How level and type of deafness affects user perception of multimedia video clips

Article

Accepted Version

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://www.springerlink.com/content/cgulyu900lya2y1y/
To link to this article DOI: http://dx.doi.org/10.1007/s10209-003-0067-5

Publisher: Springer

Publisher statement: The original publication is available at www.springerlink.com

www.reading.ac.uk/centaur

CentAUR

How Level and Type of Deafness Affects
User Perception of Multimedia Video Clips

S. R. Gulliver and G. Ghinea

Department of Information Systems and Computing,
Brunel University,
Uxbridge, Middlesex,
UB8 3PH, UK

E-mail: {Stephen.Gulliver / George.Ghinea}@brunel.ac.uk
Tel: +44 1895 27400 ex 4862 / 3635, Fax: +44 1895 251686

ABSTRACT

Our research investigates the impact that hearing has on the perception of digital video clips, with and without captions, by discussing how hearing loss, captions and deafness type affects user QoP (Quality of Perception). QoP encompasses not only a user's satisfaction with the quality of a multimedia presentation, but also their ability to analyse, synthesise and assimilate informational content of multimedia.

Results show that hearing has a significant effect on participants' ability to assimilate information, independent of video type and use of captions. It is shown that captions do not necessarily provide deaf users with a 'greater level of information' from video, but cause a change in user QoP, depending on deafness type, which provides a 'greater level of context of the video'. It is also shown that post-lingual mild and moderately deaf participants predict less accurately their level of information assimilation than post-lingual profoundly deaf participants, despite residual hearing. A positive correlation was identified between level of enjoyment (LOE) and self-predicted level of information assimilation (PIA), independent of hearing level or hearing type. When this is considered in a QoP quality framework, it puts into question how the user perceives certain factors, such as 'informative' and 'quality'.

KEYWORDS

Quality of Perception, Multimedia Video, Deafness, Hearing Level

1   INTRODUCTION

"Disability" is generally defined as being a physical or mental impairment. When presented with certain types of information, it would seem obvious to assume that the level of information assimilated will be negatively affected due to a reduced level of sensory input, caused as a result of the disability. For example, if a totally deaf individual was at home and the doorbell rang, unless assistive technology had been installed, it seems logical to assume that this individual would remain unaware of this event. Although this assumption can be applied to a number of situations, it relies on

a second assumption, namely that no assistive technology is being used to relay the information to an alternative sensory input, e.g.: coded flashing lights or a vibration device. If the same deaf individual is watching the television, using closed captions and visual alerts as an alternative to voice and sound, can we still assume that the level of information assimilated by a participant would be negatively affected? Does use of captions correct any possible reduction in the level of information assimilated by a user or not? If a reduction does occur in the level of information assimilated from multimedia video, is this reduction reflected in user perceived quality or enjoyment?

This paper investigates the impact that hearing level has on digital multimedia video Quality of Perception (QoP), with and without captions. The following questions are considered in this paper:

- What effect does hearing loss have on a user's Quality of Perception (QoP)? - We aim to identify the effect that hearing loss has on a user's ability to assimilate certain information from multimedia video (objective factor). We also look at how both a user's ability to self-predict their own level of information assimilation and a user's level of enjoyment (subjective factors) are affected by hearing loss.

- Do captions help hearing and deaf user QoP? - By identifying the effect that hearing loss has on user QoP with and without captions we can identify the impact that captions have on user QoP. This allows us to identify whether captions truly help or hinder users with different levels of hearing loss.

- How does deafness type affect user QoP? - Do groups with different types of hearing (post-lingual mildly deaf and pre-lingual profoundly deaf) have a different QoP? How is this affected by the used of captions? By addressing this question we aim to discuss the variance in QoP that exists between two specific deafness groups.

This paper is structured as follows: Section 2 introduces QoP and its implications for the deaf. Section 3 introduces the reader to the deaf community (within the United Kingdom), how different types of deafness are defined, current deaf technology and approaches to multimedia access. Section 4 presents a limited review of work relating to multimedia perceptual limitations. Section 5 describes

the experimental processes used to analyse user QoP, and how data was collected. Section 6 describes our results and, finally, Section 7 concludes by summarising results and giving their possible implications.

2   QOP (QUALITY OF PERCEPTION)

Quality of digital multimedia video clips has traditionally been measured in distributed settings using Quality of Service (QoS) technical parameters, such as jitter, delay, as well as loss and error rates. Although measurable, such objective parameters disregard the user's perception of what defines multimedia quality [19] [22]. To date, there has been a common assumption in the computer networking community that many quality issues will be resolved through objective solutions, such as increased bandwidth allocation [12] [23]. The majority of QoS research has therefore been systems oriented, focusing on factors such as traffic analysis, scheduling and routing.

Unfortunately, due to the multi-dimensional nature of multimedia, it is impossible to rely on objective factors alone when defining multimedia quality. Multimedia applications are produced for the enjoyment and / or education of human viewers, so a user's opinion of the presentation quality is important to any quality definition. Therefore, when evaluating multimedia quality, subjective testing by viewers must be considered in combination with objective testing.

Apteker et al [1] studied video clips at different frame rates over a range of different bandwidths, with user preference being used to determine 'user watchability'. Apteker et al used three dimensions, which they considered inherent in all video messages: the temporal nature of the data, the importance of the auditory and visual components. Apteker showed that 'user watchability' was significantly affected by subjective factors, such as the content of the video clips, not just the level of available bandwidth (objective).

To measure the QoS impact of multimedia video clips on user perception and understanding, Ghinea and Thomas [10] presented candidates with a series of windowed (352*288 pixels) MPEG-1 video clips, between 31 and 45 seconds long. Three frame rates were used: 25, 15 and 5 fps (frames per second). The clips were chosen to cover a broad spectrum of subject matter including: spatial

parameters, temporal parameters, and importance of audio, video and textual information in context of the clip. Their results showed that:

- A significant loss of frames (that is, a reduction in the frame rate) does not proportionally reduce the user's understanding and perception of the presentation. In fact, in some instances the user seemed to assimilate more information.

- Highly dynamic scenes, although expensive in resources, have a negative impact on user understanding and information assimilation. The level of information assimilated in this category obtained the least number of correct answers. However, the entertainment value of such presentations seems to be consistent, irrespective of the frame rate at which they are shown. The link between entertainment and understanding was found not to be direct.

Ghinea and Thomas [10] formed the notion of Quality of Perception (QoP), as they realised that objective factors alone were incapable of defining the perceived quality of multimedia video.

QoP relies on both objective and subjective factors to determine whether or not a specific user, despite their perceptual requirements, perceives a multimedia video presentation to be of 'quality'. QoP as defined by Ghinea and Thomas uses 'Information Assimilation' (IA) and 'Level of Enjoyment' (LOE) to determine the user's perceived level of multimedia quality.

