Single-Person and Multi-Party 3D Visualizations for Nonverbal Communication Analysis

Michael Kipp, Levin Freiherr von Hollen, Michael Christopher Hrstka, and Franziska Zamponi
Augsburg University of Applied Sciences
An der Hochschule 1, 86161 Augsburg, Germany
E-mail: {firstname.lastname}@hs-augsburg.de

Abstract

The qualitative analysis of nonverbal communication is more and more relying on 3D recording technology. However, the human analysis of 3D data on a regular 2D screen can be challenging as 3D scenes are difficult to visually parse. To optimally exploit the full depth of the 3D data, we propose to enhance the 3D view with a number of visualizations that clarify spatial and conceptual relationships and add derived data like speed and angles. In this paper, we present visualizations for directional body motion, hand movement direction, gesture space location, and proxemic dimensions like interpersonal distance, movement and orientation. The proposed visualizations are available in the open source tool JMocap and are planned to be fully integrated into the ANVIL video annotation tool. The described techniques are intended to make annotation more efficient and reliable and may allow the discovery of entirely new phenomena.

Keywords: annotation tools, interaction analysis, human motion visualization

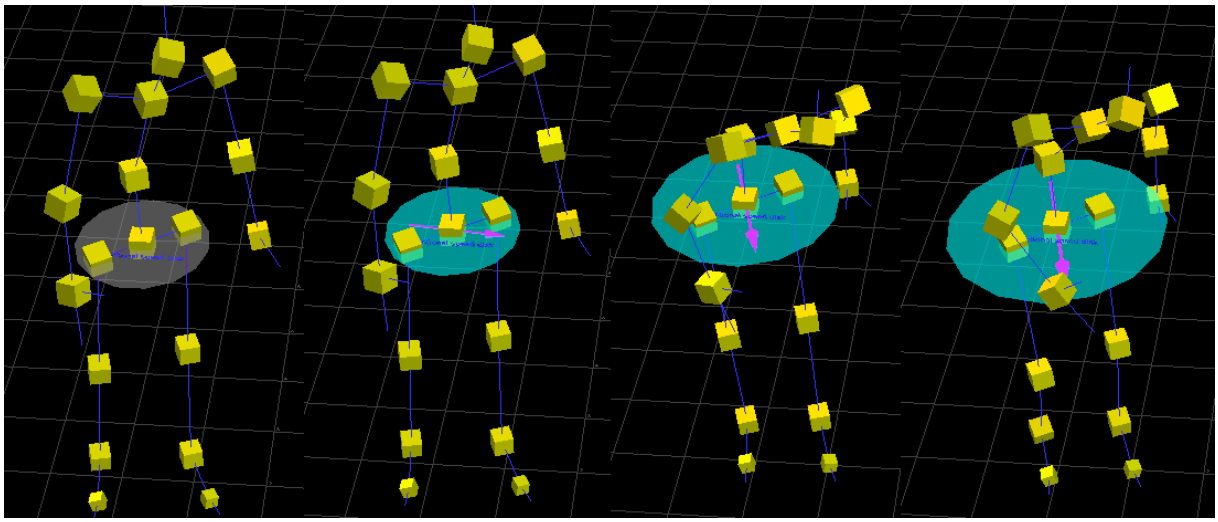

Figure 1: Direction of body motion is visualized by an arrow, the speed is shown by the size of a disc. The visualization is only active above a certain motion threshold to avoid visual clutter (grey = no motion).

1. Motivation

The most limiting factor when analyzing human subjects on video or in a 3D view, be it for gesture analysis (cf. McNeill, 1992) or interpersonal process analysis (cf. Bales, 1951), is the missing or hard-to-read depth information. While a 3D viewer allows to rotate the camera or view the scene from different angles simultaneously, we as human beings are optimized to watch a single 2D scene. Our visual system is neither made to parse true 3D information nor to integrate multiple views of the same scene. The fact that we have stereoscopic vision is only a minor enhancement to 2D vision. Therefore, the vast amount of data in 3D recordings cannot be trivially mapped to a human-readable visualization. The simultaneous viewing of multiple views of the same scene (e.g. front view, top view, side view etc.) increases the complexity and requires additional cognitive effort to fuse the different views.

