State of the Art and Future Challenges of the Portrayal of Facial Nonmanual Signals by Signing Avatar

Chapter · July 2021

DOI: 10.1007/978-3-030-78092-0_45

10 authors, including:

Rosalee Wolfe
DePaul University
101 PUBLICATIONS **552** CITATIONS

John Mcdonald
DePaul University
61 PUBLICATIONS **439** CITATIONS

Ronan Johnson
DePaul University
2 PUBLICATIONS **1** CITATION

Robyn Moncrief
DePaul University
15 PUBLICATIONS **57** CITATIONS

State of the Art and Future Challenges of the Portrayal of Facial Nonmanual Signals by Signing Avatar

Rosalee Wolfe[1,2](✉), John McDonald[1](✉), Ronan Johnson[1], Robyn Moncrief[1], Andrew Alexander[1], Ben Sturr[1], Sydney Klinghoffer[1], Fiona Conneely[1], Maria Saenz[1], and Shatabdi Choudhry[1]

[1] DePaul University, Chicago, USA
{rwolfe,sjohn165,rkelley5,bsturr,sklingho,msaenz, schoud12}@depaul.edu, jmcdonald@cs.depaul.eduf
[2] Institute of Language and Speech Processing, Athens, Greece

AQ1

Abstract. Researchers have been developing avatars to portray sign languages as a necessary component of automatic translation systems between signed and spoken languages. Although sign language avatar technology has improved significantly in recent years, there are still open questions as to how best portray the linguistic and paralinguistic information that occurs on a signer's face. Three interdisciplinary themes influence the current state of the art. The first, linguistic discovery, defines the facial activity that an avatar must carry out. The second, Computer Generated Imagery (CGI), supplies the tools and technology required to build avatars, and which determines the fidelity of an avatar's appearance. In contrast, the third theme, Sign Language Representation Systems, determines the fidelity of timing of facial co-occurrences. This paper discusses the current state of the art and demonstrates how these themes contribute to the overall goal of creating avatars that can produce legible signed utterances that are acceptable to viewers.

AQ2

Keywords: Sign language · Nonmanual signals · Avatars

1 Introduction

For the past 20 years, researchers have been developing avatars to portray sign languages. The goal of these signing avatars is to display signed languages as 3D animation, in lieu of displaying video recordings of human signers. The appeal of signing avatars is in their flexibility and consistency. It is easier to change or add a sign to an utterance when using an avatar than it is to change or add a sign to a previously recorded video. Further, when a project requires repeated production sessions over a period of several weeks or months, it is easier to maintain presentation consistency in the lighting, camera angle, clothing, and hair length of an avatar than it is when recording human signers.

Signing avatars are also a necessary component of automatic translation systems. In situations where the interaction is highly predictable but an interpreter will never be

M. Antona and C. Stephanidis (Eds.): HCII 2021, LNCS 12768, pp. 1–17, 2021.
https://doi.org/10.1007/978-3-030-78092-0_45

available, a signing avatar can work in conjunction with an automatic translation system to provide rudimentary communication. Prototypes have translated weather reports [1, 2], facilitated interactions with a post office clerk [3] and airport security personnel [4], and created Deaf-accessible public address systems [5].

Although avatar technology has improved significantly in recent years, there are still open questions about how best to display the linguistic and pragmatic information that occurs on a signer's face. The focus of this paper is a discussion of the potential for linguistics and computer graphics to work together to portray facial nonmanual signals effectively through signing avatar technology.

2 Three Themes

An in-depth understanding of the state of the art requires examining several multidisciplinary themes. The first theme is the linguistic discovery of the purpose and properties of facial nonmanual signals, and the second is a visual recounting of the developments in computer generated imagery (CGI) that make the computer display of signing possible. The third theme is sign language representation systems which direct an avatar's face to produce nonmanual signals dictated by linguistics. Combining the best practices of these three areas will contribute to an improved clarity in avatar signing, which will lead to a greater acceptance of avatar portrayal of nonmanual signals as perceived by the target users.

These three themes provide a context for considering current challenges in portraying facial nonmanual signals, which currently lag the larger manual motions of signing in current avatars. These challenges include the ramifications of choosing a cartoon style versus human realism, improving avatar motion, acquiring finer detail in corpora for more in-depth study of facial detail, adapting an avatar to produce multiple sign languages and considering legibility on different display devices.

Although there are also nonmanuals that occur on the body, this paper will focus on the linguistic discoveries involving only the facial articulators. It omits nonmanual signals arising from the body (shifts, leans), the head (tilt, nod, shake, turn), and teeth.

2.1 Linguistic Discovery

Without the insights into the structure and meaning of signed languages, it would be impossible to create avatar technology with sufficient flexibility to produce animations that would be perceived as intelligible signed language. As a discipline, sign language linguistics has evolved rapidly. Less than twenty years after the pioneering efforts of Stokoe [6] which focused mainly on a signer's hands, Baker's [7] extensive study of American Sign Language (ASL) showed that a signer's face conveys more than emotion. She observed that although affect can alter the form of a syntactic signal, the signal is still readily perceivable.

In later work, Baker noted that syntactic constituents primarily occur on the upper half of the face, including topic marking, yes/no, wh-, and rhetorical questions [8]. Consistent with this finding, subsequent researchers found that signers use eye gaze to create synatic agreement by marking referents [9, 10] and they use eye blinks to mark prosodic as well as syntactic phrases [11].