Defining Information Assimilation (IA)

After showing a participant a short video clip, the participant is asked a number of IA feedback questions. These questions are used to determine the type and level of information assimilated by each user. IA questions have a definite answer, for example: (from the rugby video-clip used in this experiment) "What was the score of the match, before the try?" As this question has a definite answer, it is possible to determine whether the participant answered this correctly or not. IA questions are designed so they can only be answered if certain information is assimilated from a specific information source (for example, the number of lions in the documentary is specific to the video). Information sources include:

C / A   : Information which is presented in both the audio stream and (transcribed) in the caption window.

D   : Dynamic information contained in the video window, for example: whether an actor exited to the left or right of the screen.

V   : Other information relating to the video window, for example: the number of lions in the documentary clip.

T   : Textual information contained in the video window, yet not contained in any captions, for example: the number on a rugby shirt.

C   : Information from captions contained in the video window, for example: the newsreader's name.

Definition of IA questions allows us to identify the percentage of correctly answered questions that relate to the different information sources within the multimedia video clip. By calculating the percentage of information assimilated correctly from a various information sources, it was possible to objectively determine where participants are gaining information. It is therefore possible to determine and compare, over any range of video clips, the differences that exist between IA for different groups with different levels of hearing level, with and without captions.

Defining Level of Enjoyment (LOE)

For each video clip, participants are asked how much they enjoyed the video clip. For the purposes of our study a scale of 0 to 10 was used, with scores of 0 and 10 representing "no" and, respectively, "absolute" user satisfaction with the multimedia video presentation. LOE is used in QoP to determine whether ability to assimilate information has any relation to user satisfaction.

Defining Predicted Level of Information Assimilation (PIA)

After answering IA questions, all participants are asked to state how much of the information in a video clip they thought they had assimilated. For the purposes of this study a scale of 0 to 10 was used, with scores of 0 and 10 representing "none" and, respectively, "all" of the information that was perceived as available. Self-predicted information assimilation (PIA), not originally used by Ghinea and Thomas, was added to this study to judge how much information participants thought they

themselves had assimilated. Although not essential to QoP, PIA was considered important to the current study as it allowed us to analyse both how much information the participants perceived to be available, as well as how proficient the participants judged themselves at answering the given questions.

3   DEAF COMMUNITY

Although there are no accurate figures, estimates suggest that in the United Kingdom one in eight persons, namely 8.7 million people, have some form of hearing loss [25]. Moreover, it is estimated that 55% of the individuals over 60 years of age have some form of hearing problems, acquired as a natural result of ageing, a problem that will only increase with the rising number of elderly.

3.1   Definition of Deafness

Deafness can occur at any time in a person's life and the impact of deafness can vary considerably depending whether deafness was acquired before or after speech development. People are therefore defined as being either pre- or post-lingually deaf [26].

"Pre-lingual deaf" is a term for deaf individuals who became deaf before learning language skills. Pre-lingually deaf often find it more difficult to acquire English, as it is an auditory–vocal language, which they can only learn through lip-reading and written word. Those who are defined as pre-lingually deaf often learn sign language as their first language and sometimes are unable to use their voice. Individuals affected by pre-lingual deafness often require specialist tuition to develop certain language skills and commonly have a lower than average reading ability. Approximately 30% of deaf students are functionally illiterate when they leave school, compared to less than 1% of hearing students [18]. People who acquire hearing loss after learning basic language skills are said to be "post-lingually deaf". This includes hearing loss acquired as a natural result of ageing. Post-lingual deaf often continue to use the spoken word, sometimes relying on lip-reading as their only medium of understanding spoken language.

The level of an individual's deafness can be defined as being *mild*, *moderate*, *severe* or *profound*:

Mild deafness: Here, the quietest sounds people can hear in their better ear (the ear which hears the greatest level of sound) averages between 25 and 40 decibels (20dB is equivalent to a whisper or quiet library). People with mild deafness have some difficulty in following speech, mainly in noisy situations.

Moderate deafness: The quietest sounds that people with moderate deafness can hear in their better ear averages between 40 and 70 decibels (60dB is equivalent to a normal conversation). Such people typically have difficulty in following speech without a hearing aid.

Severe deafness: The quietest sounds that individuals in this category can hear in their better ear averages between 70 and 95 decibels (90dB is equivalent to a lawnmower). People with severe deafness often rely a lot on lip-reading, even with a hearing aid. If they have been deaf from early in life, sign language may be the first or preferred language.

Profound deafness: Here the quietest sounds people can hear in their better ear are on average 95 decibels or more (equivalent to a chainsaw). People who are profoundly deaf commonly lip-read and if they are deaf from early in life often use sign language as a first language.

In the U.K., between fifty and seventy thousand people, representing the core deaf community, use BSL as either their first or preferred language, a number similar to that of people speaking Scottish Gaelic [24].

3.2   Deaf Accessibility to Multimedia

There are two approaches to 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 [17]. Until recently, text and GUI (Graphical User Interface) interfaces allowed similar 'direct' access to computer systems for both hearing and deaf users [8]. Unfortunately with the introduction of multimedia, access equality to computer systems can no longer be assumed. Most attempts at access provision have focused on the use of 'assistive' technologies, which helps to fill any accessibility shortfalls of the original design or technology. These 'assistive' technologies can be largely grouped into two main categories: *communication systems* and *captioning techniques*.

Communication Systems

Communication systems interpret between different communications media. Thus, they perform translation of speech to text, speech to video sign language, text to computer-generated voice or video sign language. This helps two-way communication between hearing and deaf individuals. Two such examples of communication systems include the ICommunicator [27] and TESSA [6].

Captions

Use of captions is the most common form of deaf assistive technology and is a synchronised textual alternative to audio. Captions include all aspects of audio, including a description of sounds – using symbols and icons to represent the type of content (such as a musical note to represent music). Open captions, like subtitles, are included as part of the video media and are visible to all. Alternatively, closed captions are only displayed if activated by the user. Captions are commonly displayed at the bottom of the television screen [13], although alternative positing can be used.

Although captioning was designed for the deaf and hard of hearing, captions are also used by the hearing community to either learn a second language, learn to read, or simply in noisy environments. The problem of using captions as a form of deaf assistive technology is that deaf individuals often have a lower or limited level of English literacy, as approximately 30% of deaf students are functionally illiterate when they leave school, compared to less than 1% of hearing students [18].

4 PERCEPTUAL LIMITATIONS

To understand the impact of hearing level on participant QoP, when presented a multimedia video with and without captions, a limited review is provided of work that relates to perceptual limitations.

4.1 Deaf Perception of Captions

Boyd and Vader carried out the earliest studies investigating the use of television for educational purposes [4]. They found that captions that were adjusted to the linguistic level and reading rate of the viewer, significantly improved information gain from captions. Braverman and Hertzog found that it was not the rate of captioning (60, 90, or 120 words per minute) that affected comprehension, but the language level [5]. A rate of 60 words per minute and a reduced language level was found to improve

program comprehension amongst British school children. The reduction of language level in captions, however, frustrates many within the deaf community as information is summarised or ignored, others consider a reduced language level in captions to be patronising.

