

Impact of captions on deaf and hearing perception of multimedia video clips

Conference or Workshop Item

Accepted Version

Gulliver, S. and Ghinea, G. (2002) Impact of captions on deaf and hearing perception of multimedia video clips. In: Multimedia and Expo, 2002. ICME '02., 25-28 August 2002, Lausanne, pp. 753-756. Available at <http://centaur.reading.ac.uk/31214/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1035891>

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

IMPACT OF CAPTIONS ON DEAF AND HEARING PERCEPTION OF MULTIMEDIA VIDEO CLIPS

S.R. Gulliver G. Ghinea

Department of Information Systems and Computing
Brunel University
UB8 3PH
{Stephen.Gulliver, George.Ghinea}@brunel.ac.uk

ABSTRACT

We investigate the impact of captions on deaf and hearing perception of multimedia video clips. We measure perception using a parameter called Quality of Perception (QoP), which encompasses not only a user's satisfaction with multimedia clips, but also his/her ability to perceive, synthesise and analyse the informational content of such presentations. By studying perceptual diversity, it is our aim to identify trends that will help future implementation of adaptive multimedia technologies. Results show that although hearing level has a significant affect on information assimilation, the effect of captions is not significant on the objective level of information assimilated. Deaf participants predict that captions significantly improve their level of information assimilation, although no significant objective improvement was measured. The level of enjoyment is unaffected by a participant's level of hearing or use of captions.

1. INTRODUCTION

Individuals with perceptual limitations often have limited access to multimedia technologies. Estimates suggest that approximately one in eight people suffer from hearing loss (8.7 million people in the UK), with about 673,000 of these being severely or profoundly deaf and approximately seventy thousand deaf people using British Sign Language (BSL) as either their first or preferred language [2]. It is estimated that 55% of people over 60 years old are either deaf or hard of hearing (HOH) and the number of people suffering from hearing loss is rising significantly [7], yet the effect of captions on an individual's ability to assimilate information from multimedia video clips is relatively unknown.

2. MULTIMEDIA ACCESSIBILITY

There are two distinct types of deaf access: 'assistive' and 'direct'. Assistive access uses technology to help deaf individuals use previously developed systems. Direct access provides access as part of the developed

system [6]. Until recently, the use of text and GUI (Graphical User Interface) interfaces allowed 'direct' access to computer systems for both deaf and hearing users [3]. Unfortunately with the introduction of multimedia, this access equality can no longer be assumed.

To help improve access to multimedia, current 'assistive' technologies focus on two approaches: communication systems and captioning techniques. Communication systems interpret between different communication media. Thus, they perform translation of speech to text, speech to video sign language, text to computer-generated voice or video signal language, which helps two-way communication with deaf individuals. Captions, on the other hand, are a synchronised textual alternative to audio. Captions include all aspects of audio, including descriptions of sounds – using symbols and icons to represent the type of content (such as a musical note to represent music). Although captioning was designed for the deaf and hard of hearing, captions are widely used by the hearing community to either learn a second language, learn to read, or simply in noisy environments.

3. DATA COLLECTION

The aim of our work was to investigate the impact of captions on hearing and deaf perception of multimedia video clips. Perception was measured using a parameter called Quality of Perception (QoP), which targets a user's ability to analyse, synthesise and assimilate informational content in multimedia presentations, as well as his / her level of satisfaction with the quality of the presentations [5].

The original experimental method and video clips used in this work were taken from work carried out by Ghinea and Thomas [4], used to study the quality of service impact of multimedia clips on user perception and understanding. Adaptations to the original experiment method, included:

- The addition of captions, to ensure that the impact of captions could be measured. A JAVA applet was



developed, to display captions, which could be embedded and positioned relative to the video clip within an HTML document.

- Additional feedback questions were added relating to the use of captions, experienced level of enjoyment and predicted level of information assimilation.
- Constant frame rate (25 frames per second), stereo sound (16-bit sampled at 44.1 KHz) and colour depth (24-bit) were used for all video clips. All feedback questions relating to quality change were removed.
- A reduction was made in the number of video clips. Two video clips were removed from the experiment, yet only video clips with duplicated or similar characteristics to another video were removed to limit the impact on experimental results. Also questions were removed for each clip, to ensure total experiment duration lasted no longer than 1 hour.

