Improving Video Captioning for Deaf and Hearing-impaired People Based on Eye Movement and Attention Overload

C. Chapdelaine, V. Gouaillier, M. Beaulieu, L. Gagnon[*]

R&D Department, Computer Research Institute of Montreal (CRIM),
550 Sherbrooke West, Suite 100, Montreal, QC, CANADA, H3A 1B9

ABSTRACT

Deaf and hearing-impaired people capture information in video through visual content and captions. Those activities require different visual attention strategies and up to now, little is known on how caption readers balance these two visual attention demands. Understanding these strategies could suggest more efficient ways of producing captions. Eye tracking and attention overload detections are used to study these strategies. Eye tracking is monitored using a pupil-center-corneal-reflection apparatus. Afterward, gaze fixation is analyzed for each region of interest such as caption area, high motion areas and faces location. This data is also used to identify the scanpaths. The collected data is used to establish specifications for caption adaptation approach based on the location of visual action and presence of character faces. This approach is implemented in a computer-assisted captioning software which uses a face detector and a motion detection algorithm based on the Lukas-Kanade optical flow algorithm. The different scanpaths obtained among the subjects provide us with alternatives for conflicting caption positioning. This implementation is now undergoing a user evaluation with hearing impaired participants to validate the efficiency of our approach.

Keywords: Eye tracking, motion detection, captioning systems, human-computer interaction.

1. INTRODUCTION

The aim of this paper is to report on an on-going research on eye tracking and attention overload of deaf and hearing-impaired people when viewing captioned videos, to obtain fixations and scanpaths information in order to implement a software tool for the production of smarter captioning.

Deaf and hearing impaired people rely on caption to be informed and enjoy television. Most captions are produced off-line ahead of broadcasting. But as live broadcasting increases there is a need for real-time captioning so that deaf and hearing impaired people can access all the information. Up to now, the constraint of live captioning imposed a lower presentation quality than off-line captions. Indeed, off-line captions are placed manually so that viewers can rapidly read what is being said or heard. For example, in a dialogue the caption of each person is placed directly beneath each one so that viewers can always know who is saying what. In real-time caption, the text is most often presented in a fixed region of three lines at the bottom of the screen. Another major difference with real-time caption is the unavoidable delay introduced by the time needed to translate the audio information into text. In the off-line case, the transcribed text is synchronized to the audio that is heard. This helps viewers to associate the visual content with the audio and enables efficient lips reading. For deaf and impaired people presentation quality of off-line caption over real-time is evident.

But this quality has a price; the actual production of off-line caption is a long process which can take up to sixteen times the program duration. On the other hand, real-time caption has a very short production cycle with no presentation considerations which could be greatly improved if it would be done like off-line caption. If the goal of higher quality caption presentation can be obtained in real-time, then it could also be introduced to reduce off-line captioning time. We believe that improving the quality of real-time caption and reducing production cost of off-line caption would render the television content more accessible and enjoyable to the deaf and hearing impaired community.

---

[*] langis.gagnon@crim.ca; phone 514 840 1234; fax 514 840 1244; www.crim.ca/vision

1.1. Visual attention strategies

In order to reach our goal, we needed to establish production rules that would ensure the efficiency of the produced captions. We took a human factor approach and seek our rules in the visual attention strategies of caption readers. Deaf and hearing-impaired people capture real-time information in television through visual content and caption reading. Those two activities require different visual attention strategies. There exists a huge body of literature on visual attention for reading and picture viewing (see, for instance, the overview of Rayner [17]). However, little is known on how caption readers balance viewing and reading and where would be the differences in the strategies of hearing-impaired and hearing population.

Research on cross-modality plasticity, which studies the ability of the brain to reorganize itself if one sensory modality is absent or lost, shows that deaf and hearing impaired people have developed more peripheral vision skills than the hearing people [2]. Moreover, Proksch and Bavelier [16] found that this greater resources allocation of the periphery comes at the cost of reducing their central vision. In their experiment, hearing participants (who all had the ability to sign) did not show this effect. Likewise, d'Ydewalle and Gielen [4] studied attention allocation with a wide variety of television viewers (children, deaf, elderly people). They concluded that this task require practice in order to divide attention effectively. These results suggest that visual attention strategies would be different among hearing-impaired who gained experience through usage compared to inexperienced hearing caption readers.