In her 1983 work, Baker also noted that activity on the lower part of a signer's face tends to modify individual signs or phrases and to convey adjectival or adverbial information. An example is the use of pursed lips to convey that a surface is smooth [12]. Facial nonmanual signals can also operate on the phonemic level as seen in ASL when the presence of the nonmanual 'th' changes the lexical item LATE to NOT-YET [13].

Another facial activity occuring on the lower face is *mouthing,* which is derived from words in the circumambient spoken language. Linguists recognize mouthing as part of sign languages found in the Netherlands, the United Kingdom, Sweden, Germany, France, and the German-speaking regions of Switzerland [14–17].

This short survey of linguistic discoveries demonstrates that facial articulators can convey information at all linguistic levels, ranging from the prosodic down to the phonemic. In addition, pragmatic information can co-occur with linguistic information. Figure 1 is a diagram that summarizes the participation of facial articulators in producing linguistic and paralinguistic information.

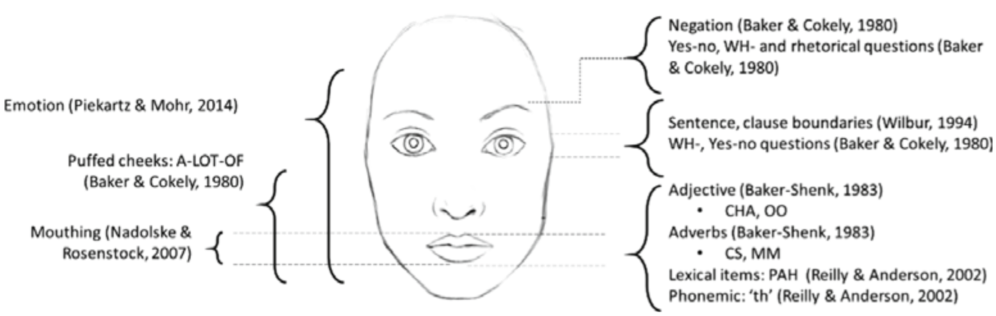

Fig. 1. Facial articulators

The segmentation of linguistic processes to particular areas of the face is a general characterization, but not a strict classification. Even though a linguistic phenomenon may be categorized as occurring on a particular facial feature, other features can participate in its production, but to a lesser degree. Figure 2 demonstrates how brow height and eye aperture can change when a signer produces the ASL adverbial modifier CS on the lower part of the face.

Another counterexample to the rule of thumb that "the upper face is used for syntactic and prosodic processes" is the role that the upper face plays in producing the ASL nonmanual VERY SMOOTH. Figure 3 demonstrates how a brow lowering and an eye squint can intensify the nonmanual adjective SMOOTH [18].

Further, facial nonmanual signals can co-occur and each can have an influence on a single facial feature. Consider a scenario where a signer poses a yes/no question about whether a surface is very smooth but does so in a concerned manner. In this case, affect, syntax and an adverbial modifier will have an influence on the brows. Lastly, these co-occurring facial nonmanual signals can have different spans and varying

Fig. 2. Comparison of a neutral face and the nonmanual marker CS. Note that the upper face participates in the production of the adverbial modifier [19].

intensities. To accommodate this complexity Wilbur [20] proposes a layering of prosodic and phonological events.

Because linguistics forms the basis of the software specifications for an avatar, there are several takeaways to keep in mind when developing an avatar capable of facial nonmanual expressions:

- Facial articulators can convey information at all linguistic levels.
- A single facial articulator can be responsible for conveying multiple co-occurring nonmanual signals.
- Co-occurring facial nonmanual signals can have different onsets, durations, and intensity envelopes.

2.2 Advancements in Computer Generated Imagery (CGI)

The limitations of Computer Generated Imagery (CGI) pose challenges to creating facial nonmanual signals that are correct, believable and acceptable. One of the contributing barriers is the difficulty of portraying a human face that is convincing in appearance and which moves in a natural, lifelike manner.

Researchers are making excellent progress in creating an appealing and realistic rendering of human skin and cartilage [21], but the most efficient and effective representation of the underlying facial musculature is still an open question. Facial musculature varies greatly among individuals and some facial muscles are not present in all humans.

CGI depicting human figures began with primitive representations. The first recording of a computer-generated animation depicting human figures was *A Computer Generated Ballet* [22]. Figure 4 shows a frame from the movie. The dancers were simplistic stick figures drawn using white lines on a dark background.

Fig. 3. A lowered brow and squinted eyes intensity the nonmanual SMOOTH to convey VERY_SMOOTH.

It would be another twenty years twenty years before animators and developers would tackle the challenge of portraying emotions through facial expressions to support a storyline. *Tony de Peltrie* (Fig. 5) features an aging lounge pianist. His entire face reflects his thoughts and moods as he wistfully remembers the success of his youth [23].

