

Gestures in human-computer interaction – just another modality?

Antti Pirhonen

Department of Computer Science and Information Systems
FI-40014 University of Jyväskylä, Finland
pianta@jyu.fi

Abstract. The role of physical gestures in human-computer interaction has mostly been neglected. In this paper, we argue that gestures are not just one input modality. Rather, they should be seen as a unifying phenomenon, in terms of which all interaction with physical reality could be conceptualised.

1 Introduction

The tradition of analysing multimodality in the field of human-computer interaction is strongly anchored to a somewhat superficial conception of modality. In fact, a more precise term for referring to that conception is sensory modality [1]. Human being is seen as a technical construction with input and output devices. In this view, multimodality refers to interaction in which more than one sensory modality is required.

Many observations of human being in everyday situations challenge the oversimplified conception of modalities. Using one sensory modality in interaction does not necessarily mean that other modalities could be handled as free resources. A famous example is the use of hands-free equipment in a mobile phone. It was introduced as a solution to the distracting effect of using a phone when driving a car – but there is convincing evidence that there is no big difference between a normal use of a phone (holding the device with hand) and hands-free conditions. This is a concrete example how inappropriate it is to handle interaction modalities as separate, independent mechanisms.

The oversimplification of modality conception has its origin in mechanical view of human being. The dominating metaphor of human cognition is still a computer. Therefore, it is not surprising that modalities have traditionally been divided to input and output modalities, analogously with input and output devices of a computer. However, that kind of classification of means with which human beings interact in their environment, does not provide with adequate framework for understanding processes which don't fall into these two simple categories of sensory modalities. Physical gestures, in particular, are phenomena which contain intensive processes in both directions, in and out.

Rather than trying to explain gestures as a combination of input and output, we suggest taking them as a 'meta-modality'. By this we mean that all interaction with

our environment can be composed in terms of physical gestures. A theoretical framework for this claim can be found in the notion of embodied cognition [2]. We argue that by articulating user-interface elements to be designed, in terms of physical action, we would have a sound basis for design which results in coherent wholes. In other words, when user-interface element is being designed in a design group, the properties of each element should be expressed as gestures. Gestures, in turn, can be composed in the next stage to any modality, whether audio, visual, or something else.

An important starting point for this proposal of using gestures as the unifying modality in the design of multimodal interaction was our observation in various experiments some years ago. In those experiments [e.g., 3], we discovered human-computer interaction with a portable music player application. The application was run in a PDA with a touch screen. The player was supposed to be controlled without gaze, utilising the gesture control of the application. The device was hanging on the side of the user, fixed in a seam of a pocket. Browsing the playlist was performed with simple sweeps across the screen: sweeps forward (in walking direction) meant browsing forward, and sweeps backwards meant browsing backwards. In addition, each function had a non-speech feedback sound. The sounds were panned in left-right dimension and were heard through the headphones. There was certain logic in panning – the directions illustrated a control panel of typical music player: left means backwards and right means forward. The combination of gestures and feedback sounds was found unsuccessful; the participants of the experiment did not notice any difference between different feedback sounds. We concluded, that the left-right metaphor of the sounds and the forward-backward metaphor of the gestures, were in conflict with each other, and the user had to choose which one to follow. The result of the analysis was that all of our 20 participants followed the gestural metaphor. We were surprised how strong modality gestures were, at least in that particular application.

This experience is a good example of how human conceptualisation processes can be understood in concrete, physical terms. Ultimately, we argue that the understanding of bodily experience is key issue in strive to figure out how humans interact with their environment. This has important implications in product design. When anticipating the use of a product, it should be kept in mind that the user is a physical creature, who constructs his or her reality through bodily experience.

References

1. Bernsen, N. O. 1995. A toolbox of output modalities: Representing output information in multimodal interfaces. Esprit Basic Research Action 7040: The Amodeus Project, document TM/WP21.
2. Lakoff, George and Johnson, M. 1999. *Philosophy In The Flesh: the Embodied Mind and its Challenge to Western Thought*. Basic Books.
3. Pirhonen, A., Brewster, S. & Holguin, C. 2002. Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of CHI2002*, 20-25 April 2002, Minneapolis, Minnesota.