Jelinek, Lewis and Jackson [13], whilst assessing deaf and hearing students' comprehension of captions with and without visual/video, found that comprehension test scores for students who were deaf were consistently below the scores of hearing students. They also found that caption video provided deaf students with a significantly better comprehension of the script. Deaf students all had hearing loss greater than 60dB for the unaided, better ear across speech frequencies (60dB is equivalent to a normal conversation).

4.2   Acceptable Caption Synchronisation

Captions are used by speakers in many environments, such as CSCW (Computer Supported Cooperative Work), to point out individual elements of graphics and therefore require synchronisation with audio media. Steinmetz measured pointer skews in a CSCW environment where participant were working on a shared map and technical sketch [3] [20].

The 'in-sync' region for pointer skews spans between -750ms (pointer behind audio) and +500ms (pointer ahead of audio). The 'out of sync' region, which is considered as "not acceptable", spans beyond −1000ms and beyond +1250ms. Within the 'transient' area candidates notice the 'out of sync' effect, but it was not mentioned as annoying.

4.3   Influence of Deaf User Experience

Piaget's cognitive theory suggests that learning requires transition through three stages: assimilation, accommodation and equilibration. Assimilation is the process of incorporating a new object, piece of information or event into an existing schema. A schema is something that people construct for dealing with the world and can be defined as either behavioural (grasping or driving a car) or cognitive (solving problems or categorizing objects). If an object does not directly fit a current schema, then accommodation of an existing schema is required. This accommodation, the changing of an existing scheme to fit a new object or event, causes an imbalance between what is understood and what is encountered. To remove this lack of imbalance, new schemas are developed or old schemas are

adapted until equilibrium is found. This concludes that each stage of learning is derived from the previous stage and prepares for the next stage. This process is simulated by a state of disequilibrium. Language development, for example, can therefore not progress without a rich linguistic environment in which to do so. Garrison, Long and Dowaliby [9], whilst examining how working memory affects language comprehension for deaf students, found that reading comprehension depends on readers' background knowledge and lexical abilities. Language abilities increase with interaction, yet due to their reduced level of hearing deaf individuals often have a reduced or limited variety of learning interactions [11].

4.4   Attention limitations

Visual attention can be divided into two stages: pre-attentive processing and selective attention [16]. Pre-attention processing brings together information into objects of perception, which allows objects to be recognised and perceived. Selective attention operates on these objects, yet is limited in capacity. Although pre-attention mechanisms operate outside voluntary control, selective attention requires conscious focus. Location of visual focus at a specific point in time can be used to determine the 'region of interest' [21].

Humans also have difficulty in assimilating audio, visual and textual information concurrently and therefore tend to focus on only one of these media at any one moment, yet focus may switch between different media as long as the information is considered contextually important [10] [11].

If a user is expected to assimilate pieces of information from multiple sources, it may leave the user in a state of confusion. This implies that an increased number of information sources does not necessarily equate to improved levels of IA.

4.5   Type of Media

In general, new information can be held in memory for a few seconds before it starts to deteriorate. However, auditory presentation of information leads to a better level of overall retention [7] [15]. Information can be assimilated into long-term memory, but this often requires considerable effort on the part of the user [2]. For multimedia video it has been shown that the loss of audio data packets has

a more noticeable effect on the assimilation of information than the loss of video frames, and users are less likely to notice degradation of video content if they are shown low quality audio media [14]. Visual information is usually available for longer periods of time than audio information, allowing the user longer periods of time to assimilate the presented information. In previous QoP (Quality of Perception) experiments [10], it was found that a significant loss of frames (a reduction in the frame rate) does not proportionally reduce a user's understanding and perception of a given presentation. In some instances more information can be assimilated, due to increased time viewing each frame.

5 DATA COLLECTION

5.1 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 both consisted of a 10 / 10 split of pre-lingual profoundly deaf BSL users and post-lingual mild and moderately deaf non-BSL users. Although these two deaf groups alone are not representative of the total deaf community [26], they represent the extreme range of user's within the deaf community. Participants defined in this paper as 'pre-lingual profoundly deaf' used BSL (British Sign Language) as their first or preferred language. Participants defined in this paper as 'post-lingual deaf', were post-lingual and were either mild or moderately deaf. Many post-lingual deaf participants still possessed some level of hearing and relied on the English language as their first or preferred language. In this study, when no specific definition is given, profoundly deaf and post-lingual deaf participants are considered together as 'deaf'. Hearing participants were aged between 17 and 58. Post-lingual deaf participants were aged between 15 and 92. Profoundly deaf participants were aged between 28 and 48. Pre- and post-lingual participants were contacted through deaf and hard of hearing clubs and were offered a financial incentive (£10 worth of shopping vouchers) for the use of their time. Participants with varied backgrounds and ages were used to help provide a range of different users within each group.

5.2   Experimental Method

In order to ensure the consistency of experimental conditions, a series of measure were rigously followed in our study. Thus, as the location and working conditions were unknown (since many of the deaf participants had to be interviewed at their homes or at a local deaf club) the same portable computer (233MHz MMX, with 16 bit 3D sound and a Trident Cyber 9397 3D Video Graphics card) was used throughout all interviews. The level of lighting, seating angle (90° to screen) and distance from the screen (60-80cm) were considered in each location to maintain a high level of experimental constancy throughout. Headphones were used to reduce background noise and ensure a consistent audio level for all participants. A consistent audio level (approximately 60dB) was provided to ensure that participants did not increase the volume to a point where any hearing deficiencies might have been attenuated.

To make sure that users' understood their role within the experiment the following instructions were given to all participants. If a participant normally required spectacles, then they were told to use them for the duration of the experiment. All participants, independent of hearing type, were asked to read the menu text at the top of the screen. If a participant was unable to read this writing (even with spectacles) the participant was not included in this study. It was explained that the experiment involved watching 10 short video clips and that some of the video clips would be shown with captions and that some would be shown without. Participants were informed that after each video clip they would be required to stop and answer a number of questions that related to the video clip that had been presented to them. To ensure that the participants did not feel under test conditions, it was made clear that their intelligence was not being tested and that they should not at all be concerned if they were unable to answer any of the IA questions. To check that participants were put in to the correct hearing groups, all participants were asked about their hearing level. To prevent lengthy hearing tests or participant embarrassment, self-defined deafness level, relating to experience, was used to characterise participants. A BSL interpreter relayed all instructions and IA questions into BSL, to ensure that deaf BSL users fully understood what was required of them (BSL interpretation did not extend to the information contained within captions).