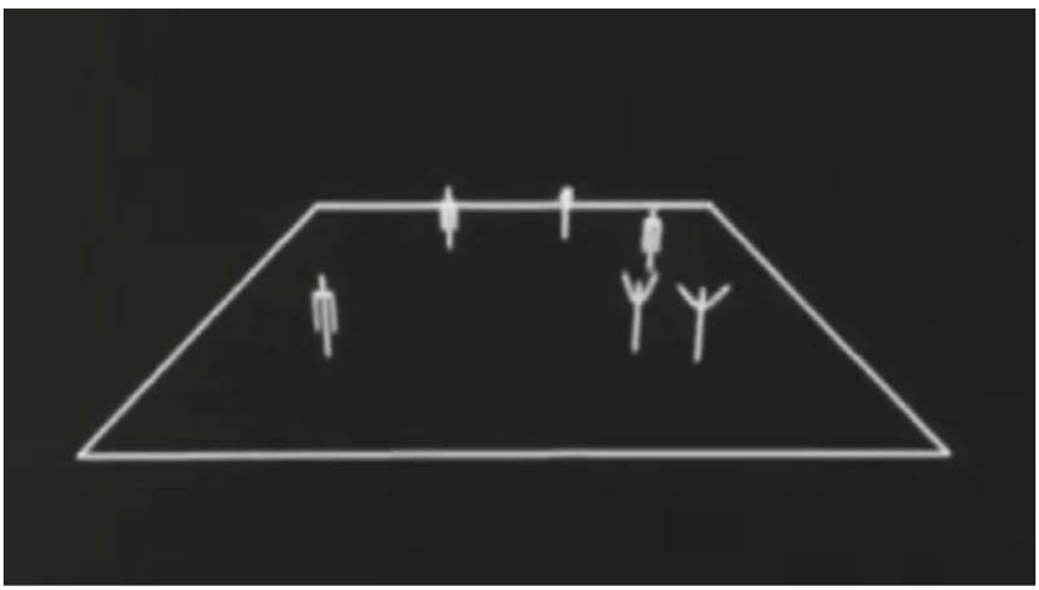

Fig. 4. *A Computer Generated Ballet* [22].

With this expressivity came greater complexity. Whereas the stick figures of Noll's *Ballet* used less than fifty line segments, *Tony* required thousands. It took three years to create the 7.5 min film.

Fig. 5. Tony de Peltrie [23].

Although *Tony* won numerous international prizes for innovations in computer animation and was hailed as groundbreaking, in recent years, younger viewers have often expressed an aversion to the Tony character. This is an example of the phenomenon known as the *uncanny valley* [24].

The robotics researcher Mori proposed that a person's reaction to a robot would change from empathy to aversion as its face approached, but failed to obtain, a lifelike appearance. He described this change as the uncanny valley, as seen in Fig. 6. This effect is intensified if the robot is in motion. If it appeared eerie in a still pose, it will appear even eerier in motion. Researchers have since found that this relationship also applies to computer generated human characters as well.

This is the reason that the first computer-generated movies deliberately portrayed eerie or alien humanoids. The pseudopod character in *The Abyss* (1989), capitalized on the uncanny valley. Another approach is to deliberately choose to give characters a cartoon-like look, as in as in *Toy Story* (1995) and *Avatar* (2009). Since their appearance is less human-like, they avoid the uncanny valley, and viewers are more accepting of their motion.

Although there was much enthusiasm in the early days of CGI, the reality is that creating animations takes a tremendous amount of manual labor, and the results aren't always that great to look at. Since the advent of CGI, two avenues of research have attempted to ameliorate the problem. One of them reduces the amount of manual labor required and the second attempts produce better visual effects. Several approaches to reducing manual labor include using blend shapes, a method of using static facial poses (the shapes) that are blended together to create motion moving from pose to pose [25], Improved 3D scanning techniques to record the shape of character faces and facial movement through *motion capture* now permit artists to combine the face of one virtual actor with the movements of another [26].

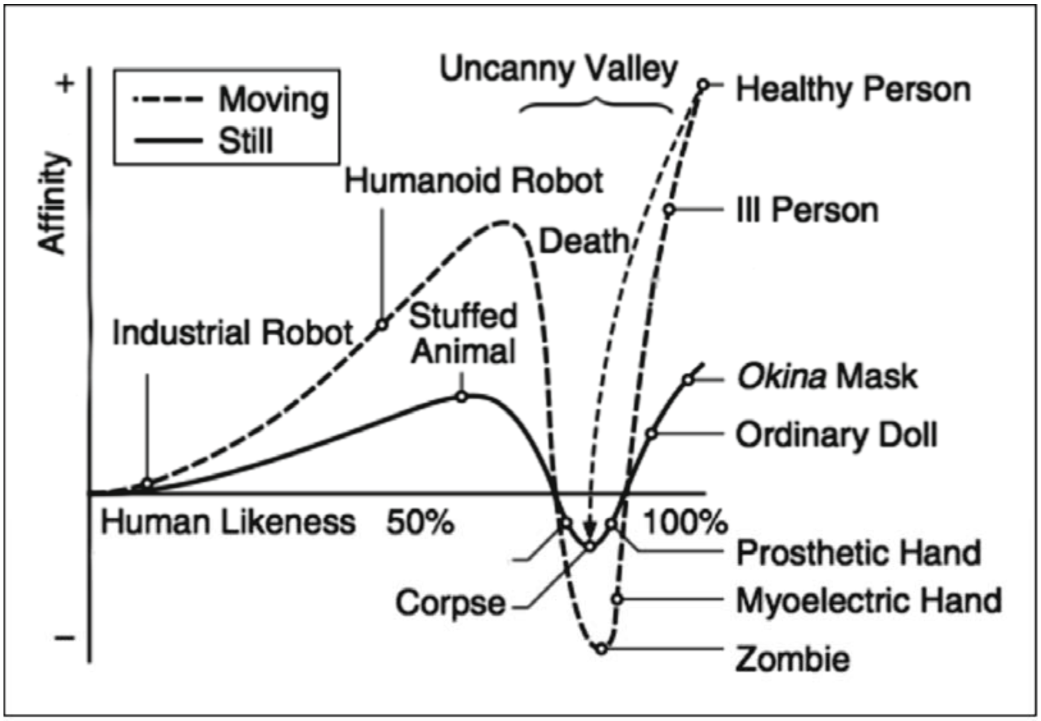

Fig. 6. The uncanny value is intensified by movement. After [24].

