learning/abstraction (in contrast with SignWriting [SignWriting 1974]). These advantages are discussed below after we review the difficulties with traditional methods.

2.2 The semantroid model

We have developed a new 3D model of the semantroid and a toon shader using Maya 6.0 software. In general, toon shaders give 3D surfaces a flat, 2 dimensional appearance by eliminating effects such as realistic shading and highlights. The application of the toon shader to those parts of the body (neck, ears, and hair) not relevant to the signing motion (see fig. 1) has allowed us to create a clear contrast between the 2D, stylized look of such body parts and the 3D, realistic look of hands and face thus maximizing the readability of hands and face configurations.

The rational for the use of the semantroid instead of full avatar has been given in [Adamo-Villani and Beni 2005]. The justification is based on comparing the semantic intensities of the

∗ nadamovi@purdue.edu
• beni@ee.ucr.edu

semantroid with the semantic intensity of the avatar. Similarly we can compare the semantic intensity of the semantroid with partial toon shading with the all-3D shaded semantroid. Since the only parts which are toon-shaded are the neck, the hair, and ears, the comparison is restricted to these parts. This is a special case of the general problem of choosing between 2D and 3D representations. In [Adamo-Villani and Beni 2005] it was suggested that semantic intensity calculations could be used to resolve such issues. Although the problem is still not resolved in general and no special cases have been considered yet, we have used the basic concepts of semantic intensity as guidance.

Basically, semantic intensity is a measure of the ratio of the quantity of 'meaning' conveyed to the quantity of 'effort' required to perceive such a meaning. 'Meaning' is closely related to information but is constrained by the requirement that the information must be represented directly (visually) without inference/abstraction/analysis on the part of the perceiver. 'Effort' is related to the perceptual effort of the observer in perceiving the meaning, and it is affected by two main factors:
(1) The spatial distribution of the image. The intuitive notion is that it takes more effort to perceive widely scattered elements than more compactly located elements.
(2) The distribution of the meaning-conveying property. A measure [Adamo-Villani and Beni 2005] of this effort is the ratio of the standard deviation of the possible meanings to the standard deviation of the nuances of meaning, i.e. the different values of the meaning-conveying property for a given meaning.
Semantic intensity is most effectively applied in cases of drastic data reduction, as when entire parts of a scene are removed (e.g., in going from avatar to semantroid). But the concept can be used also in less drastic cases of data reduction as in going from 3D to 2D in the toon-shading transformation.

In the case of the toon-shading of the hair, neck, and ears of the semantroid, there is no loss of meaning since all the meaning is conveyed by the hands and face. There is, however, a reduction in the perception effort. And of the two factors affecting effort mentioned above, only the second is relevant, as discussed below.

The properties that convey the meaning for the hair, neck, and ears may be taken to be the shape and the color. Assume that these are the only meaning conveying properties. The hair, neck, and ears shapes/colors change (very slightly) from one image to another but the change of shape/color does not represent a change of meaning. Thus the set or all the images of the neck (hair/ears) convey the same meaning of 'neck' (hair/ears). There is no effort required in distinguishing among possibly different meanings conveyed by the neck (or hair or ears) shapes/colors.

Thus, the only perceptual effort to consider is the effort of neglecting the variability in the shapes/colors of the neck (hair/ears). In fact, when more than one image may have the same meaning, the variation in the meaning-conveying properties (in this case shape and color) has an effect on the perceptual effort. Generally, in a set of images that should convey the same meaning, any variability of the meaning-conveying properties requires an additional perception effort. Therefore the more variety there is in the neck (hair/ears) shape and color, the more effort is made in attributing the same meaning to the neck (hair/ears).

So which type of image, 3D or toon-shaded, has more variability of shape/color? Without quantitative analysis it is rather intuitive to argue that the 3D case has more variability in color because of the variety of hues, levels of brightness and saturation caused by the shading. The 3D case has also more variety in shape if the images are segmented according to color and then the shapes of the various parts are considered separately. No such segmentation is possible in the toon-shader case.

