Corpus for Multimodal Presentations

Milena Pesheva,[1] Adel Elsayed[2]
and Roger Hartley[3]

Abstract

Multimodal presentations form a special class of multimedia content where the emphasis is on capturing a presenter delivering a message to an audience. Such corpus would be useful for researchers involved in the analysis of the organic structure of such communication messages. In our research, we are particularly interested in the role of gestures in assimilating the communication message into a coherently integrated multimodal content (CIMC).

There are three multimodal components in these presentations: speech, represents the *verbal* part of the message, visual objects, usually displayed on a convenient medium, represent the *pictorial* part of the message and hand gestures which integrate the former modalities into CIMC, hence referred to as the *integrating object*. In our study we have identified four main *delivery modes*. This categorization is based on the technology employed in capturing the presentations and the medium used for displaying the pictorial elements. The first is a *board mode*, captured by video, where the presenter uses a board, usually standing up as in a typical classroom delivery. The second, also captured by video, is a *desk mode*, in which the presenter sits on a desk using paper as a display medium. The third and fourth modes are computer-based, in that the pictorial content is stored in a computer and is displayed on the computer screen when delivering the presentation. In the third mode the gestures are entered using a digitizer tablet (or a mouse) and the presentation is captured (audio and visual) using screen capturing software. This mode is referred to as the *Active Multimodal Presentation (AMP) mode*. The fourth mode is the same, but does not use gestures, hence referred to as a *slide show*.

In order to analyse the above presentations, a gesture coding scheme has been developed which is then used, together with a transcript of the verbal content, to produce a descriptive record of the communication message. The verbal transcript is also annotated with the prosodic features of the verbal message. It is proposed that such transcript, supported by the gesture and prosodic annotations, would be sued to supplement the multimedia content of the proposed corpus.

[1] University of Bolton
e-mail: mp6ect@bolton.ac.uk
[2] University of Bolton
[3] University of Leeds