Other time-savers involve building libraries of frequently used gestures for repeated use in animation. One example is automatic lip sync. Manually animating lips to move when a character speaks is time consuming. First, artists create a library of *visemes*. A viseme is a shape that lips take when a speaker is producing a particular phoneme. See Fig. 7 for an example set of visemes for spoken English. Software deconstructs the character dialog into phonemes and chooses a corresponding viseme from the library for portrayal in the animation [27].

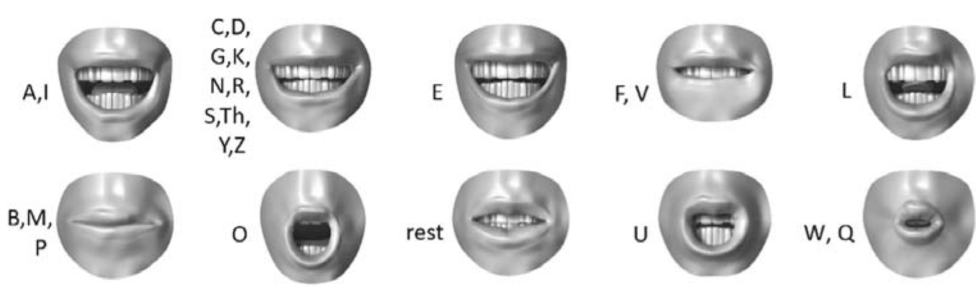

Fig. 7. Visemes and corresponding phonemes commonly used in automatic lip sync of spoken English. After Preston Blair [28].

As a result of the introduction of these time-saving techniques and the advent of ever-faster computers, movie effects have become increasingly realistic. Figure 8 shows the first extreme close-up of a CGI face, which appeared in the "Super Burly Brawl" scene in *The Matrix Revolutions* when the character Smith sustains a violent face punch [29].

A question naturally arises, "If movie effects are so realistic, why are avatars so unrealistic?" The answer lies in the time it takes to create the movies from the artist's work. Although many time-saving techniques have made it quicker for artists to animate, render times are still a bottleneck. Quoting Craig Good, a digital artist who worked at Pixar [30]:

> There's something I call The Law of Constancy of Pain: Back in 1983 it took between half an hour and around 8 hours to render a frame… Today, computers are literally millions of times more powerful. And guess how long it takes Pixar to render a frame [today]? Yup. Between half an hour and around 8 hours, with a typical average of a couple or three hours. Rendering time has stayed essentially flat for three decades.

Fig. 8. "Super Burly Brawl" The Matrix Revolutions [29].

Such long rendering times are incompatible with interactive graphics applications, such as video games and signing avatar technology which must be rendered at a rate of at least 25 frames per second. Interactive graphics must respond instantly to user commands. Compare Fig. 9, a frame from the "Super Burly Brawl" scene of the video game *Matrix Online 2005* and Fig. 8, a frame from the movie. Figure 9 contains more primitive characters in a simpler environment. These reductions in realism result in greater interactivity.

Fig. 9. "Super Burly Brawl" *Matrix Online* video game released in 2005 [31].

The discussion of CGI has revealed the following insights into developing avatar technology:

- Avoid the uncanny valley,
- Avatars need the same level of responsiveness found in video games,
- Avatars also need the realism of CGI in movies,
- Creating animation is still an expensive process, even with automation.

2.3 Sign Language Representations

The third theme considers alternatives in representation which can direct an avatar's face to produce nonmanual signals. Multiple approaches to representing sign language have been explored in the past twenty years, varying degrees of success. Categories of these approaches include.

1. general-purpose representations adapted for sign languages
2. sign language notation and annotation systems
3. sign language representations supporting prosody.

A representation used in early analyses of facial nonmanual signals is the Facial Action Coding System (FACS), which records movements of individual muscles on a face from visually observing small instantaneous changes that occur [32]. Although used extensively in psychology and computer vision research, FACS could not encode the motivation for facial movements. For example, although FACS can effectively represent raised brows, it does not facilitate recording *why* the raise occurred. It may be due to syntactic or affective reasons or both [33].

Another general-purpose representation is MPEG-4 H-Anim, the Humanoid Animation International Standard [34]. It facilitates the display of 3D avatars in web browsers on any device, which means that applications using this representation will work on any computer or smart phone running an H-Anim enabled browser. H-Anim's Facial Animation Parameters dictate an avatar's facial movement. See Fig. 10 for a list of facial feature points in the representation.

Unfortunately, H-Anim is a coarse representation that does not capture important facial details. Figure 10 shows that each eye lid only has one control to open or close it. This suffices for animating an eye blink but is insufficient even to represent basic emotions. On the left side Fig. 11 is an eye of an angry face and on the right is the eye of a sad face. In each case, the upper lid forms a complex curve. A single lid control is not be able to define a curve.