Thus, a semantic intensity analysis, albeit qualitative, justifies the choice of toon shading vs. 3D for those parts of the semantroid which are not meant to convey a variety of meanings but simply convey 'static' notions, such as neck and hair; such notions being necessary only to support the meaning conveying parts , i.e., hands and face.

Besides the simplification via toon-shaders, the semantroid model has been 'optimized' for color. Again such 'optimization' is only qualitative and it is not meant to be definitive. The choice of colors has been guided also by semantic intensity considerations. Returning to factor (2) in the effort of meaning, it is clear that the effort of distinguishing among the various meanings conveyed by the hands and face is reduced by increasing the differences in color/shape of the possible configurations of the hands and face. Because of the need to maintain a minimum level of realism, the differences cannot be emphasized by modifying the shapes of fingers and facial features. On the other hand, color can be used to enhance contrast without loss of realism. This has determined our choice of a female model so that the nails can be colored without loss of realism. Bright colored nails makes it easier to discern the finger configurations especially on a reduced scale. Similar considerations apply to brightly colored lips which help discern different mouth configurations.

For the eyes, besides the brightness of the color, the choice of the hue is also a factor that affects the perception effort.
For the hue factor, we have followed research results in color perception [Hill and Scharff 1997]. These results were obtained for readability of websites with various foreground/ background color combinations, using different fonts and text sizes. The measure of readability was reaction time of the reader. No combination of font, size, or color was found to be optimal but for foreground/background the color combination with fastest reaction time was green on yellow out of 6 possible combinations which included yellow/blue. Based on these results, we have selected green rather than blue eyes, since the general hue of the face approaches a shade of yellow. Of course this choice is very tentative and based on results that may not be applicable; further studies should be done to determine the highest 'readability' of facial expression using different color combinations [Adamo Villani and Beni - in preparation].

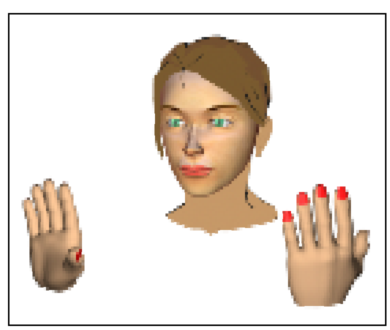

Figure 1. Partially toon-shaded semantroid

2.3 Difficulties with traditional methods

Traditional subtitling methods present several difficulties. First we consider methods which are easily scalable so that the subtitles can fit in a small portion of the screen. The requirement of fitting in a small area limits the subtitles to the use of static symbols. The most advanced of such systems is SignWriting [SignWriting 1974], developed by Valerie Sutton in 1974 and used worldwide for writing, reading and researching signed languages.

SignWriting has many uses but as a subtitling method has the disadvantage of being (1) static, as written English, and (2) highly abstracted so that it requires significant amount of learning. Because of these two factors, it is not easy to see the advantage over written English for subtitling. If effort has to be invested in teaching a system of abstract symbols, it might be more efficient to teach the Deaf how to read English and use English for subtitles. The purpose of subtitling in sign language is to communicate with persons whose only means of communication is sign language.

A more intuitive alternative to SignWriting could be the use of static images of signers as represented, e.g., in ASL dictionaries. But this method would require too many static images to follow in real time the messages communicated by voice. Even in using English subtitles, the speed is often not enough to keep up with the spoken word; and since using images for words requires a much larger screen space, it would be impossible to follow the spoken message by presenting such images as in comic strips at the bottom of the screen. To sum up, static images cannot be used because they must necessarily be small and as such must be abstract symbols which require a learning effort comparable to learning to read English.

Turning to methods that use dynamic (i.e., moving) signing, consider both human signers and avatars. In either case the first difficulty is the size. Can the problem be solved simply by cutting the signer at the waist so that the screen space occupied is sufficiently reduced ?