4. EXPERIMENTAL METHOD

4.1. Experimental Structure

50 participants (20 deaf {**D**} and 30 hearing {**H**}) were divided into groups 1, 2, 3 and 4. Groups 1 and 3 were of equal size and were made up of deaf participants. Groups 2 and 4 were also of equal size and were made up of hearing participants. Group 1 and 3 consisted of individuals with different hearing abilities, yet had an even split of profoundly deaf and HOH participants.

When video clips, without captions, were tested on a pilot study involving deaf participants, the participants soon voiced frustration. This frustration became a resistance to continue with the experiment. To limit this resistance, it was decided that captions should be presented on alternate clips. For detailed information, concerning when captions were shown, see Table 1. For detailed information, concerning video clips, see [3].

Group	1 (D)	2 (H)	3 (D)	4 (H)
Space	Off	On	On	Off
Band	On	Off	Off	On
Snooker	Off	On	On	Off
Cooking	On	Off	Off	On
News	Off	On	On	Off
Rugby	On	Off	Off	On
Pop	Off	On	On	Off
Doc	On	Off	Off	On
Animation	Off	On	On	Off
Weather	On	Off	Off	On

Table 1: Group descriptions and caption information.

4.2. Ensuring Consistency

Consistent experimental conditions (level of lighting, seating angle {90° to screen} and distance from screen {60-80cm}) and equipment (233MHz MMX, with 16 bit 3D sound and a Trident Cyber 9397 3D Video Graphics card) were used throughout all interviews. To reduce background noise and ensure a consistent audio level for all participants, headphones were used. All clips were 352*288 pixel MPEG-1 video.

4.3. Actual, Predicted Information Assimilation and Level of Enjoyment

To understand our results, it is important to see how an individual's level of information assimilation, self predicted level of information assimilation and user defined level of enjoyment were analysed.

Approximately nine questions for each clip had definite answers, for example: "How many pieces of news are there in the bulletin?" As this question has a single definite answer, it is possible to determine what percentage of questions each participant answered correctly. As questions can only be answered if certain information is assimilated from specific information sources (for example, the words of a song can only be gained from audio or caption information) it is possible to determine the proportion of correctly answered questions, which relate to the different types of information. For each question the source of the information, required to answer the question was determined and each question was labelled with the source of the question answer, as indicated below.

- C/A : Information which is presented in both the **Audio** stream and (transcribed) in the **Caption** window (21 questions in total).
- C : Information from **Captions** contained in the video window, for example: the newsreader's name (4 questions in total).
- T : **Textual** information contained in the video window, yet not contained in any captions, for example: the number on a rugby shirt (12 questions in total).
- D : **Dynamic** information contained in the video window, for example: whether an actor exited to the left or right of the screen (12 questions in total).
- V : Other information relating to the **Video** window, for example: the number of lions in the documentary clip (40 questions in total).

By calculating the percentage of correctly absorbed information, it was possible to determine and compare, over a range of video clips, the relative abilities hearing and deaf participants at assimilating certain types of information.



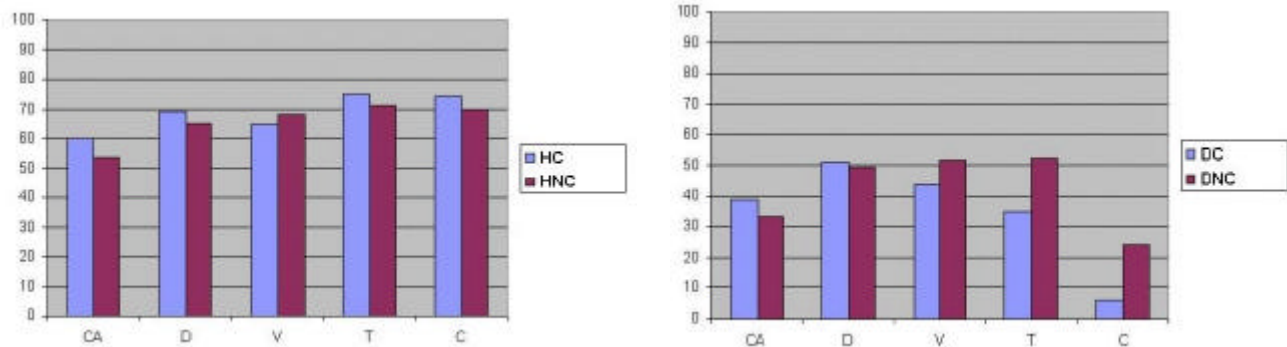


Figure 1: Percentage of correctly assimilated information, taken from hearing and deaf information sources (with and without captions).