The second category mentioned in the beginning of this section is sign language notation and annotation systems. Designed by researchers with in-depth knowledge of sign languages, these systems have the advantage recording linguistic intent. The earliest of these was Stokoe notation, the first phonemic notation for signed language [35]. Created for ASL, it provided for the notation of handshape (dez), location (tab), and movement (sig), but it did not provide for nonmanual signals.

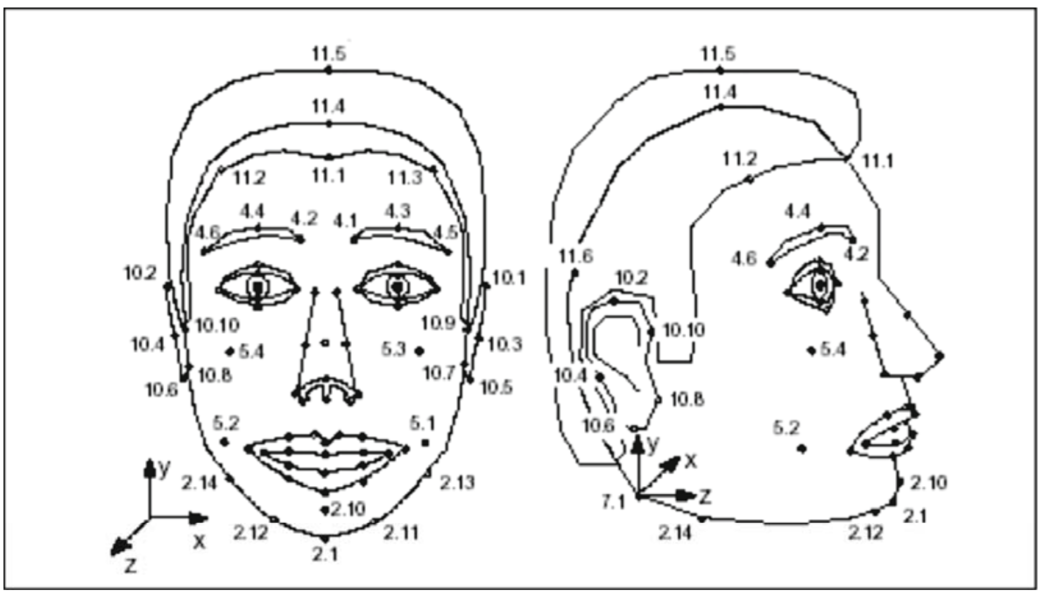

Fig. 10. H-Anim feature points

In contrast, HamNoSys (the Hamburg Notation System) can specify handshape, palm orientation, location, movement and nonmanual signals [36]. See Fig. 12 for an example. It is a phonetic system, so it is not limited to a particular sign language language. In fact, the Dicta-Sign project [37] used HamNoSys to create corpora in German, French, British and Greek sign languages.

One advantage of notation systems is that they are amenable to automatic analysis through statistical tools. However, these notation systems can express only sparse information about the timing of facial events within a sign or throughout a sentence. They do not effectively specify timing of co-occurrences.