Consider first the human signer. For a human signer the aesthetic/emotional appeal is not easily controlled or changed by a simple menu in software. Human signers' appeal to different ages, genders, and ethnic groups varies and cannot be controlled. An avatar's appearance instead, can be easily modified in the user interface. Moreover, human signers, unlike avatars, cannot easily be made artificially emphatic without appearing ridiculous. Features that can be emphasized for enhanced communication include: nails, color and size of eyes, eyebrows, lips etc. These types of emphasis can be realized easily in an avatar but not in a human signer. A third difficulty with a human signer is that the background interferes unless the signer is clothed in black against a black background, as in a pantomime. But the dark background confuses the shadows at the edges of the hands and makes the gesture less clear than if the background were light. It is also in practice difficult to realize a very neutral darkly clothed signer on a dark background. Some details always remain and tend to stand up and be distracting. Another problem is present also in avatars and it is considered next.

The full bodied avatar, like the human signer, must be cut at the waist in order to fit in the restricted space at the bottom of the display. In doing so two problems arise. First the trunk is still visible and remains a distractive factor. Second, the hands are positioned at a natural distance from the head and thus in order to be included they require a significantly higher vertical size than the semantroid. The semantroid, in fact, positions the hands as close as possible to the head without significant loss of the meaning of the gesture. This would be tiring for a human signer and is not realized in a full bodied avatar since its purpose is to look as much as a human signer as possible. A comparison is shown in figure 2 where the semantroid image requires a 16.7% shorter vertical dimension. It is also clear from the figure that, if necessary, the semantroid vertical length can be reduced further by shifting the hands toward the head without major loss of meaning.

Figure 2. Size comparison between a full-body avatar cut at the waist and the semantroid (the image on the left is taken from [Vcom3D 2000])

2.4 The scrolling method

In addition to using the semantroid as a subtitling tool, we have devised an efficient scrolling method that allows for the simultaneous display of 4 animated signed sentences at the bottom of the screen throughout the entire video playback (see fig 3). This allows review -- a feature non existent in any sign language presentation method so far.

Traditional methods of captioning in sign language use one human signer occupying a corner of the screen (usually at the bottom right). This method has several drawbacks as discussed above. But, in addition, whether a human signer is used or an avatar, the traditional method displays the message as a sequence of signs so that only the last sign is visible. This is basically the method that must, by necessity, be used by a human signer in direct sign language conversation.

But, in displaying subtitles, there is no reason to limit the display to only the last sign which is being communicated. In fact, if the message is broken down in signed sentences, more than one sentence could be playing on the screen simultaneously. This method has both advantages and disadvantages.

It has the advantage that, similarly to written language, more than one sentence is readable at a given time. This allows for review of information. Without such review, it is very likely that, due to a lapse of attention, some meaning is lost. The same happens in listening to a message in comparison with reading it. Since the probability of missing one word or sentence is generally non-zero, in a long and complex message, listening is most likely to lead to some loss of meaning. In reading, such a loss of meaning due to lapse of attention is prevented by the possibility of re-reading the sentence which had not been attended to properly. In captioned sign language, even more so than in listening to a message, it is likely that attention will lapse. This is because the viewer must pay attention simultaneously to both the main scene on the screen and to the captioned signs. This situation is particularly severe when the viewer must pay attention to complex explanations, as, e.g., in learning about science.

Thus the possibility of re-reading signed sentences is very useful in general and especially in cases of complex messages. And this is the basic advantage of displaying simultaneously more than one signed sentence.

On the other hand, the method has the disadvantage of taking up a larger portion of the screen and thus interfering with the main scene. This is certainly the case if full body avatars or human signers were used. But even if the signers were shown only from the waist up, as discussed previously, showing multiple signed sentences still can disrupt the main scene significantly unless the signing display is considerably simplified This is achieved in our method by using the partially toon-shaded semantroid discussed above, which minimizes the screen area occupied, while maintaining a high level of signing clarity.

Even with the toon shaded semantroid, ultimately the number of signed sentences shown must be limited by the display area available. We have found that a reasonable compromise is to display 4 signed sentences simultaneously. This allows for recovering from lapses of attention while at the same time not occluding the main scene significantly. The method of displaying the sentences is chronological from right to left so that the method of reading the sentences is from left to right as in reading English. The scrolling method can easily be adapted to reading from right to left as in Hebrew or from top to bottom as in (some) Japanese writing. The latter case may, due to the screen aspect ratio, require that the sentences displayed be reduced to three.