Participants were then presented with a series of 10 windowed (352*288 pixel) MPEG-1 video clips, each between 31 and 45 seconds long. The clips were varied in nature and chosen to cover a broad spectrum of subject matters ranging from dynamic action movies to a weather report (see Table 1 for characteristic weightings). After showing each video clip, the video window was closed and the participant was asked a number of IA questions about the video that they had just seen. The questions were designed to examine the type of information assimilated by the user (dynamic (D), video (V), captions and audio (C/A), textual (T), video captions (C)), in accordance with QoP definition. The total distribution of IA questions is as follows: D – 12, V – 40, C / A – 21, T – 12 and C – 4.

➔ **INSERT TABLE 1**

Once a user had answered all questions relating to a video clip, and all responses had been noted, users were asked to rate the quality of the clip using a scale of 0-10 (with scores of 0 and 10 representing the worst and, respectively, the best perceived possible user level of enjoyment).

Many of the video clips used in the experiment had a high audio characteristic weighting (see Table 1), therefore captions were added to identify the impact on both deaf and hearing QoP. Originally it was intended that each participant watched all video clips with or without captions, yet during a pilot study, involving four pre-lingual profoundly deaf users, the participants voiced frustration at the total lack of captions. This frustration soon became a resistance to continue with the experiment. To limit this resistance, it was decided that captions should be presented on alternate clips. Captions, used in this experiment, contained the full contents of the audio and were not summarised or reduced in any way. Therefore, complexity of the captions was largely dependent on the video audio characteristic weighting (see Table 1). For detailed information concerning which captions were shown to which groups, see Table 2.

➔ **INSERT TABLE 2**

The division of participants into four groups, with different combinations of clip-order and captioning, was applied to minimise order-effects and participant frustration, whilst allowing the effect of captions on user QoP to be defined. Although level of literacy is recognised as an issue when using captions, the addition of participant literacy as a experimental factor was considered outside the

scope of this paper. Any reduction in level of information assimilation, as a result of a reduced level of English literacy, was considered as a characteristic of that hearing group as a whole.

➔ **INSERT FIGURE 1**

6   RESULTS

6.1   What effect does hearing loss have on user QoP?

Although it seems common sense that a reduced level of sensory input, as a result of a disability, will cause a reduced level of information assimilation, it relies on the assumption that no assistive technology is being used to relay the information to an alternative sensory input. If a deaf individual is watching a multimedia video using assistive technology, can we still make such an assumption?

To determine the effect that hearing loss has on user QoP, a comparison between deaf and hearing QoP was required. To make this comparison, data from groups 1 to 4 were categorised into combined groups 'hearing' (H) and 'deaf' (D). The D group was made up of the 20 deaf participants from groups 1 and 3, with a 10 / 10 split of pre-lingual profoundly deaf BSL user and post-lingual mild and moderately deaf non-BSL users. The H group was made up 20 hearing participants from groups 2 and 4.

Information Assimilation (IA)

Hearing participants answered, on average, 64.9% of IA questions correctly, while deaf participants answered only 46.5% of IA questions correctly. A relative difference of 39.5% was therefore identified between the hearing and deaf levels of information assimilation, despite duplication of full audio information in captions. Statistical analysis, using hearing level and use of captions as independent variables and average participant IA as a dependent variable, showed that level of hearing causes a significant difference in participant information assimilation $F(1,20) = 39.697$, P< 0.001. This significance difference between hearing and deaf IA raises two questions, namely:

- is it consistent across all video clips and QoP information sources?
- is it caused because of improper use of assistive technology?

A breakdown was made of hearing and deaf IA, to see whether difference in IA occurred consistently across all QoP information sources. Results showed that hearing participants absorbed and retained more information from each of the QoP information sources (C / A, D, V, T, C), as depicted in Figure 2. Initially this seemed surprising, as IA from certain information sources, especially video, does not appear dependent on a user's hearing ability. However, discussion made later in this paper will help the reader to understand why this occurred.

➔ **INSERT FIGURE 2**

Captions contained all available audio information in a textual format, which meant that video clips with a high audio characteristic weighting (see Table 1) often had a high number of captions. Video clips with a low audio characteristic weighting (see Table 1) often have few or no captions. Many will argue that a significant difference found between hearing and deaf IA is simply due to the variance in speed of captions or level of deaf literacy. Although these factors need to be considered, results show that it is not possible to simply assign this difference between hearing and deaf IA specifically to these factors. Due to a lower level of IA across all QoP information sources, deaf participants assimilate relatively less information than hearing participants for all video clips, independent of audio weighting (see Figure 3). As quantity and rate of captions is largely affected by audio weighting, it is clear that the difference between hearing and deaf IA is not due to the implementation and use of assistive technology. Therefore, if a deaf individual was watching a multimedia video, using assistive technology, a reduced level IA is still expected.

➔ **INSERT FIGURE 3**

Level of Enjoyment (LOE)

Hearing level, using hearing level and use of captions as independent variables and average LOE as a dependent variable, was found not to have a significant impact on overall participant LOE $\{F(1,20) = 0.63, p=0.804\}$. A positive correlation was found between PIA and LOE, for both deaf (N=20, CC=0.327, p=.047) and hearing (N=20, CC=0.327, p=0.026) participants. A positive correlation exists between predicted level of information assimilation and level of enjoyment (both subjective factors) independent of hearing level.

16

Predicted Information Assimilation (PIA)

By investigating whether hearing and deaf groups were able to accurately judge their own level of information assimilation (across all video clips), it was possible to identify whether deaf participants were aware of their lower level of QoP PIA (see Figure 4).

→ INSERT FIGURE 4

Deaf participants predicted themselves as having answered 67.7% of answers correctly. Hearing participants predicted themselves as having answered only 62.6% of answers correctly. Statistical analysis, using hearing level and use of captions as independent variables and average PIA as a dependent variable, showed that hearing level does not have a significant impact, across all video clips, on the overall level of predicted information assimilation $F(1,20) = 2.150$, $p=0.151$.

A significant positive correlation was measured between actual and predicted IA for hearing participants, with and without captions (N=20, CC=0.327, p=0.010). No significant correlation was measured between the deaf level of information assimilation and deaf predicted level of information assimilation. Further work is required to clarify this loss of information and what impact this consistently high level of PIA has on deaf perception. This suggests that hearing level affects the relationship between IA and PIA, causing a loss of correlation. One possible implication of this result is that deaf participants, as a result of hearing loss, perceive less information as being available (approximately 70% of the total information available).

6.2   Do captions help hearing and deaf user QoP?

To investigate whether captions help hearing and deaf QoP, the following data categories were created: DC (Deaf participants with Captions), DNC (Deaf participants with No Captions), HC (Hearing participants with Captions) and HNC (Hearing participants with No Captions).