The visualizations we propose in this paper are aimed at a single integrated view with visual enhancements that can be switched on/off depending on the current target of analysis. We also add derived information computed from the underlying 3D data like location, speed and angles to enhance our visual markup. We thus extend the existing 3D visualization of the ANVIL video annotation tool which can display figures as skeletons in 3D space with color-coded motion trails of the hands (Kipp, *in press*; Kipp 2012, 2012b; Heloir et al., 2010).

2. Related Work

Nonverbal communication researchers primarily rely on video data for their analyses. Videos are manually annotated with meaningful data like gesture occurrences according to an annotation manual (cf. McNeill, 1992) and these data can then be quantitatively analyzed. Quek et al. (2002) used computer vision techniques to derive such data automatically (to some degree) and to support qualitative analysis with continuous data visualization like motion curves, e.g. the position of the hand along the up-axis over time. Motion capture data provides such data without the need for extraction techniques and much with higher precision. In (Heloir et al., 2010) we presented visualizations both as curves and as 3D markups attached to the 3D stick figure that represents the speaker in the 3D scene. Similar techniques are used in the analysis of sports motion (e.g. swimming motion analysis, see Kirmizibayrak et al., 2011) and for 3D computer games (cf. Zammitto, 2008). Unfortunately, the scientific documentation of such visualizations is rare.

The rich motion caption data is especially useful in multi-party interactions, when computing relationships between people in space, e.g. whether they are oriented toward their interlocutor at particular stages of the interaction (Battersby and Healey, 2010). Systems based on motion capture data are intended both to support qualitative analysis with informative visualizations and to automatically annotate data for quantitative analysis (e.g. the PAMOCAT system by Brüning et al., 2012). In this paper we focus on visualizations for qualitative analysis. However, by automatically writing the visualization information into our hand-made annotations, we can also use these data for quantitative analysis.

3. Single-Person Visualizations

Our first suite of visualizations concern body and hand motion of single subjects.

3.1 Directional Body Motion

Our first visualization is concerned with body motion. We define body motion as the motion of the *hip* through space. The motion direction is indicated by an arrow and the magnitude of the speed is shown with a disc whose diameter is proportional to the magnitude (Fig. 1).

To avoid distraction by small movements we define a speed magnitude threshold below which we do not indicate motion. In this case, the arrow vanishes and the disc turns gray.

This visualization can be used when analyzing the movement patterns of a single person or when analyzing crowds.

3.2 Hand Direction and Speed

Gesture researchers are mainly interested in the movement of the hands. In our data this corresponds most closely to the movement of the *wrist* through space (as opposed to hand-internal motion which is currently still hard to capture). In previous work, we have visualized the path of the hand motion as a color-coded trail of spheres through space (Heloir et al., 2010). Color-coding allows to show the different movement phases (preparation, stroke, retraction etc.) along this trail. In previous work, we visualized the speed of the motion by adding orthogonal 2D circles where the diameter is proportional to the hand's speed at that point. We now added a representation in the form of an arrow whose length is proportional to the hand's speed to make the current direction and speed more visible (Fig. 2).

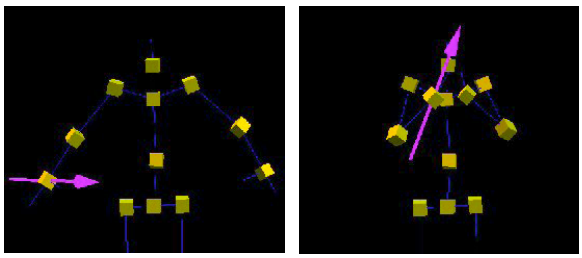

Figure 2: Direction and speed of hand movement is depicted by an arrow along the tangent of the current movement where the arrow's length is proportional to the hand's speed.

This visualization can be useful in the analysis of single gestures.

3.3 Gesture Space

In gesture research the location of the hands during the decisive phases of the gesture (typically stroke and independent hold) is a meaningful aspect of the gesture. McNeill (1992) suggested a scheme called *gesture space* that decomposes the frontal plane into various sections on the extreme periphery, periphery, center and center-center (Fig. 3).

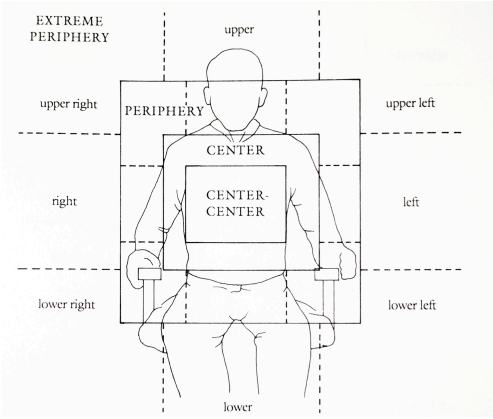

Figure 3: Gesture space (taken from McNeill, 1992).