Furthermore, the work of Jensema [8-10] covers almost every aspect of caption reading. Studies spanned from the reaction to caption speed to the amount of time spent reading caption. But mostly, it is his eye movement study on visual attention patterns that is more pertinent to us. This study involved six participants (three deaf/impaired hearing and three hearing people) viewing video excerpts (of 8 to 18 seconds duration) with and without captions. Jensema [10] found that the coupling of captions to a moving image created significant changes in eye-movement patterns. In some cases, reading became the most dominant activity to the detriment of viewing. He also found preliminary evidence proposing that at high caption rate, reading is favored over viewing. However, he did not actually measure reading time spent on caption which makes it difficult to compare any results with our study. All his work suggests, when coupling reading and viewing activities, that eye tracking of visual attention can provide elements of the viewer's strategies. These in turn would evoke more efficient ways of producing captions and eventually develop a software tool for the production of smarter captioning. Then, our approach is to obtain the related visual attention data using techniques of eye tracking and attention overload paired with an information retention level test. .

1.2. Eye tracking

Eye tracking is used to identify the patterns of visual attention exhibited by viewers when attention is divided between visual content and caption reading. Attention overload is recorded to detect critical conditions when viewer's attention is saturated by the information to be processed. Eye tracking has proven its value in determining the correlation between gaze directions and informative regions of an image. Henderson [7] gives reviews of the research done since 1967 where human rating was matched to fixation density. Eye tracking also enables us to obtain scanpaths information.

We base our eye tracking analysis on specific *a priori* and *a posteriori* visual regions of interest (ROI) in the video. *A priori* ROI are defined either as static or dynamic. Static ROI are the two possible regions where one usually position captions (i.e. in the upper and lower parts of the screen). Dynamic ROI are either human face or moving objects. These ROI are manually identified in our video dataset. Those identifications are later used as the ground truth for automatic motion and face detection. *A posteriori* ROI emerge from the eye tracking data. These ROI are not anticipated in the dynamic *a priori* ROI (see Section 2.1).

The paper is organized as follows. Section 2 provides the technical background. Section 3 presents the methodology used in this study. Section 4 gives the results obtained so far from the analysis of ROI and scanpath data acquired from hearing-impaired and hearing persons. Finally, we conclude in Section 5 and discuss our future works.

2. TECHNICAL BACKGROUND

2.1. Fixation

Eye tracking is obtained using a pupil-center-corneal-reflection system. Gazepoint are recorded at a rate of 60 Hz. Systematic errors is controlled by calibration and on-going surveillance during experimentation. Data collected during experiments include the time and coordinates of the gazepoints. The time is given in milliseconds and the coordinates are normalized with respect to the size of the stimulus window. Given a quality index issued by the tracker, data is filtered to delete bad quality results due to occasional bad pupil detection by the system. These missing gazepoints are estimated by a cubic spline interpolation. Analysis of eye tracking data allowed measuring many different eye movement behaviors such as fixation frequency and duration, saccade frequency, velocity and smooth pursuits [5]. We restricted the scope of our analysis to fixation of both a priori and a posteriori ROI and to scanpaths comparison.

Fixations correspond to gazepoints for which the eye remains relatively stationary for a period of time, while, saccades are rapid eye movements between fixations. Fixation identification in eye tracking data can be achieved with different algorithms [18]. We use a dispersion-based approach in which fixations correspond to consecutive gazepoints that lie in close vicinity over a determined time window. In our analysis, duration threshold for fixation is set to 250 milliseconds. This concurs with values found in literature that range from 66 to 416 milliseconds [11,15,18] with an average length between 100 and 300 ms. Every consecutive points in a window of 250 ms are labeled as fixations if their distance with respect to the centroid corresponds to 0.75 degree of viewing angle. This dispersion threshold is within the range proposed by Salvucci and Goldberg [18].

In order to study the repartition of visual attention over captioned videos, we analyze shot-by-shot the fixation patterns of each participant. Firstly, in an attempt to discriminate between relevant and irrelevant visual content for viewers, ratios of hits on each *a priori* ROI are computed. A hit is defined as one participant having made at least one fixation in the specified ROI. We then identified *a posteriori* ROI, i.e. regions that received hits, but that are outside the *a priori* ROI. Secondly, the sharing of visual attention between captions and visual content is investigated. We accumulated the fixations falling within dynamic ROI versus caption region and computed the ratios of total fixation time in caption region to shot duration.