Hearing QoP, With and Without Captions

The average level of IA experienced by categories HC and HNC was 65.7% and 64% respectively, while the average LOE experienced by HC and HNC was 55.2% and 55.6% respectively and the average PIA experienced by HC and HNC was 63.8% and 61.3% respectively. Statistical analysis, using hearing level and use of captions as independent variables and IA, LOE and PIA as a dependent

variables, showed that captions have no significant impact on either IA{F(1,20)=.33, p=0.57}, LOE F(1,20) = 0.004, p=0.949} or PIA {F(1,20) = 0.417, p=0.526}.

Deaf QoP, With and Without Captions

The average level of IA by groups DC and DNC was 43.85% and 49.1% respectively. Analysis between DC and DNC, showed that captions did not cause a significant increase in the overall average level of deaf information assimilation F(1,20) = 1.059, p=0.317. Interestingly, a detailed breakdown of user IA, using analysis of variance with hearing and captions as fixed factors and C/A, C, T, D and V as dependent variables, showed that captions significantly impacted the type of information being assimilated by deaf participants (see Figure 5). Results showed that use of captions has a significant affect on both C/A and V information assimilation (C/A: F(1,21)=5.673, p=0.023; V: F(1,40)=19.914, p<0.001). However, the affect of captions on information assimilation from QoP IA information sources seems to cancel, resulting in no overall significant change in the objective level of deaf IA (see Figure 5). It is interesting to note that questions relating to C/A and V information assimilation account for 68% of total QoP IA questions.

➔ **INSERT FIGURE 5**

Application of captions seems to cause a shift in deaf IA between the video (V) and caption / audio (C/A) elements of the video clip (see Figure 5). When captions are added, attention is drawn to captions, resulting in a higher level of IA from the C/A information source. When captions are not used attention is drawn to the video, resulting in a higher level of IA from the video information source. The overall level of deaf IA is not significantly affected by the used of captions.

The level of enjoyment, experienced by deaf participants, varies considerably between different video clips. Overall averages in categories DC and DNC were 63.3 % and 40.1 % respectively. Despite this variation, statistical analysis showed captions to have no significant impact on deaf level of enjoyment {F(1,20) = 2.783, p=0.113}.

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, with and without captions $F(1,20) = 7.585$, p=0.022.

Jelinek, Lewis and Jackson [13] found that captioned video provides deaf students with significantly better comprehension of the script (the story). This helps us to account for why deaf participants predict captions to have such a significant impact on their level of information assimilation. Captions cause a shift in attention from video information to caption information, which results in an increased level of information assimilation from caption / audio information sources. This increase in caption / audio information seems to provide the user with a greater level of context of the clip, despite an equivalent reduction in the level of video IA. This would account for the commonly held belief that captions provide deaf users with a greater level of information from video. QoP analysis suggests that captions do not necessarily provide deaf users with a 'greater level of information' from video, but that the information assimilated from captions provides a 'greater level of context of the video'.

6.3   Effect of Deafness Type on User Perception

Deaf participants used in this experiment can be grouped as being either pre-lingual profoundly deaf BSL users or post-lingual mild and moderately deaf non-BSL users. Although these deaf groups do not demographically representative the deaf community [26], they do represent the range of possible deafness. To consider the difference between these two deafness types, a comparison of QoP was made.

The average levels of information assimilation in profoundly deaf and post-lingual deaf categories were 46.9% and 44.56% respectively. Despite very different hearing levels, analysis showed that there was not a significant difference between the average level of participant IA as a result of deafness type {$F(1,20) = 0.285$, p=0.597}. Although the average level of IA was not significantly affected by hearing type, a breakdown of user IA, with and without captions, provides interesting results about the type of information assimilated by different deafness groups.

Use of captions caused pre-lingual profoundly deaf participants to experience a shift in IA, between caption and video elements (see Figure 6). Pre-lingual profoundly deaf participants assimilated significantly less video (V), textual (V) and video caption (C) information when captions were

displayed. However, use of captions caused participants to assimilate significantly more caption / audio (C/A) information, resulting in no significant overall change in average IA. Unlike pre-lingual profoundly deaf participants, captions did not cause post-lingual mild and moderately deaf participants to experience a significant difference in IA from specific QoP information sources (see Figure 7). This suggests that post-lingual mild and moderately deaf participants are not distracted when captions are used. We conclude that post-lingual mild and moderately deaf participants assimilate the C/A element of video from audio information, using residual hearing, and only use captions if audio is lost or confused. This would explain the difference in information assimilation between pre-lingual profoundly deaf and post-lingual mild and moderately deaf groups, when captions were and were not displayed.

➔ **INSERT FIGURE 6**

➔ **INSERT FIGURE 7**

The average level of enjoyment (LOE) experienced by profoundly deaf and post-lingual deaf participants was 57.5 % and 70.75 % respectively. A positive correlation was identified between deaf LOE and PIA for both profoundly deaf and post-lingual deaf groups. This result has important implications on a user's definition of 'informative' as it implies that deaf PIA (prediction of IA) is actually more related to level of enjoyment, the second essential component of QoP, than level of information assimilation.

Pre-lingual and post-lingual deaf PIA was 57% and 71% respectively (122 and 157.4 percent of the actual result). A significant difference in PIA was identified between for the two deaf groups $F(1,20) = 4.643$, p=0.038. Analysis showed that post-lingual deaf participants were less able to predict their level of IA than profoundly deaf participants, despite post-lingual deaf participants still possessing a level of residual hearing. This result is interesting as it implies that other factors, other than hearing loss and shift of attention result in high predicted PIA. Further work is needed to identify the causes of high participant PIA within deaf perception and how correlation with LOE affects user definition of 'informative'.

7   CONCLUSIONS

This paper explored the impact of hearing level on multimedia video QoP, with and without captions. The following questions were considered: What effect does hearing loss have on user QoP? Do captions help hearing and deaf user QoP? How does deafness type affect user QoP?

7.1   What effect does hearing loss have on user QoP?

Hearing participants assimilated significantly more than 'deaf' participants, across all QoP IA information sources (see Figure 2). Although a lower average level of IA was expected for certain information sources, especially without the use of captions, a reduced level of IA over all information sources was surprising. A reduction in especially deaf IA of V, T and C information sources was only later explained by the shift, due to the use captions, in pre-lingual profoundly deaf attention.

It is interesting that this lower average IA, across all QoS IA information sources, appears consistent across all video clips, independent of audio weighting (see Figure 3). As quantity and rate of captions is largely affected by audio weighting, it is clear that the difference between hearing and deaf IA is not due to the implementation and use of assistive technology within the experiment. A reduction in deaf IA is therefore not due alone to the use or absence of caption. This implies that captions cannot be used to correct deaf levels of information assimilation. From the above results we can conclude that, as long as participant focus is consistent, a deaf individual watching a caption-assisted multimedia video, independent of audio weighting or clip type, would still assimilate less information than a hearing individual. This conclusion supports Jelinek et al [13], who whilst assessing deaf and hearing student's comprehension of captions with and without visual/video, found that comprehension test scores for students who were deaf were consistently below those of hearing students. Further work is required to identify the impact of other factors, such as participant literacy, caption complexity and social background, on 'deaf' IA.