3 Example of application and initial evaluation

As a practical example, we have applied the partially toon-shaded scrolling semantroid as a subtitling tool for an educational video of a chemistry experiment (the determination of pH of two common household products).The explanation of the experiment requires 8 signed sentences, thus 8 animated sequences:

1. Determining pH
2. You will need a bottle of vinegar
3. You will need a bottle of household detergent such as Mr. Clean
4. You will need two beakers containing a "Universal Indicator"
5. Pour the vinegar in beaker 1
6. Pour Mr. Clean in beaker 2
7. A color of red to yellow indicates a pH<7
8. A color of green to purple indicates a pH>7

An ASL (American Sign Language) signer has performed the above signs and the signing motion has been captured on videotape, blocked, imported into the 3D software, and used as a motion reference for the creation of the keyframed animations. The animations have been composited with the video footage and the timing of the scrolling motion has been synchronized with the explanations of the various phases of the experiment.

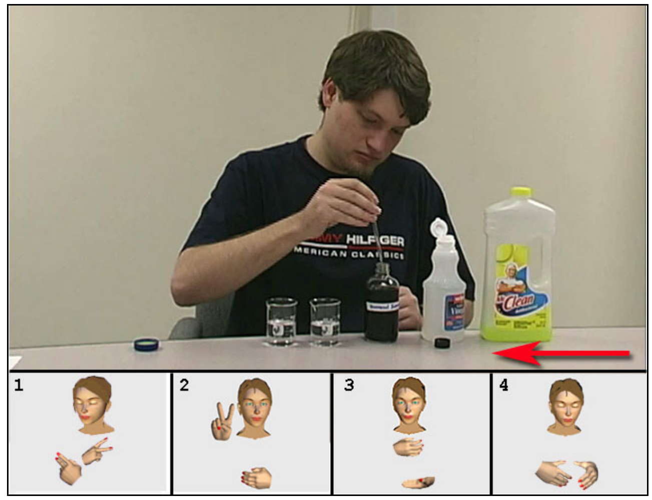

Figure 3. Video shot of the pH experiment with simultaneous display of 4 animated signing sequences. The arrow indicates the scrolling direction. The numbers indicate the order of appearance on the screen.

In Figure 3 the animations correspond to different steps of the experiment. In frame 1 the semantroid is signing "determining" in sentence 1, in frame 2 she is signing "V" of vinegar in sentence 2, in frame 3 she is signing "bottle" in sentence 3, and in frame 4 she is signing "beaker" in sentence 4. The three sections of the educational video with the subtitling scrolling semantroid and all animated clips are available at:
http://www.tech.purdue.edu/cgt/I3/ph_clips.htm

The subtitling scrolling semantroid has been evaluated throughout its development by deaf adults, Purdue faculty and students knowledgeable in sign language and deaf related issues who have provided positive feedback on the readability of the signs and the effectiveness of the method. Full-scale evaluation of the method with children age K-3 is currently being carried out in collaboration with the Indiana School for the Deaf (ISD), one of the leading institutions in Deaf Education. We are currently evaluating the method's success based on: (1) deaf children's reactions (emotional appeal of the model and willingness to use), and (2) teachers'/parents' feedback on the degree to which the subtitling scrolling semantroid improves deaf children's understanding of science concepts presented on video.

4 Conclusion and future work

In this paper we have introduced a new method of subtitling motion pictures for deaf persons who cannot read written languages and can only communicate via a sign language. The method is based on the use of semantroids [Adamo-Villani and Beni 2004] which are 3D avatars reduced to hands and head only in order to optimize the semantic intensity [Adamo-Villani and Beni 2005] of the images. Because of the small size of the semantroids, it is possible to fit four windows at the bottom of the screen where subtitles normally appear. Each window displays a movie of one signed sentence; hence the subtitling windows display simultaneously 4 sentences. The windows are scrolled from right to left so that reading proceeds from left to right as in English (the order can be reversed for other languages).