With motion capture data, the location of the hands in gesture space can be automatically determined and visualized. For visualization, we attached a planar *gesture*

space grid in front of the figure (Fig. 4). We compute whether a hand is within a section. If this is the case the section is highlighted in either yellow (right hand), green (left hand) or red (both hands).

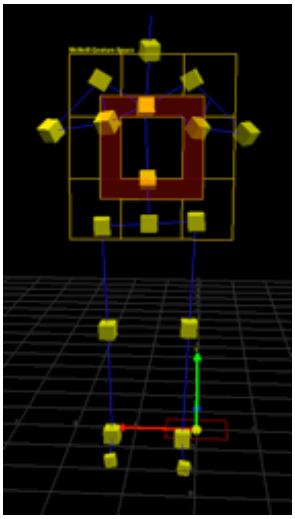

Figure 4: Gesture space is attached as a grid to the skeleton. Sections are highlighted whenever a hand is located within it (yellow: right hand, green: left hand, red: both).

We decided to keep the gesture space grid upright at all times, i.e. it does not bend when the figure's upper body bends. This not only corresponds to McNeill's methodology where a 2D video view is annotated but also avoids visual motion clutter where the grid would constantly make small tilting movements. Also, if a subject bends the concept of gesture space is of limited use and our priority was to make our visualizations as easy-to-read as possible, adding as little distraction as possible. Of course, our grid does follow the figure as it is always positioned in front of and in parallel to the shoulders.

This visualization is useful in gesture research. The automatically detected location in gesture space can easily be exploited for the automatic annotation of gesture location.

4. Multi-Party Visualizations

The following visualizations concern the relationship between multiple people. These visualizations can be used in the context of proxemics (Hall, 1966) and/or when studying social interactions in terms of e.g. F-formations (Kendon, 1990) or micro-territories (Scheflen, 1975).

4.1 Interpersonal Distance

In his theory of proxemics, Edward Hall (1966) introduced *interpersonal space* as a meaningful aspect of nonverbal communication. He divided the possible *distance* between two interlocutors into four functionally different spaces: intimate, personal, social, public. The exact sizes of these zones differ across cultures, e.g. sizes are larger in northern European countries and smaller in southern European countries.

In 3D, it can be hard to see how far apart people are from another unless one looks at them from a bird's eye view, which comes at the cost of having multiple views. Therefore, we visualize distance with an ellipsoid between the feet of the interlocutors that becomes thicker (more circular) if people are closer and slimmer if people are farther away from another (Fig. 5). Thus, there is two shape cues for distance: the length of the ellipsoid and the thickness. Moreover, we color-code the proxemic zones (Fig. 6), i.e. for each of the four zones the ellipsoid changes to a specific color. Finally, we put the precise distance as text into the ellipsoid.

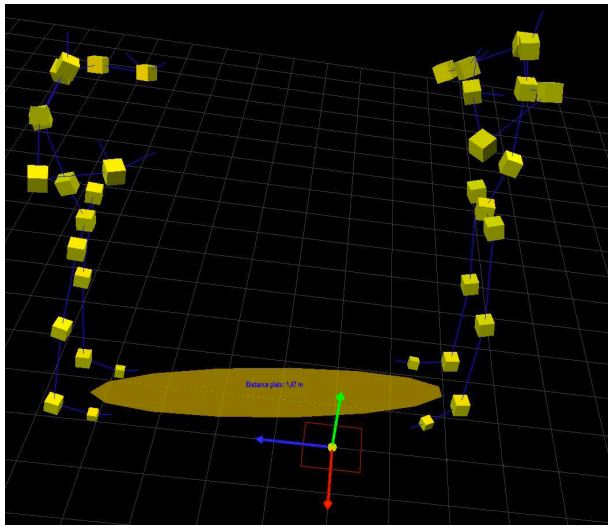

Figure 5: Distance between figures can be hard to read in 3D. Our proxemic visualization displays a color-coded ellipsoid between the feet of the figures which becomes more circular (thicker), the closer the figures are.