The average predicted level of information assimilation between hearing and deaf participants, showed no significant difference (see Figure 4). One implication of this high level of deaf PIA is that deaf participants, as a result of hearing loss, perceive less information as being available

(approximately 70% of the total information available). Further work is required to clarify this loss of information and what impact this consistently high level of PIA has on deaf perception.

A positive correlation was identified between LOE and PIA, independent of hearing level or hearing type. When this is considered in a QoP quality framework, it puts into question how factors, such as informative content, are defined by the user. If a greater correlation exists between PIA and LOE than between PIA and IA, can enjoyment be purposefully adapted to aid a user's perception of informative video? This opens an interesting avenue of further research, which explores how this correlation affects user definition of 'informative'.

7.2   Do captions help hearing and deaf user QoP?

Captions were found not to have any significant affect on the average overall level of information assimilated, for either deaf or hearing participants. However, a detailed breakdown of user IA showed that captions significantly impacts the type of information being assimilated by deaf participants. Results showed that use of captions significantly affects both C/A and V information assimilation. Captions cause a shift in deaf IA between the video (V) and caption / audio (C/A) elements of the video clip (see Figure 5). When captions are added, attention is drawn to captions, resulting in a higher level of IA from the C/A information source. When captions are not used, attention is drawn to the video, resulting in a higher level of IA from the video information source. The overall level of deaf IA is not significantly affected by the used of captions. We can conclude that captions do not significantly change the deaf overall level of IA, however, they do cause assimilation from different information sources.

Results showed that deaf participants predicted captions to cause a significant improvement in their level of information assimilation, even though no significant change in overall deaf IA was measured. This result supports our argument that captions do not necessarily provide deaf users with a 'greater level of information' from video, but that the information assimilated from captions provides a 'greater level of context of the video'. Jelinek, Lewis and Jackson [13] found captioned video provides deaf students with significantly better comprehension of the script (the story). This helps account for why deaf participants predict captions to have such a significant impact on their level of

information assimilation. Results show that captions cause a shift in attention from video information to caption information. This shift causes an increased level of information assimilation from caption / audio information sources, which seems to provide 'deaf' participants with a greater level of context of the clip, despite an equivalent reduction in the level of video information assimilated. This would account for the commonly held belief that captions provide deaf users with a greater level of information from video. Further work is required to investigate whether variation in captions can be used to help focus deaf attention to provide increase overall deaf IA.

7.3 How does deafness type affect user QoP?

No significant difference was found between pre-lingual profoundly deaf and post-lingual mild and moderately deaf average IA. However, a breakdown of IA for both deaf categories, with and without captions, provides interesting results about the type of information assimilated by different deafness groups. Captions caused significant changes in the levels of video (V), textual (V), video caption (C) and audio / caption IA for pre-lingual profoundly deaf participants. Alternatively, post-lingual mild and moderately deaf participants do not experience significant shifts in IA from any specific QoS information source. Unlike pre-lingual profoundly deaf participants, this suggests that post-lingual mild and moderately deaf participants do not experience a significant shift in attention as a result of added captions. We concluded that post-lingual mild and moderately deaf participants, unlike pre-lingual profoundly deaf participants, assimilate the C/A information from the audio media, using residual hearing, and only use captions if audio is lost or confused. This suggests that post-lingual mild and moderately deaf, like hearing participants, do not naturally rely on captions for certain information. As post-lingual participants lose hearing once reliance on spoken language skills has been developed, a reduction in volume causes a loss of information. Further work is required to investigate the QoP impact of reduced levels of audio on post-lingual mild and moderately deaf. To pre-lingual profoundly deaf participants, the absence of sound is largely irrelevant as reliance on other methods of IA (for example: captions) that do no rely on hearing are developed from an early age. This reliance on captions, as the only source of context rich information, explains the significant shift

in profoundly deaf IA in experimental results. Further work is required investigating the impact of different captioning techniques and use of video sign language translation on deaf QoP IA.

A significant difference in PIA was identified between for the two deaf groups. Results showed that post-lingual deaf participants were less able to accurately predict their IA than profoundly deaf participants, despite residual hearing. This result is interesting as it implies that other factors, other than hearing loss and shift of attention result in high predicted PIA.

A positive correlation was identified between deaf LOE and PIA for both profoundly deaf and post-lingual deaf groups. This result opens an interesting avenue for further research, which would explore how this correlation affects user definition of 'informative video'.

In conclusion, we can see that QoP allows detailed analysis of both hearing and deaf perception of multimedia, with and without captions. Many questions and areas of further research have been raised by our results, yet the authors consider QoP as the most effective method of analysing multimedia perception, because both subjective and objective factors are considered. Multimedia holds both promise and danger in the area of user perception. If limitations to user accessibility are not effectively considered, the development of future multimedia systems may exclude direct access for users without full sensory and perceptual capabilities. The authors see a need for an appropriate adaptive infrastructure that supports varied representation of information, allowing users to directly interact with multimedia systems on their own perceptual terms. Whatever the requirements, the implementation of adaptive multimedia technologies would allow provision of multimedia access to match the perceptual criteria of a specific user.

REFERENCES

1. Apteker RT, Fisher JA, Kisimov VS, Neishlos H (1995) Video Acceptability and Frame Rate. IEEE Multimedia, 2(3), pp 32-40.

2. Atkinson RC, Shiffrin RM. (1968) Human Memory: A proposed system and it's control processes. In K. W. Spence and J. T. Spence, (Eds.), The Psychology of learning and motivation. New York: Academic.

3. Blakowski G, Steinmetz R (1996) A Media Synchronisation Survey: Reference model, specification and case studies. IEEE Journal on Selected Areas in Communications, 14(1), pp 5-35.

4. Boyd J, Vader EA (1972) Captioned television for the deaf. American Annals of the Deaf. 117(1), pp 34-37.

5. Braveman BB, Hertzog M (1980) The effects of caption rate and language level on comprehension of a captioned video presentation. American Annals of the Deaf. 125(7), pp 943-948.

6. Cox. S, Lincoln. M, Tryggvason. J, Nakisa. J, Wells, M, Tutt. M and Abbot. S., TESSA, a system to aid communication with deaf people. *ASSETS 02,* Edinburgh, Scotland, 2002.

7. Crowder RG (1970) The role of one's own voice in immediate memory. Cognitive Psychology, pp 157-178.

8. Finney EM, Dobkins KR (2001) Visual Contrast Sensitivity in deaf versus hearing populations: exploring the perceptual consequences of auditory deprivation and experience with a visual language. Brain Research and Cognitive Brain Research, 2001, pp 171-83.