In order to optimize the visibility of the signing semantroid we have developed a new model. We have used recent results on semantic intensity and on color contrast readability. Based on semantic intensity [Adamo-Villani and Beni 2005] considerations, we have simplified further the semantroid model by toon-shading hair, neck, and ears, three parts which don't play an essential role in conveying meaning. Based on recent results [Hill and Scharff 1997] on color contrast readability we have produced a model that, while maintaining realism, has salient features with colors that make them easily detectable on a small scale.

This being a first step in proposing the semantroid subtitling method, many aspects need to be tested and improved. The optimization of the model is still semi-quantitative. Further studies on the readability of gestures and facial expressions as a function of color, shape, and shade parameters are required. Limits of scalability to smaller images also need to be quantified. Optimal number of simultaneous signed sentences is also a factor to analyze. Could be better to have only 3 sentences but in larger size? Because of the aspect ratio of the screen, the 3 sentences could be displayed vertically. Would this affect the readability? Would this be an advantage for persons who eventually will learn to read oriental languages which are written from top to bottom?

Although many questions remain to be answered, this method of subtitling in sign language helps bridging the gap between two rather different views on presenting information to the Deaf. The two views are at opposite ends in the scale of abstraction. On one hand (more concrete) we have human signers, on the other (more abstract) we have symbolic representations of signs. In between we have the method of subtitling by semantroid, which may be regarded, in fact, as intermediate between subtitling realized via a human signer (or a signing avatar) [Adamo-Villani et al. 2005], and subtitling realized with an abstract system of sign representation, such as SignWriting [Signwriting 1974]. The table below shows the relation of the present method to the other two.

	Level of abstraction	**Motion**	**Size**
Human signer or avatar	Low (concrete)	yes	large
Semantroid	Medium (parts removed + toon shading)	yes	small
Sign Writing	Very high (symbolic)	no	small

Table 1.

Thus table 1 shows that the subtitling semantroid fills a niche left by the two extreme cases used so far to help the Deaf understand explanations of visual information.

Acknowledgements

This research is partially supported by the School of Technology at Purdue University (I3 grant - Proposal #00006585 - http://www.tech.purdue.edu/cgt/I3/), by the Envision Center for Data Perceptualization, and by gh, LLC Inc., IN. We thank Professor Ronnie Wilbur from the Department of Audiology and Speech Sciences at Purdue University for her valuable contributions to the project. We also thank Marie Nadolske for performing the signs and for her continuous help and feedback on the project. In addition, we are grateful to all the signers who have participated in the evaluation of the method and to the Indiana School for the Deaf (ISD) for providing the testing ground for a thorough evaluation of the ASL subtitling technique.

References

ADAMO-VILLANI, N. and BENI, G. 2004. Keyboard Controlled Signing Semantroid . In *IEEE Proceedings of ICARCV- 8th International Conference on Control, Automation, Robotics and Vision,* China, 741-746.

ADAMO-VILLANI, N. and BENI, G. 2005. Semantic Intensity: a Measure of Visual Relevance. *International Journal of Human-Computer Studies* (under review).

ADAMO-VILLANI, N., DOUBLESTEIN, J., MARTIN Z. 2005. "Sign Language for K-8 Mathematics by 3D Interactive Animation". *Journal of Educational Technology Systems,* 33, 3.

CAPTIONING WEB available at:
http://www.captions.org/

HILL , A. and SCHARFF, L. V. 1997. Readability of websites with various foreground/background color combinations, font types and word styles. In *Proceedings of the Eleventh National Conference in Undergraduate Research* 2, 742-746.

HOLT, J. A., TRAXLER, C. B., ALLEN, T. E. 1997. Interpreting the Scores: A User's Guide to the 9th Edition Stanford Achievement Test for Educators of Deaf and Hard-of-Hearing Students. Washington, D.C., Gallaudet Research Institute.

OSMOND, A. and THACKER, J. 2004. 3D into 2D: Inside Story. *3D WORLD,* 28.

Preparing our children: Math and science education in the national interest 1999. Washington, DC, National Science Foundation, National Science Task Force on Mathematics and Science Achievement.

SIGNWRITING available at:
http://www.signwriting.org/about/what/what02.html

VCOM3D available at:
http://www.vcom3d.com/