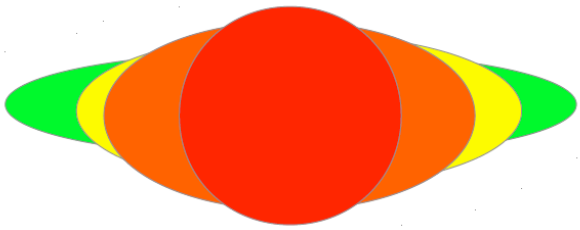

Figure 6: Hall's (1966) *proxemic zones* are visualized by color and shape. From center to periphery: intimate (red), personal (orange), social (yellow), public (green).

The exact sizes of the zones, e.g. at how many meters does the "personal zone" start and end, can be changed in a configuration file to keep the visualization adjustable to different cultures (Hall, 1966).

4.2 Relative Body Movement

In a two-person situation (dyad) it is meaningful whether person A is approaching person B or moving away or moving sideways. To visually clarify this relation, we

combine the single-person visualization for body motion (disc and arrow, see Sec. 3.1) with a small marker which shows the position of the interlocutor. With each figure having a "little radar" around its hip, it is easy to see how the figure is moving relative to another figure. To clarify which figure the little marker is referring to we color-code the marker. In Fig. 7 figure A has a yellow disc and figure B has a blue disc. On figure A's "radar" disc, figure B is then represented with a blue marker. The relation between the figure's own speed arrow and the marker makes clear whether the figure is approaching the other figure or moving sideways etc.

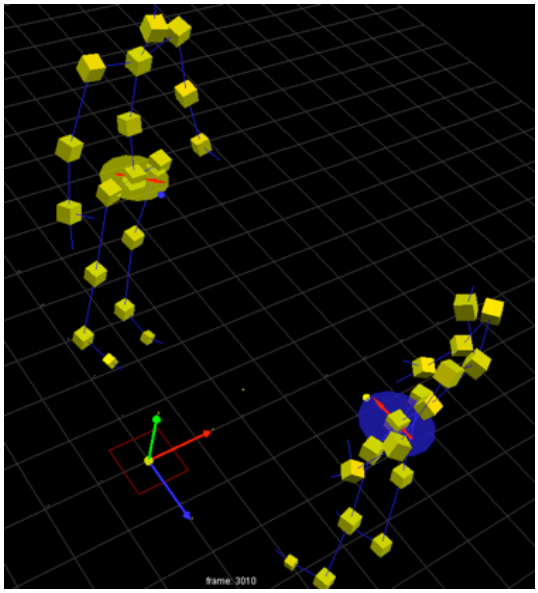

Figure 7: To visualize relative movement we use the body motion discs and add a little marker which represents the other figure. The little blue marker on the left figure's disc corresponds to the right figure.

4.3 Relative Body Orientation

When two people are communicating with each other it is meaningful how they are oriented toward each other (or turned away from another). This can be used to determine the F-formation according to Kendon (1990). Orientation for two people has two aspects: the orientation of a single speaker toward the other (is he facing the other or looking away) and the total angle between the two (if both are looking away, how much so). Therefore, we conceived one visualization for each figure and one for both figures which is positioned exactly in the middle between the two (Fig. 8).

The individual figure's visualization shows two arrows: the first arrow points in the direction of the *other* speaker, the second arrow shows the figure's *own* upper body orientation. The angle between the two arrows shows how much the figure is averted from or facing his interlocutor. Instead of upper body orientation the second arrow could show the direction of the head or eye gaze direction (not implemented).

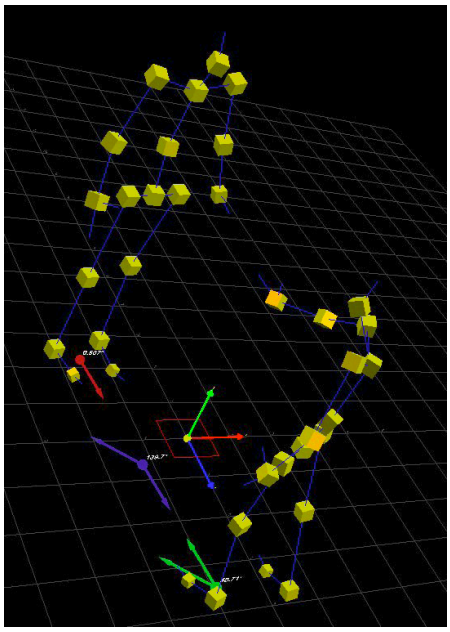

Figure 8: Two visualizations for interpersonal orientation. At the feet of a figure there is one arrow pointing to the other speaker and one arrow for the upper body orientation. Between the figures two arrows show the "overall" orientation relation, i.e. how much the two face each other, taking both figure's orientations into account.