9. Garrison W, Long G, Dowaliby F (1997) Working memory capacity and comprehension processes in deaf reader". Journal of Deaf Studies and Deaf Education, 2(2), pp 78-94

10. Ghinea G, Thomas JP (1998) QoS Impact on User Perception and Understanding of multimedia Video Clips. Proc. of ACM Multimedia '98, Bristol, UK, pp. 49-54.

11. Hapeshi K, Jones D (1992) Interactive Multimedia for Instruction: A cognitive Analysis of the role of Audition and Vision. International Journal of Human-Computer Interaction, 4(1), pp 79-99.

12. Jayant NS (1990) High-quality coding of telephone speech and wideband audio. IEEE Communications Magazine, Jan. 1990, pp 10-20.

13. Jelinek Lewis MS, Jackson DW (2001) Television Literacy: Comprehension of Program Content Using Closed Captions for the Deaf. Journal of Deaf Studies and Deaf Education, 6(1), pp 43-53

14. Kawalek J A (1995) A User Perspective for QoS Management. Proc. Of QoS Workshop aligned with the 3rd International Conference on Intelligence in Broadband Services and Network (IS&N 95), Crete, Greece.

15. Murdock BB, Walker KD (1969) Modality effects in free recall. Journal of Verbal Learning and Verbal Behaviour, 8, pp 665-676.

16. Neisser, U (1967) Cognitive psychology. New York: Appleton-Century-Crofts.

17. Nielsen J (1995) Advances in Human-Computer Interaction", Volume 5, (editor), Intellect Publishers, 1995, ISBN 1-56750-199-0.

18. Paul PV, Quigley SP (1994) Language and deafness (2nd ed.). San Diego: Singular Publishing Group.

19. Steinmetz R, Nahrstedt K (1995) Multimedia: Computing, Communications and Applications" Englewood Cliffs, NJ: Prentice-Hall.

20. Steinmetz R (1996) Human Perception of Jitter and Media Synchronisation. IEEE Journal on Selected Areas in Communications, 14(1), pp 61-72.

21. Van den Branden Lambrecht CJ, Verscheure O, (1996) Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System. Proc. of the SPIE, 2668, pp 450 - 461.

22. Watson A, Sasse MA (1997) Multimedia conferencing via multicasting: determining the quality of service required by the end user. Proc. of AVSPN '97, Aberdeen, UK, pp. 189-194.

23. Zhang L, Deering S, Estrin D, Shenker S, Zappala D. (1995) RSVP: A new resource ReSerVation Protocol. IEEE Network Magazine, 7(5), pp 8-18.

24. British Deaf Association (November 13, 2001) What is BSL. Available at http://www.britishdeafassociation.org.uk/bslpolicy.html

25. Hearing Concern (January 25, 2002) Understanding Deafness. Available at http://www.hearingconcern.com/factsheets/factsheet4.htm

26. Royal National Institute for the Deaf (February 08, 2002) Fact sheet - Statistics on Deafness. Available at http://www.rnid.org.uk/html/info_factsheets.html

27. Teach the Deaf (November 26, 2001) Description of ICommunicator. Available at http://www.teachthedeaf.com/icomm/overview.php3

	Dynamic	Audio	Video	Text
Space (Action Sequence)	black	grey	black	white
Band	grey	grey	grey	white
Snooker	white	grey	grey	black
Cooking	grey	black	black	white
News	white	black	grey	grey
Rugby	black	grey	black	black
Pop Music	grey	black	black	black
Documentary	white	grey	black	white
Animated Movie	grey	grey	black	white
Weather Forecast	white	black	grey	black

Table 1: Videos and Characteristic weightings originally defined by Ghinea and Thomas [10] (white – no importance, grey – some importance, black – important)

Video Clip	Group 1	Group 2	Group 3	Group 4
	(D)	(H)	(D)	(H)
Space (Action Movie)	Off	On	On	Off
Band (Jazz Band)	On	Off	Off	On
Snooker	Off	On	On	Off
Cooking	On	Off	Off	On
News (Local news)	Off	On	On	Off
Rugby (World Cup)	On	Off	Off	On
Pop (MTV style music)	Off	On	On	Off
Doc (Lion Documentary)	On	Off	Off	On
Animation (101 Dalmatians)	Off	On	On	Off
Forecast (National Weather)	On	Off	Off	On

Table 2: Use of Captions for Groups 1 to 4 (ON – Captions; OFF – No Captions)

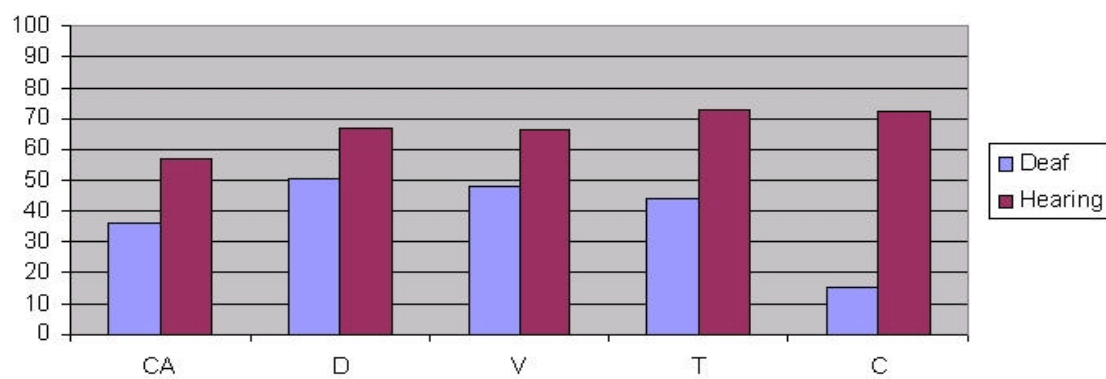Figure 1: Screen shot of experiment (Pop Music).

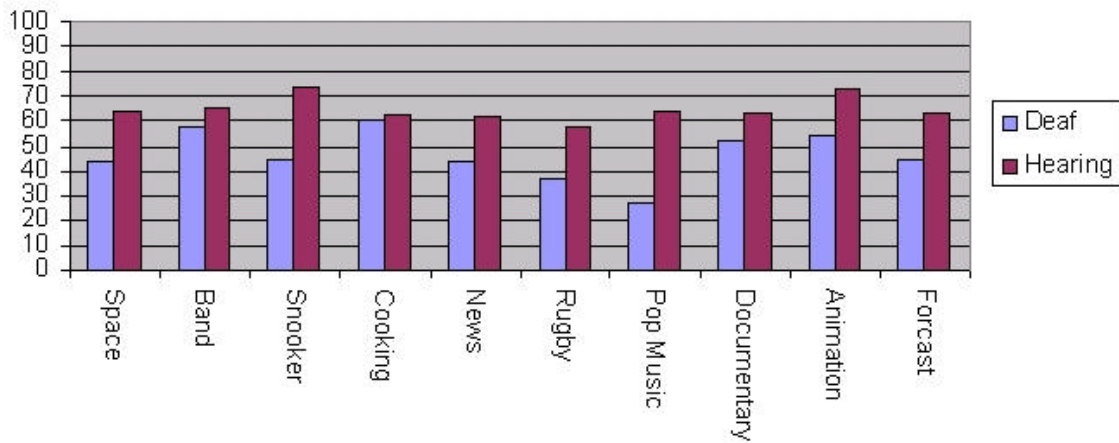Figure 2: A detailed breakdown of the deaf / hearing IA (%).

