This paper should be cited as: Wieczorek, A. M., Kłyszejko, Z., Szarkowska, A., & Krejtz, I. (2011). Accessibility of subtitling for the hearing-impaired. Proceedings of the Conference: Interfejs użytkownika - Kansei w praktyce, Warszawa 2011 (pp. 27–33). Warsaw: Wydawnictwo PJWSTK.

Accessibility of subtitling for the hearing-impaired

Anna Maria Wieczorek
Interactive Technologies Laboratory, Institute for Information Processing; Department of Psychology, Warsaw School of Social Sciences and Humanities
anna.wie4@gmail.com

Zuzanna Kłyszejko
Interactive Technologies Laboratory, Institute for Information Processing; Department of Psychology, Warsaw School of Social Sciences and Humanities
zklyszejko@gmail.com

Agnieszka Szarkowska
Institute of Applied Linguistics University Of Warsaw
aszarkowska@gmail.com

Izabela Krejtz
Interdisciplinary Center for Applied Cognitive Studies
Warsaw School of Social Sciences and Humanities
iza.krejtz@gmail.com

Abstract
The paper aims at answering the question how to improve accessibility of subtitling for the hearing-impaired. One of the most frequently asked questions in captioning is whether captions should be edited or verbatim. e paper reports on the results of an eye-tracking study on captioning for the hearing-impaired, reading different types of captions. By examining eye movement patterns when watching clips with verbatim, standard and edited captions, we tested whether the three different caption styles were read differently by deaf, hard of hearing and hearing participants. In terms of group differences, deaf participants differed from the other two groups only in the case of reading verbatim captions. The results are discussed with reference to classical reading studies, audiovisual translation and a new concept of viewing speed.

Keywords
accessibility, eyetracking, reading, subtitling, verbatim/standard/edited captions, hearing-impaired, deaf, hard of hearing,

Interfejs użytkownika – Kansei w praktyce 2011
ISBN 978-83-89244-99-4

Introduction

Recent years has seena big amount of efforts aimed at improving accessibility of audiovisual programs for people with hearing imapirments through closed captioning. Countries that have taken the lead, such as the US or the UK, are especially focusing on the quantity of captioning and on increasing statutory targets. Others, such as Poland, where this study takes place, have tried to follow suit – both in terms of quantity and quality of captioning, but are still lagging behind. Regardless of the development stage of captioning in particular countries, it seems that certain themes are recurrent in captioning. One of them is the question whether captions should be rendered verbatim, i.e. be a literal and faithful transcription of the dialogue list, or should they rather be edited, i.e. reduced and simplified in order to foster comprehension and facilitate the reading process. It is this question that we address in our paper.

In this paper we report on the Polish results of eyetracking research on accessibility of captioning for the hearing-impaired carried out within the framework of the EU-funded project *Digital Television for All* (DTV4ALL). By examining eye movement patterns of 40 deaf, hard of hearing and hearing viewers when watching clips with verbatim, standard and edited captions, we aim to establish which type of captions would be optimal for the hearing-impaired. With this goal in mind, we analyze the overall comprehension in the three captioned clips, the percentage of time spent on captions vs. on image, the number of times the participants moved their eyes from captions to the image (deflections) and fixation patterns in the three caption styles.

As mentioned above, one of the most frequently recurring themes in captioning is whether captions should be edited or verbatim. According to Romero Fresco [1] and Neves [2], there are three groups of stakeholders in the verbatim vs. edited captioning debate. First, there are hearing-impaired viewers and deaf organizations who often treat any editing of captions as censorship and demand equal access to dialogues. The second group of stakeholders are broadcasters, many of whom also support verbatim captioning. This is due to financial considerations since verbatim captions are cheaper and faster to produce, particularly in countries using speech recognition technology. The third group of stakeholders are researchers, many of whom have called for the editing of captions on the grounds that the reading rates in verbatim captions can be so high that they are actually almost impossible to follow.