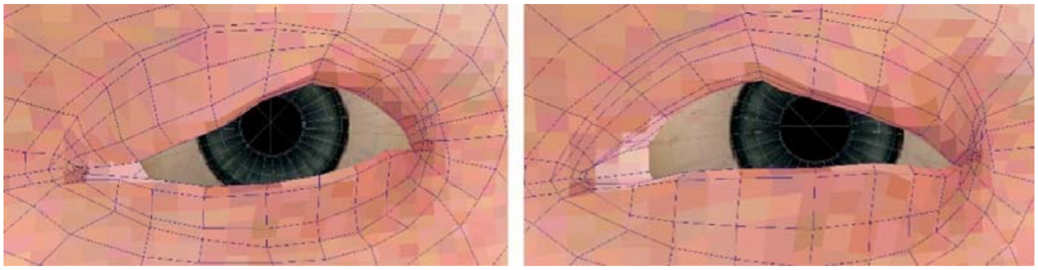

Fig. 11. Left: Angry eye. Right: Sad eye. Courtesy of Ronan Johnson

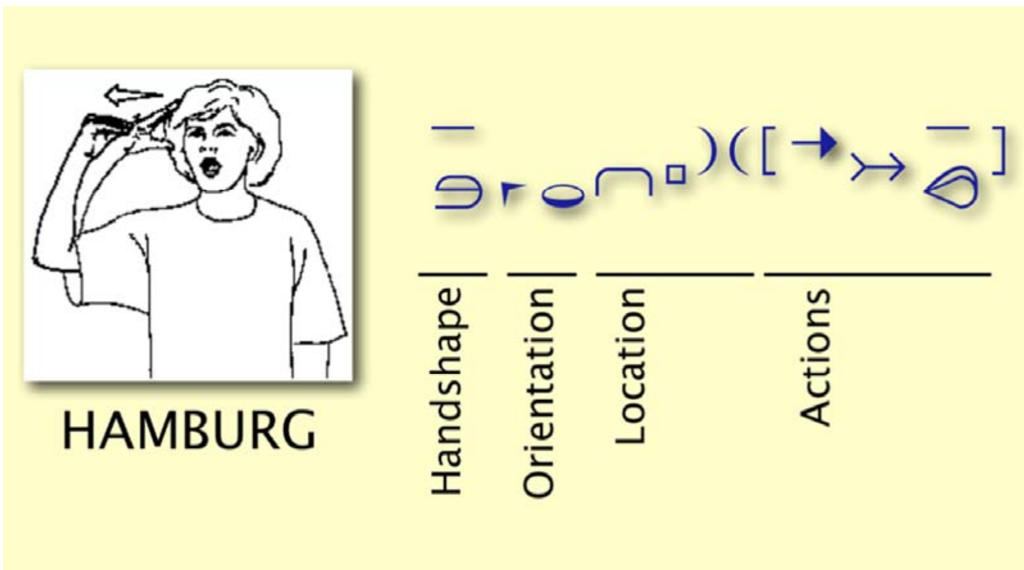

Fig. 12. Annotating the DGS sign HAMBURG: sign sketch and HamNoSys notation

Linguists have posited that are as many as fourteen channels of activity – on the head, hands and arms – where behaviors could co-occur [12]. Most of these occur on the face. Producing the onset, duration and intensity envelope, of each facial event is essential for clear communication via avatar.

Two useful annotation systems that can record co-occurrences are iLex [38] and ELAN [39]. Both systems support time-based annotations of video. They allow for the definition of tiers that correspond to the linguistic channels identified by researchers. Figure 13 is a partial screenshot from an ELAN file which contains multiple tiers representing nonmanual signals. There are many instances of co-occurrence in this short example.

The last category of representations support prosody. Knowledge of prosody enables avatars to produce natural, easy-to-read signed sentences. Without it, the animated signing "will be as unacceptable and potentially as difficult to understand as robotic speech lacking cues to phrasing, stress and intonation." [41].

Although Wilbur had previously identified several prosodic elements that are predictable enough for automation, many nonmanual signals are not as predictable and require additional information. To accomplish this, Adamo-Villani and Wilbur created

Fig. 13. An ELAN annotation with 14 tiers [40].

the representation ASL-Pro (ASL with prosody). The representation is time-based, similar to ELAN or iLex, but the annotations contain sufficient geometric information that an avatar can produce more lifelike motion.

A second representation supporting prosody is a hybrid approach [42] that relies on a mixture of traditional and procedural animation to build the basic elements for a discourse produced by avatar. A hierarchical description called AZee provides the coordination and timing of the animation data to create co-occurrences [43]. The goal is to build directly from linguistic descriptions rather than sequences of individual phonemes. Because it works with larger segments of the discourse, it achieves more natural animation.

Researchers have a wide variety of alternatives for representing sign languages for avatar display. When choosing, the following questions may prove useful:

– Generality of representation: Is it one that is project specific, or does it lend itself to data reuse for additional linguistic research?
– Availability of supporting software: Are there facilities to input, validate, store, retrieve, analyze, and display signed utterances?
– Level of detail of the representation: Does the representation facilitate linguistic abstraction or details of the timing and coordination of facial movements, or both?
– Acceptance within the research community: is the representation prevalent enough that there is a community to discuss challenges and future directions?

3 State of the Art

From the above description of the three themes – linguistic discovery, the limitations of CGI and the suitability of sign language representations – it is clear that substantial challenges remain. Recall that the goal is to create avatars that can portray facial nonmanual signals sufficiently clearly and correctly that the utterances they produce are acceptable to viewers.

All current avatars remain works-in-progress. Although there are avatars whose appearance is appealing and there are avatar systems that offer different characters, no avatar system to date can create convincing motion by drawing from a library of signs to synthesize the full range of new signed utterances. To be sure, there are many avatars that can display previously animated utterances but no one system can.

- retrieve lexical items from a database,
- modify their motion to add adverbial or adjectival modifiers,
- designate the characters in constructed dialog,
- respect the prosody,
- but still paying attention to the co-occurrence of facial articulators.

All of these are necessary for viewer comprehension and acceptance.

In the end, the avatar must be judged by the naturalness of its communication, the comprehensibility of the signing it generates and its acceptance by the Deaf community. To date, there has been little in the way of evaluations by end users of avatar comprehensibility [44]. The best published results from commercial efforts to date put comprehension rates at 52% [45]. This comprehension rate is not sufficient for effective communication.

The acceptability issue for current avatars is a tradeoff between the ability to generate new utterances and the naturalness of resulting motion. Motion capture systems, that supply some of the most natural human motion for pre-recorded signing, produce motion that is very difficult to edit and re-combine. Pre-recorded hand animation also produces very natural motion and is easier to edit but suffers from the high cost of animator time. On the other hand, synthesis driven directly from linguistic specification via mathematical procedural models, while very flexible, produces highly robotic motion. Hybrid systems that try to combine the naturalness of pre-recorded motion with the flexibility of procedural animation fall in between these two extremes.

4 Current Challenges and Future Directions

Several research directions will aid in creating an avatar that will produce credible facial nonmanual signals. The first direction is toward *integration.* Building on current work will achieve a new representation that will accommodate any sign language at any level of detail, ranging from the sublinguistic to the linguistic to the paralinguistic through the use of multiple channels. Researchers will still have the option of annotating the linguistic events that are pertinent to their study, but their data will be open to further modifications that add or refine annotations. Representing multiple channels is essential for specifying events that co-occur but do not coincide. All of the channels contribute to building message clarity for a viewer. Without this data, an avatar would be limited to producing motion based on developer's heuristics or an animator's best estimate.