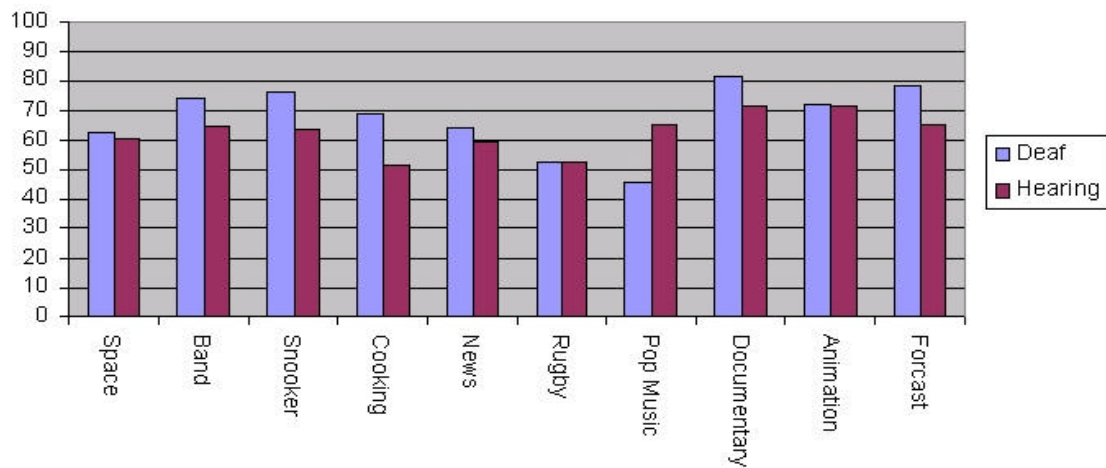Figure 3: Percentage of correct answers for Deaf and Hearing groups, over all Video Clips.

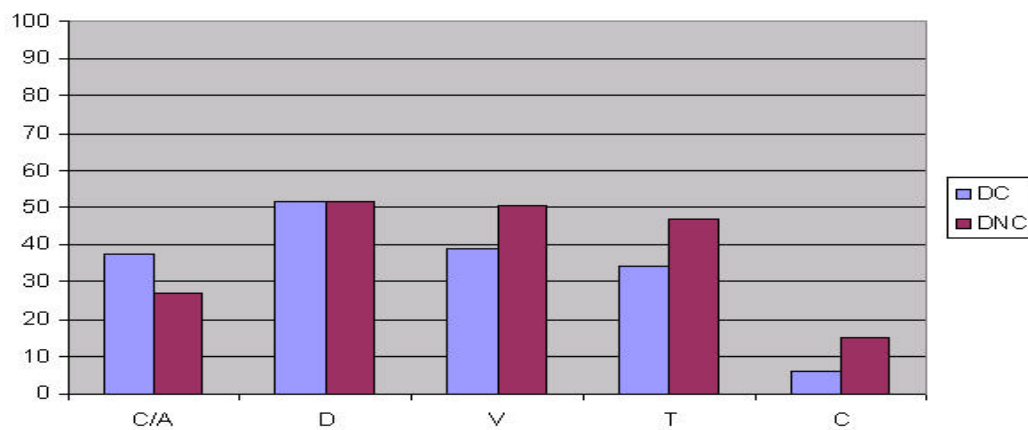Figure 4: Predicted level of Information Assimilation (%), for Deaf and Hearing groups.

Figure 5: A detailed breakdown of the deaf IA, with (DC) and without (DNC) captions (%).
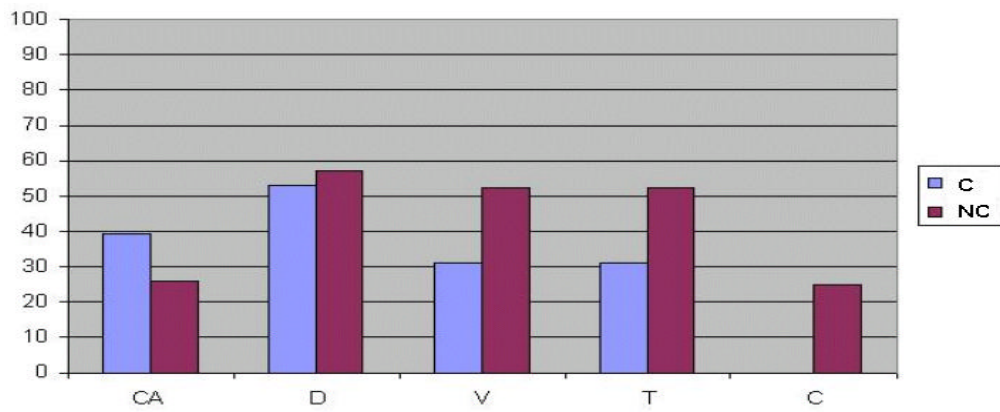
Figure 6: A detailed breakdown of the pre-lingual profoundly deaf IA,
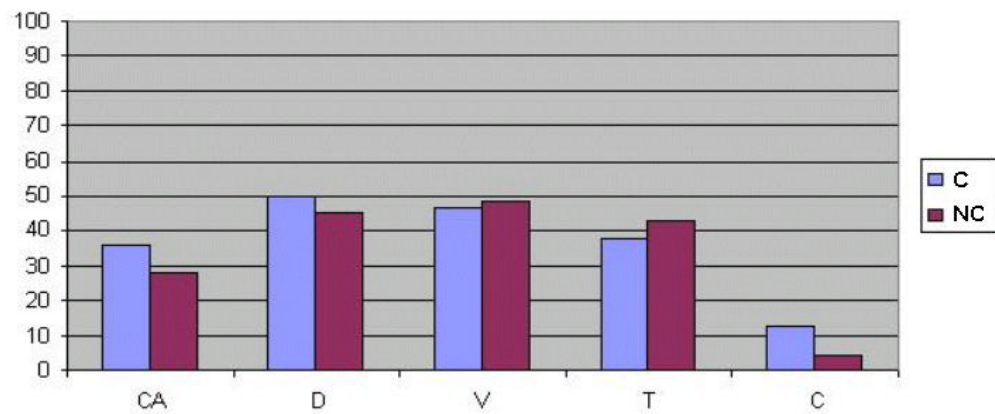with (C) and without (NC) captions (%).

Figure 7: A detailed breakdown of the post-lingual mild and moderately deaf IA,
with (C) and without (NC) captions (%).