Rationale

The present study contributes to prior work by evaluating eye-movement characteristics of deaf, hard of hearing and hearing viewers when watching clips with three different captioning styles: verbatim, standard and edited. The approach adopted here differs from most previous studies in the treatment of particular captioning styles (we have examined three styles and not just two), the use of sound during the tests (our participants had the sound turned on) and the language of the experiment (Polish).

As mentioned above, three captioning styles were used in the experiment: verbatim, standard and edited, based mainly on their linguistic characteristics and display times. Verbatim captions included every single word from the dialogue – even words which usually do not find their way to captions, such as repetitions, hesitations and other elements typical of spoken language

Method

Forty two participants aged between 18-65 took part in the study. They were divided into three groups based on the declared degree of hearing loss: (1) the deaf (*N*=9), (2) the hard of hearing (*N* =20), and (3) the hearing (*N* = 11) as the control group. Data from one hard of hearing and one deaf participant were excluded from analyses due to poor quality of their eye movement recordings, giving the total N = 40. There was a similar proportion of men and women (46% and 54% respectively). The test lasted approximately one hour, after which all participants were rewarded with promotion kits from the two institutions conducting the study: the University of Warsaw and Warsaw School of Social Sciences and Humanities.

Participants watched 23 short clips with the sound turned on, lasting about 1 minute each, taken from *Shrek* dubbed into Polish. Participants were tested individually and asked to watch the clips as they would watch a regular film. First, they filled out a questionnaire with personal details and questions concerning their views on captioning in Poland. Then, they were seated in front of the eyetracker at a viewing distance of ca. 70 cm. Calibration, performed before and in between viewing video clips, required a participant to visually follow nine dot targets displayed sequentially. After watching each clip, participants answered three comprehension questions on (1) general comprehension of the clips, (2) textual elements included in subtitles, and (3) visual image. They were also asked about their caption preferences (e.g. whether they preferred verbatim, standard or edited captions).

Participants' eye movements were recorded with an Eye-Link CL eyetracking system (SR Research, Canada) with a sampling rate of 500 Hz. Participants were seated in front of a monitor (1024 X 768 resolution; 17-inch LCD, refresh rate 60 Hz) at a distance of ca. 70 cm. Their heads were positioned on a chin rest in order to minimize head movements.

Dependent variables

The proportion of dwell time on caption reading relative to scene viewing

Dwell time, or *gaze duration*, is here defined as the sum of all fixations in the caption area of interest [3, 4].

Fixation count

A fixation can be simply defined as a period of time between two consecutive rapid eye movements, called saccades [3]. Fixation duration as well as fixation count can be treated as indices of information processing difficulty.

Deflections from image to captions

Deflections are here defined [5] as the "number of times a viewer's eyes deflected away from scene viewing to focus on the image". As in d'Ydewalle and Bruycker's study [6], who examined what they called 'back and forth shifts', we excluded from the analysis the first saccade to the subtitle area after subtitle presentation onset. In other words, we calculated how many times when watching a clip, participants were coming back with their eyes to the caption area of interest.

Overall comprehension

Edited captions were expected to have the highest comprehension scores as they were linguistically less

complex and had long display times, thus giving the viewers ample time to read and process the caption content. Verbatim subtitles, in contrast, were expected to render lower comprehension scores since they were deemed to be more demanding both in terms of linguistic content and short display times.

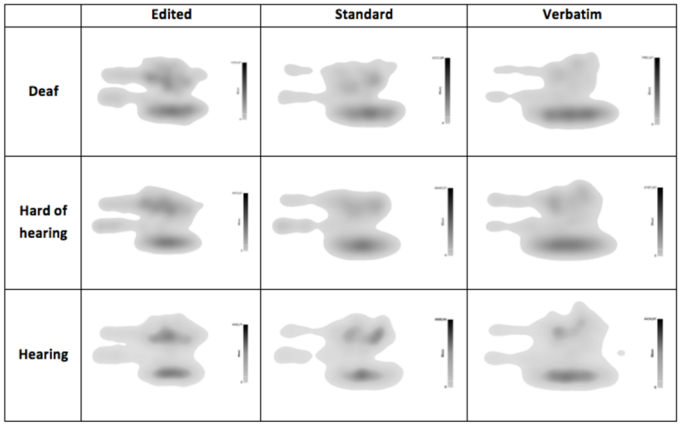

Figure 1. Heat maps based on fixation durations - presenting division of time between scene viewing and subtitles reading.