The sublinguistic level must be part of the integration efforts because it is essential to legibility and acceptability of the signing produced by an avatar. Developers of a new sign language representation will need to work closely with avatar developers to ensure that the multiple methods for characterizing motion mentioned in Sect. 3 are specified, including motion capture, manual animation, procedural modeling, and constraint specification. Each of these has its strengths and will be beneficial to an avatar system.

A second research direction is toward *flexibility.* Any sign language representation needs to be sufficiently flexible to accommodate new discoveries in linguistics. Further, the avatar representation needs the flexibility to adapt its appearance to the needs of the audience. These can be surface changes such as modifications in the color of the hair, skin and clothing, or deeper changes such as accommodating diversity in ethnicity, age, and gender.

The last research direction is the development of evaluation techniques leading to standards. Most of the studies to date that have been shared with the research community involve such metrics as comprehensibility and appeal. These are common summative measures which are evaluated after a project has been completed. Since avatar technology is still a work-in-progress, developers need more *formative* feedback. Granted, much user feedback from a formative evaluation is qualitative which is more difficult to analyze.

Vital to creating a basis for formative evaluation is a way to generate the test stimuli (rendered animations) that an evaluation would require. Developing a set of *challenge data* for the avatar to perform would aid researchers carrying out formative evaluations. A set of challenge data could contain examples of facial nonmanual signals that have proved to be problematic in the past, such as lip coarticulation and eye aperture, as well as combining multiple channels of co-occurring linguistic events.

Developing challenge data and new methods for carrying out formative evaluation will aid in more rapid advancement in avatar technology. Such tools will save developers the work of creating test instruments "from scratch" and will provide a standard of comparison between systems.

5 Conclusion

Avatar technology is a new field, and the display of facial nonmanual signals is still an open question. The three themes that will support new innovations toward this goal are sign language linguistics, computer generated imagery and sign language representation. Possible directions of new research include the integration of current sign language representations to incorporate not only linguistic events, but sub-linguistic and paralinguistic events, and planning for sufficient flexibility to accommodate necessary changes based on linguistic discoveries and audience preference. Finally, developing challenge data and new evaluation techniques will help define the specification of new standards for avatar development.

References

1. Grieve-Smith, A.B.: English to American Sign Language machine translation of weather reports. In: Proceedings of the Second High Desert Student Conference in Linguistics (HDSL2), Albuquerque, NM (1999)

2. Verlinden, M., Zwitserlood, I., Frowein, H.: Multimedia with animated sign language for deaf learners. In: EdMedia+ Innovate Learning (2005)
3. Cox, S., et al.: The development and evaluation of a speech-to-sign translation system to assist transactions. Int. J. Hum.-Comput. Interact. **16**, 141–161 (2003)
4. Furst, J., Alkoby, K., Lancaster, G., McDonald, J., Wolfe, R.: Making airport security accessible to the deaf. In: Proceedings of the Fifth IASTED International Conference on Computer Graphics and Imaging, Kaua'i, HI (2002)
5. Ebling, S., Glauert, J.: Building a Swiss German Sign Language avatar with JASigning and evaluating it among the deaf community. Univ. Access Inf. Soc. **15**(4), 577–587 (2015)
6. Stokoe, W.C.: Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf (Studies in Linguistics, Occasional Papers 8). University of Buffalo, Buffalo, NY (1960)
7. Baker, C.: Focusing on the nonmanual components of American Sign Language. Understanding Language Through Sign Language Research (1978)
8. Baker-Shenk, C.: A Microanalysis of the Nonmanual Components of Questions in American Sign Language (1983)
9. Bellugi, U., Fischer, S.: A comparison of sign language and spoken language. Cognition **1**, 173–200 (1972)
10. Bahan, B.: Non-manual Realization of Agreement in American Sign Language (1997)
11. Wilbur, R.B.: Eyeblinks and ASL phrase structure. Sign Lang. Stud. **1084**(1), 221–240 (1994)
12. Wilbur, R.B.: Effects of varying rate of signing on ASL manual signs and nonmanual markers. Lang. Speech **52**, 245–285 (2009)
13. Reilly, I., Anderson, D.: FACES: the aquisition of non-manual morphology in ASL. Direct. Sign Lang. Acquisit. **2**, 159–182 (2002)
14. Crasborn, O.A., Van Der Kooij, E., Waters, D., Woll, B., Mesch, J.: Frequency distribution and spreading behavior of different types of mouth actions in three sign languages. Sign Lang. Linguist. **11**(1), 45–67 (2008)
15. Elliott, E.A., Jacobs, A.M.: Facial expressions, emotions, and sign languages. Front. Psychol. **4**, 115 (2013)
16. Sallandre, M.: Simultaneity in French sign language discourse. In: Amsterdam Studies in the Theory and History of Linguistic Science Series 4, vol. 281, p. 103 (2007)
17. Braem, P.B.: Functions of the mouthing component in the signing of deaf early and late learners of Swiss German Sign Language. In: Brentari, D. (ed.) Foreign Vocabulary in Sign Languages: A Cross-Linguistic Investigation of Word Formation, pp. 1–47. Erlbaum, Mahwah (2001)
18. Shumaker, C.: NMS Facial Expression, 2 Feb 2016. https://www.youtube.com/watch?v=NbbNwVwdfGg. [Accessed 15 Apr 2020]
19. Foster, H.: Non-Manual Markers in ASL/North Carolina Division of Services for the Deaf and Hard of Hearing, 12 Sept 2019. https://www.youtube.com/watch?v=8HIc0IRe-dE&feature=youtu.be
20. Wilbur, R.: Phonological and prosodic layering of nonmanuals in American Sign Language. In: Emmrey, K., Lane, H.L., Bellugi, U., Klima, E. (eds.) The Signs of Language Revisited: Festscrift for Ursula Bellugi and Edward Klima, pp. 213-241 (2000)
21. Nunes, A., Maciel, A., Meyer, G., John, N., Baranoski, G., Walter, M.: Appearance modelling of living human tissues. Comput. Graph. Forum **38**(6), 43–65 (2019)
22. Noll, A.M.: Computer-generated Three-dimensional Movies. Bell Telephone Laboratories (1965)
23. Bergeron, P., Lachapelle, P., Langlois, D., Robidoux, P. (Directors): Tony de Peltrie. [Film] (1985)
24. Mori, M.: The uncanny valley. Energy **7**(4), 33–35 (1970)

