

The Mona Lisa Gaze Effect as an Objective Metric for Perceived Cospatality

Jens Edlund, Samer Al Moubayed, and Jonas Beskow

KTH Speech, Music and Hearing, Stockholm, Sweden
{edlund, sameram, beskow}@speech.kth.se

Abstract. We propose to utilize the Mona Lisa gaze effect for an objective and repeatable measure of the extent to which a viewer perceives an object as cospatial. Preliminary results suggest that the metric behaves as expected.

Keywords: Copresence, Face-to-face interaction.

Face-to-face interaction evolved between humans present in the same physical space (cospatial) and time (cotemporal; not discussed further here). This copresence is essential for face-to-face interaction. Cospatality concerns interaction between people and ECAs as well, and a potential design goal for ECAs is that they be perceived by persons interacting with them as cospatial. In some ECA implementations, the virtual and physical spaces are implicitly connected via a conduit such as a monitor, which acts as a window onto virtual space. In others, designers attempt to blend virtual and physical space to create a mixed reality. In immersive virtual realities the viewer is pulled from physical space into the virtual reality, and in the holographic projections of science fiction (e.g. the holodeck of Star Trek), virtual characters and objects are moved from virtual to physical space. Some ECAs are clearly perceived as being cospatial: Pepper's ghost displays (e.g. Disney's Haunted Mansion ghosts) give a very strong notion of being present in the room. Others, like traditional paintings, clearly display characters a different space than that of the viewer. There is a range of less obvious examples in-between. The Robot receptionist at CMU [1] features a talking head displayed on a monitor, but the monitor itself moves to face the person it is targeting. One interpretation is that the entire monitor is to be taken for a head – then clearly present in physical space. The image on the monitor, on the other hand, can be interpreted as a head residing in a virtual space. [2] has capabilities that suggest to the user that it is present in the room, such as using gaze to address one out of two persons standing in front of it, while the graphical representation is that of a traditional ECA, suggesting that the character is present in virtual space. In general, any ECA that is represented in virtual space while behaving as if it were present in physical space risks being ambiguous. It is likely that the more properties suggesting that the ECA is present in the room, the more likely it is to be perceived as cospatial. But there is no obvious metric to measure to what extent a viewer perceives cospatality, short of asking viewers post-interaction, which is notoriously unreliable [3].

In [4], we presented an interpretation of the well-documented Mona Lisa gaze effect in which gaze in 2D images is interpreted accurately, although in relative terms

from the perspective of the viewer, whereas gaze projected on a 3D surface – an object in physical space – is interpreted accurately in absolute terms. We propose to capitalize on this to achieve an objective and repeatable measure of the extent to which a viewer perceives an object and the viewer as being cospatial. Given a set of observer judgments of gaze direction, we can score these against the intended gaze target in both relative and absolute terms. A match against the target in relative terms would indicate that the gaze is interpreted as being situated in a virtual space, and not cospatial. A match in absolute terms would indicate that the gaze is interpreted as cospatial. If neither a relative nor an absolute interpretation yields a match with the intended target, we may conclude that the gaze direction was generally difficult to perceive. [4] also presents an experiment paradigm for acquiring large amounts of gaze direction judgments with relative ease.

In the following preliminary experiments, the relative-absolute score range has been normalized so that a perfect relative score (lowest cospatiality) gives -1 and a perfect absolute score (full cospatiality) gives a score of 1. The first experiment tests two varieties of an animated talking head [4]: the first is displayed on a 2D monitor, and we expect of low cospatiality. The second projects the face of the talking head on a 3D model of a head, creating the impression of a physical head in the room. We expect this to be considerably more cospatial. For cospatiality, the 2D monitor scores -0.15 and the 3D head model 0.49, corresponding to expectations. The second experiment utilizes the same monitor, but a back projected mask for the second case. A group of five engaged in dialogue with the ECA, and we noted carefully who answered each question, taking this as an indication of perceived gaze direction. Scores for cospatiality was 1.0 for the back projected mask and 0.75 for the 2D monitor. The first number is as expected, and indicates that the mask is very successful. The second number is lower than the first, which is as expected, yet it is relatively high. We interpret this as an effect of the situation: when engaging in an actual dialogue with the talking head, people are more inclined to think of it as present in the room. This interpretation would require more cognitive effort, as people are overriding the initial interpretation. The 2D case shows significantly longer response times, which is consistent with this interpretation. As a final sanity test, we ran an experiment in which a live person gazed at different subjects. The result, again, was a full 1.0 for cospatiality, which is exactly as expected.

References

- [1] Michalowski, M.P., Sabanovic, S., Simmons, R.: A spatial model of engagement for a social robot. In: Proc. of the 9th International Workshop on Advanced Motion Control (AMC 2006), pp. 762–767 (2006)
- [2] Bohus, D., Horvitz, E.: Facilitating multiparty dialog with gaze, gesture, and speech. In: Proc. ICMI 2010, Beijing, China (2010)
- [3] Nisbett, R.E., Wilson, T.D.: Telling more than we know: Verbal reports on mental processes. *Psychological Review* 84(3), 231–259 (1977)
- [4] Al Moubayed, S., Edlund, J., Beskow, J.: Taming Mona Lisa: communicating gaze faithfully in 2D and 3D facial projections. *ACM Transactions on Interactive Intelligent Systems* (in press)