Results

The proportion of dwell time on caption reading vs. scene viewing

The first eyetracking variable related to the process of reading captions was the proportional index of time spent on reading subtitles and watching a scene. We examined the differences in the percentage of dwell time on captions between the three groups of viewers and the three types of captions. Our prediction was that group differences would be most prominent when watching the clip with verbatim captions, with deaf and hard of hearing participants spending more viewing time on reading subtitles, as they are generally considered in literature to be slower readers than hearing people.

First of all, in line with our expectations, deaf participants tended to dwell on captions significantly longer (M = 68.04, SE = 5.89) than hearing participants (M = 45.14, SE = 5.33), however only during watching the video clip with verbatim subtitles, $F(2,37) = 4.23$, $p < .05$ ($p = .022$), $eta^2 = .186$. Between-group comparisons for the other two types of captions were not statistically reliable. This result suggests a similar strategy of allocating attentional resources for all groups since the proportion of dwell time on captions was nearly the same, except for verbatim captions. To sum up, the results of dwell time scores confirmed our first hypothesis, expecting differences in reading time between three different captions, with higher proportion of dwell time on verbatim captions. Furthermore, the between-group comparisons demonstrated that deaf people spent more time than other participants on reading verbatim and standard captions, but not on edited ones. These findings suggest that edited captions are relatively the easiest to process for all groups of viewers. To understand the nature of longer dwell time on captions in the deaf group, a series of further analyses was conducted.

Fixation count

In addition to testing dwell time, fixation count was analyzed. A higher rate of fixations might suggest slower reading process and smaller amount of time devoted to viewing a scene. In line with our second hypothesis, we expected to observe relatively small between-group

differences for edited captions, suggesting this caption type is less cognitively demanding for viewers, and a relatively high number of fixations for verbatim captions. Deaf and hard of hearing participants also had higher mean fixation counts both per caption and per word than hearing participants. On average, they fixated 8.72 and 7.42 times respectively on each caption whereas the hearing had ca. 6.29 fixations per one caption. A similar pattern is seen in the fixation-per-word rate, where deaf participants needed more fixations to read one word (1.24 on average) than the hard of hearing (1.05) and hearing (0.89).

Deflections to captions

The final question we asked is whether the higher frequency of fixations for deaf participants is a consequence of more deflections between the caption area and the image, whereby the deaf are looking for better understanding of a scene (which slows down the reading process), or rather a result of their reading pattern.

A two-way mixed analysis of variance (ANOVA) was performed with deflection count to captions as a dependent variable and viewers (hearing, hard of hearing, deaf) as the between-subjects factor and caption type as the within-subjects factor (edited, standard, verbatim). The analysis did not reveal any significant results except for the main effect of captions, $F(2, 74) = 63.51$, $p < .001$, $eta^2 = .632$. All participants were moving their eyes more frequently to the captions area while watching the video clip with verbatim subtitles ($M = 229.71$, $SE = 9.83$), comparing to clips with edited or standard captions ($M = 164.38$, $SE = 6.01$, $M = 175$, $SE = 8.18$, respectively). A simple explanation of the difference is that in the verbatim clip there were more captions ($N = 32$) comparing to standard ($N = 24$) and edited ($N = 25$). To rule this out, the total number of deflections to the caption area was divided by number of captions in each of the three video clips and the analysis was repeated.