25. Marmor, D.: Facial Rigging Blend Shape, 22 Oct 2011. http://www.cgfeedback.com/cgfeedback/showthread.php?t=2119. Accessed 30 Mar 2020
26. Edwards, G. (Director): Rogue One: A Star Wars Story [Film]. Lucasefilm; Walt Disney Pictures; Allison Shearmur Productions, USA (2016)
27. Xu, Y., Feng, A.W., Marsella, S., Shapiro, A.: A practical and configurable lip sync method for games. In: Proceedings of Motion on Games, pp. 131–140 (2013)
28. Martin, G.C.: Preston Blair phoneme series, 4 May 2018 http://www.garycmartin.com/mouth_shapes.html
29. Wachowski, L., Wachowski, L. (Directors): The Matrix Revolutions [Film]. Warner Bros, USA (2003)
30. Good, C.: When CGI artists say it takes "X" amount of time to render a frame, are they talking literally or figuratively? 11 Sep 2016. https://www.quora.com/When-CGI-artists-say-it-takes-X-amount-of-time-to-render-a-frame-are-they-talking-literally-or-figuratively
31. Royce, B: The Matrix is getting a fourth movie – so can we have The Matrix Online back now please?, 21 Aug 2019. https://massivelyop.com/2019/08/21/the-matrix-is-getting-a-fourth-movie-so-can-we-have-the-matrix-online-back-now-please/. Accessed 16 Apr 2020
32. Ekman, R.: What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford University Press, USA (1997)
33. Weast, T.P.: Questions in American Sign Language: A quantitative analysis of raised and lowered eyebrows. ProQuest, Ann Arbor, MI (2008)
34. Garchery, S., Boulic, R., Capin, T.: Standards for Virtual Human Animation: H-ANIM and MPEG4 (2004)
35. Stokoe, W.C., Casterline, D.C., Croneberg, C.G.: A Dictionary of American Sign Language on Linguistic Principles. Gallaudet College Press, Washington, DC (1965)
36. Hanke, T.: HamNoSys – Representing sign language data in language resources and language processing contexts. In: Fourth International Conference on Language Resources and Evaluation (LREC 2004). Representation and Processing of Sign Languages Workshop, Paris (2004)
37. Efthimiou, E., et al.: Dicta-sign–sign language recognition, generation and modelling: a research effort with applications in deaf communication. In: Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (2010)
38. Hanke, T., Storz, J.: iLex – A database tool for integrating sign language corpus linguistics and sign language lexicography. In: Workshop on the Representation and Processing of Sign Language, at the Sixth International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco (2008)
39. Brugman, H., Russel, A.: Annotating multi-media/multi-modal resources with ELAN. In: Proceedings of the 2nd Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios, Paris (2004)
40. Fabian Benitez-Quiroz, C., Gökgöz, K., Wilbur, R., Martinez, A.: Discriminant features and temporal structure of nonmanuals in American Sign Language. PLoS ONE **9**(2), e86268 (2014)
41. Adamo-Villani, N., Wilbur, R.B.: Asl-pro: American sign language animation with prosodic elements. In: International Conference on Universal Access in Human-Computer Interaction (2015)
42. Filhol, M., McDonald, J.: Extending the AZee-Paula shortcuts to enable natural proform synthesis. In: 8th Workshop on the Representation and Processing of Sign Languages, pp. 45–52 (2018)
43. Filhol, M., Hadjadj, M.N.: Juxtaposition as a form feature; syntax captured and explained rather than assumed and modelled. In: Language Resources and Evaluation Conference (LREC), Representation and Processing of Sign Languages, Portorož, Slovenia (2016)

44. Kipp, M., Nguyen, Q., Heloir, A., Matthes, S.: Assessing the deaf user perspective on sign language avatars. In: The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility. ACM (2011)
45. Pauser, S.: Prototypentest SiMAX im Rahmen des Innovationsschecks, 19 Mar 2019. https://www.equalizent.com/fileadmin/user_upload/News/2019_04_Avatar_Projektbericht.pdf

Author Queries

Chapter 45

Query Refs.	Details Required	Author's response
AQ1	This is to inform you that corresponding authors have been identified as per the information available in the Copyright form.	
AQ2	Please check the deletion of duplicate keyword 'Nonmanual signals'.	