All participants were asked (out of ten), for each video clip, how much information they thought they had assimilated. By asking this question we were able to judge what proportion of the available information the participant thought they assimilated. This is important, as it allowed us to analyse both how much information the participant perceived to be available and how accurately they judged themselves to be at information assimilation. This information allowed us to determine whether the participant was aware of the available information.

For each video clip, all participants were asked (out of ten) how much they enjoyed the video clip. This information was used to determine whether ability to assimilate information has any relation to personal preference and level of enjoyment, the second essential component (beside information analysis, synthesis and assimilation) of QoP.

5. RESULTS

It is the aim of this work to investigate the impact of captions on hearing and deaf QoP of multimedia video clips. Therefore as data with and without captions were combined in categories D (groups 1 and 3) and H (groups 2 and 4), data was extracted to form categories DC, DNC, HC and HNC.

- DC** : Deaf with Captions
- DNC** : Deaf with No Captions
- HC** : Hearing with Captions
- HNC** : Hearing with No Caption

5.1. Hearing Perception

The average level of information assimilation experienced by categories HC and HNC was 65.7% and 64% respectively. Comparison of HC and HNC (see Figure 1) showed that captions increase a hearing participant's ability to assimilate all types of information, except video (i.e. C/A, C, T, D). No significant difference was found between hearing level of information

assimilation, when captions were and were not used ($F=0.461$, $df=1$, $P=0.514$). This shows that the ability of hearing participants to assimilate information was not significantly affected by the use of captions.

The average predicted levels of information assimilation experienced by HC and HNC were 63.8% and 61.3% respectively. No significant difference was found between hearing predicted level of information assimilation, when captions were and were not used ($F=0.598$, $df=1$, $P=0.459$). This shows that hearing participants were able to accurately predict their own level of information assimilation regardless of caption use.

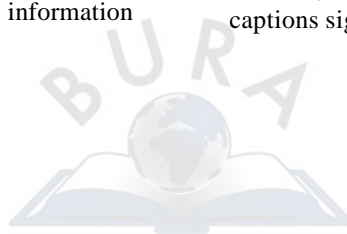
The average level of enjoyment experienced by HC and HNC were 55.2% and 55.6% respectively (see Figure 2). No significant difference was found between hearing level of enjoyment when captions were and were not used ($F=1.498$, $df=1$, $P=0.253$). This shows that the addition of captions has no significant impact on the level of enjoyment experienced by hearing participants.

5.2. Deaf Perception

The average level of information assimilation experienced by DC and DNC was 43.85% and 49.1% respectively. Comparison between DC and DNC showed that, with an even profoundly deaf and HOH split, captions cause a non-significant reduction in the average level of information assimilation ($F=2.346$, $df=1$, $P=0.160$).

A detailed breakdown of user perception (see Figure 1) showed that although the use of captions helps deaf participants absorb more audio / caption and dynamic information, the overall level of information absorbed from the video, textual and captions sources is reduced.

The average predicted level of information assimilation experienced by DC and DNC was 72% and 63.3% respectively. Statistical analysis, using repeated measures, showed a significant difference between 'deaf' predicted levels of information assimilation, when captions were and were not used ($F=7.585$, $df=1$, $P=0.022$). Therefore, deaf participants predict that captions significantly improve their level of information