Overall comprehension

As mentioned earlier, participants were asked to complete three comprehension questions after each clip. A one-way ANOVA with mean comprehension accuracy for all questions (three for each video clip) as a dependent measure and viewers group as an independent factor showed no significant differences. There was a tendency for deaf participants to have higher accuracy ($M = 65.74$, $SE = 9.08$) than hearing ($M = 54.17$, $SE = 8.76$) and hard of hearing participants ($M = 61.46$, $SE = 6.37$), however it did not reach a significance level. This tendency may appear quite surprising, especially given the fact that deaf people are generally considered poorer readers than hearing people.

Another finding which ran contrary to our expectations was that the highest accuracy scores were attained in the case of verbatim captions and lowest in the case of edited captions. This may result from the characteristics of the visual scenes analyzed – the clips with verbatim and standard captions were slightly more static, whereas in the clip with edited captions there was more dynamic action to follow. However, a similar result was recorded in the study by de Linde and Kay (1999), where participants had lower comprehension scores for the clip with slower subtitles.

One possible explanation for the fact that deaf viewers scored slightly better than hard of hearing and hearing viewers in the case of edited captions may stem from the fact that – unlike hearing and hard of hearing people – deaf viewers

were not disturbed by the inconsistencies between the sound and the content of subtitles. They may also have been more attentive to reading captions as their only source of information on the dialogue.

Discussion

The first hypothesis predicted that differences in the proportional dwell time on captions between deaf, hard of hearing and hearing participants would be most pronounced for verbatim captions, compared to edited and standard captions. In general, the proportion of dwell time devoted to reading verbatim captions was longer for all participants in comparison with the dwell time on standard and edited captions, which is not surprising as they contained more text and had shortest display times. A significant two-way interaction showed, however, that deaf participants dwelled on verbatim captions significantly longer than hearing participants, but not more than hard of hearing viewers.

The second hypothesis tried to disentangle differences in reading strategies among the three groups of viewers and was tested in two separate analyses: using fixation count and deflections to captions as dependent measures. Again, we found significant differences among verbatim, standard and edited captions. Not surprisingly, in line with the previous analyses, all viewers had highest fixation rate on verbatim captions. The difference between standard and edited captions was also significant, confirming a linear trend. The analysis of the deflection rate allowed us to accept the assumption that deaf participants had a different way of information processing than the other two groups. The general effect of caption type was not quantified by differences between participants. In general, all participants made more deflections to verbatim and standard captions when compared to edited captions. The relatively high deflection rate for all participants informs us about attentional resources allocation. With this study in mind, we hoped to provide a simple recommendation for the use of edited, standard or verbatim captions. In the light of our results, it seems that when considering the proportion of dwell time on captions, the best choice would seem to be edited captions as they allow viewers to spend more time on watching the image and not only on reading. Standard captions, however, appear to be almost equally good in this respect, as it took the participants about 50% of time to watch the image and about 50% to read standard captions. A large disadvantage of edited captions seems to be that in spite of being read faster, they do not render high comprehension, as their processing may be hampered by discrepancies between the dialogue and the caption text. It therefore seems reasonable to promote standard captions as the optimum solution. In our study, standard captions did not differ significantly from edited captions in terms of favorable eyetracking measures. They gave viewers ample time both to read the text and look at the image (ca. 50%/50%).

References

[1] Romero Fresco, P. (2009). More haste than speed. Edited vs. verbatim respoken subtitles. *Vigo International Journal of Applied Linguistics* 6, 109-133.

[2] Neves, J. (2007). Of pride and prejudice. The divide between subtitling and sign language interpreting on television. *The Sign Language Translator and Interpreter* 1(2): 251-74.

[3] Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research.

[4] Duchowski, A. T. (2003). *Eye Tracking Methodology: Theory and Practice.* London: Springer-Verlag.

[5] De Linde, Z. and Kay, N. (1999). *The Semiotics of Subtitlin*g. Manchester: St. Jerome.

[6] D'Ydewalle, G., J. V. Rensbergen and J. Pollet (1987) Reading a message when the same message is available auditorily in another language: The case of subtitling. In: J. K. O'Reagan and A. Lévy-Schoen (eds.) *Eye Movements: From Physiology to Cognition*. Amsterdam / New York: Elsevier Science Publishers, 313-321.