The second visualization which is placed in the middle shows the two upper body orientation arrows of the two speakers together. This represents how much averted both are. If the arrows form a straight line, the speakers face each other. Otherwise, the stronger the divergence from the straight line, the more averted they are. All angles are also shown numerically.

5. Conclusion and Future Work

Although 3D recordings have much to offer in terms of information, it is challenging to visualize this information in a way that makes analysis easier and potentially more substantial. In this paper, we have presented six visualizations that aim at supporting the manual annotation and analysis of nonverbal communication.

The presented techniques are implemented in Java and integrated into the publicly available open-source JMocap software [1] . In future work, we will integrate the visualization techniques into the ANVIL video annotation tool[2] (Kipp, 2001, 2012, 2012b, *in press*). The challenge will be to offer controls to combine various visualizations easily and to optimize them according to the user's needs. Moreover, it is necessary to ensure that visualizations are also compatible and configurable with three and more figures without making the view visually cluttered.

[1] https://code.google.com/p/jmocap
[2] http://www.anvil-software.org

For the future, multi-party interactions with 3+ people may necessitate new visualizations for group constellations like in *interaction process analysis* (Bales, 1951). Another relevant addition would be the automatic analysis and visualization of *posture* (e.g. open vs. closed). Moreover, our techniques need be to evaluated by annotation/analysis experts in two regards: first, whether the visualizations increase coding reliability - i.e. manual annotations become more consistent - and, second, whether the discoverability of new phenomena is facilitated or enabled because of the richer information and information visualization.

6. References

Bales, R. F. (1951). *Interaction Process Analysis*. Chicago University Press, Chicago.

Battersby S., Healey P. G. T. (2010). Head and hand movements in the orchestration of dialogue. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*.

Brüning, B., Schnier, C., Pitsch, K. and Wachsmuth, S. (2012). Integrating PAMOCAT in the research cycle: linking motion capturing and conversation analysis. *Proceedings of the 14th ACM international conference on Multimodal interaction (ICMI '12)*. ACM, New York, pp. 201-208.

Hall, E. T. (1966). *The Hidden Dimension*, Doubleday, New York.

Heloir, A., Neff, M. and Kipp, M. (2010). Exploiting Motion Capture for Virtual Human Animation: Data Collection and Annotation Visualization. *Proc. of the LREC Workshop on "Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality"*, ELDA.

Kirmizibayrak, C., Honorio, J., Jiang, X., Mark, R. and Hahn, J. (2011) Digital Analysis and Visualization of Swimming Motion. *The International Journal of Virtual Reality*, Vol. 10 (3), pp. 9-16.

Kendon, A. (1990) Spatial organization in social encounters: the F-formation system. In A. Kendon, *Conducting interaction: Patterns of behavior in focused encounters*, Cambridge University Press, pp. 209-237.

Kipp, M. (in press). ANVIL: A Universal Video Research Tool. In J. Durand, U. Gut, G. Kristofferson (Eds.) *Handbook of Corpus Phonology*, Oxford University Press.

Kipp, M. (2012). Annotation Facilities for the Reliable Analysis of Human Motion. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC),* ELDA, Paris.

Kipp, M. (2012b) Multimedia Annotation, Querying and Analysis in ANVIL. In M. Maybury (Ed.) *Multimedia Information Extraction: Advances in Video, Audio, and Imagery Analysis for Search, Data Mining, Surveillance and Authoring*, Chapter 21, John Wiley & Sons, pp. 351-368.

Kipp, M. (2001). Anvil - A Generic Annotation Tool for Multimodal Dialogue. *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, pp. 1367-1370.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thoughts*, University of Chicago Press.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C., McCullough, K. E. and Ansari, R. (2002). Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction*, Vol. 9 (3), pp. 171-193.

Scheflen, A. E. (1975). Micro-Territories in human interaction. In A. Kendon, R.M. Harris and M.R. Key (Eds.) *The Organization of Behavior in Face-to-Face Interaction*. The Hague: Mouton Publishers, pp. 159-173.

Zammitto, V. (2008) Visualization techniques in video games. *Proc. of Electronic Visualisation and the Arts (EVA 2008)*.