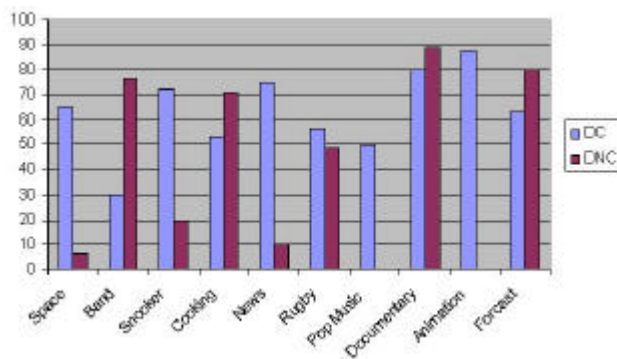
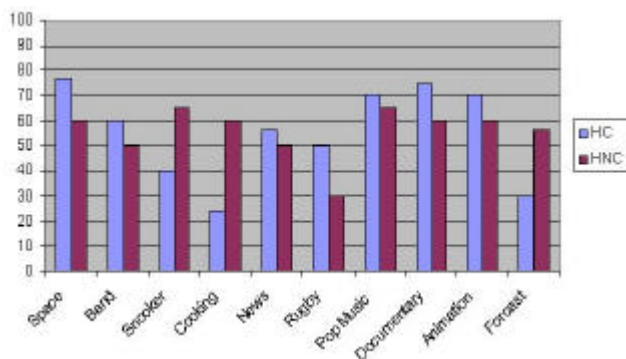


Figure 2: Comparison between levels of enjoyment for all clips (%) for deaf and hearing groups.

assimilation, although no significant objective improvement in information assimilation was measured.

The level of enjoyment, defined by deaf participants, varies considerably between different video clips, as shown in Figure 2. Overall averages in categories DC and DNC were 63.3 % and 40.1 % respectively. Despite these variations, statistical analysis showed no significant difference between deaf level of enjoyment, when captions were and were not used ($F=1.380$, $df=1$, $P=0.270$).

Analysis of variables, with hearing and captions as fixed factors and C/A, C, T, D and V as dependent variables, showed that whilst hearing level has a significant effect on the level of caption / audio and video information assimilated (C/A: $F=5.673$, $df=1$, $P=0.023$; V: $F=19.914$, $df=1$, $P<0.001$) captions don't have a significant effect on the level of information assimilated.

These results show that although hearing level has a significant effect on information assimilation, the effect of captions is not significant on the objective level of information assimilated.

6. CONCLUSION

In this paper, we have examined the effect of captions on hearing and deaf perception of multimedia clips. In summary, our results showed that when captions were used, neither hearing nor deaf participants experienced a significant difference in the level of information assimilation. However, deaf participants did predict a significant change in the level of information that they perceived they had assimilated from the multimedia presentations. No significance difference was measured between the levels of enjoyment experience by deaf and hearing groups. This suggests that, despite assimilating significantly lower level of information, deaf participants experience similar levels of enjoyment to that of hearing participants.

In concluding, the impact of captions has no relative significant effect on the objective level of information assimilated by either deaf or hearing participants. Interestingly, the fact that deaf participants predicted

captions to have a significant impact on information assimilation has important implications to the way in which multimedia satisfaction is measured.

If user perceptual trends are not effectively considered the development of future multimedia systems could exclude users without full sensory and perceptual capabilities.

Alternatively, by providing an appropriate infrastructure that supports multiple representation of information, users could interact 'directly' with multimedia systems on their own terms. Whatever the requirements of the users, the implementation of adaptive multimedia technology would allow the provision of multimedia to match the perceptual criteria of a specific user.

7. REFERENCES

- [1] R.T. Apteker, J.A. Fisher, V.S. Kisimov and H. Neishlos, "Video Acceptability and Frame Rate", *IEEE Multimedia*, 2(3), pp. 32-40, 1995.
- [2] BDA, "What is BSL?", available from <http://www.britishdeafassociation.org.uk/bslpolicy.html>, *British Deaf Association*, November 2001.
- [3] E.M. Finney and K.R. Dobkins, "Visual Contrast Sensitivity in Deaf versus Hearing Populations", *Cognitive Brain Research*, 11, pp. 171-183, 2001.
- [4] G. Ghinea and J.P. Thomas, "QoS Impact on User Perception and Understanding of Multimedia Video Clips", *Proc of ACM Multimedia '98*, Bristol UK, 1998
- [5] G. Ghinea and J.P. Thomas, "Multimedia, Network Protocols and Users – Bridging the Gap", *Proc. Of ACM Multimedia '99*, 1999.
- [6] J. Nielsen, *Advances in Human-Computer Interaction*, Vol. 5, Intellect Publishers, 1995.
- [7] RNID, "Fact sheet - Statistics on Deafness", available from http://www.rnid.org.uk/html/info_factsheets.htm, *Royal National Institute for the Deaf*, September 2001.