2.2. Scanpath

In the study of participants' scanpaths, only fixations are retained to compose fixation sequences. A string-edit technique [12] is used to compare these sequences. An ASCII character is assigned to each ROI. Fixation sequence for each participant is then translated into a string where each character corresponds to the ROI in which the fixation occurred. Fixations outside all ROI are assigned the NULL character. Similarity between the resulting coded sequences is measured by the Levenshtein distance. This metric computes the smallest number of insertions, deletion and substitutions necessary to transform one sequence into another. We normalize the distance between each pair of coded sequences by the length of the longest. Distance computation is achieved with an Optimal Matching Analysis technique.

Our goal through this analysis is to determine if deaf and hearing-impaired people share a common viewing strategy that would provide more precise guidelines caption positioning. However, in our experiment, due to the relatively small number of viewers for each video, clustering techniques cannot be applied. To overcome this difficulty, we have chosen to compare participant fixation sequences based on their similarity to reference sequences. As reference, we chose the scanpaths of the participants who demonstrated the "best" and the "worst" viewing strategies based on the results of the information retention test. Each participant fixation sequence is ranked according to their distance to these two reference sequences. The results are analyzed to determine if deaf and hearing-impaired fixation strategy tend to align with one of these representatives.

2.3. Motion activity for automatic detection of dynamic ROI

Dynamic ROI are defined as regions in video frames where high motion activity is recorded or where faces are detected. In this study, the dynamic ROI are manually positioned. They also serve as ground truth for the implementation and test of a process that would automatically detect dynamic ROI. Usually, captions should avoid those regions because they

might hide important information to the viewer. Our approach consists then to produce a map where captions should not be allowed to be put on. Figure 1 gives a block diagram of our implementation.
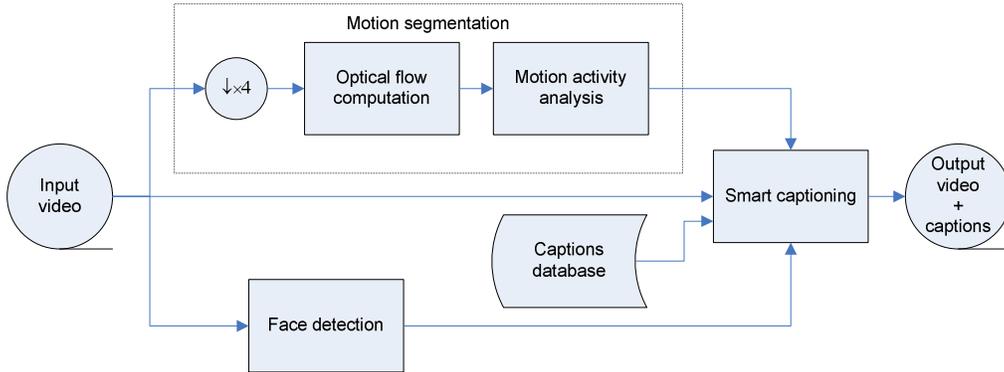

Figure 1: Block diagram of the ROI detection implementation

The map is computed by combining the output of two algorithms. A motion detection algorithm based on the Lukas-Kanade optical flow techniques is applied on the input video stream [13]. Optical flow is computed between two frames at each pixel, giving the velocity magnitude, segmentation in blobs is done. Blobs with large average magnitude are then labeled as regions where captions are *non-grata*. In parallel, faces are detected through a classifier trained with Gabor responses at different scales and orientations of face images. The accuracy of the classifier is improved by combining with a particle filter time tracking of the detected face [19]. This algorithm returns ROI also labeled as *non-grata*. At the end of the process, we combine both outputs in one map to generate the possible locations for a caption. We select the most appropriate one given the context of the video. The caption is finally encoded in the output video stream. The overall system is implemented as plug-in filter of the open-source video editing tool VirtualDub.

3. METHODOLOGY

3.1. Video stimulus (data)

In a prior experiment done on French caption rate [3], debriefing with hearing-impaired participants showed that the difficulty is not just on caption speed but also on motion in image. In fact, news with intense visual coverage such as riots and war scenes, action sports and action film were identified as more challenging content. Those findings motivated us to design an experiment with varying caption rates and also with different motion level. The video stimulus corpus contains 6 source videos of various types with caption and without audio. The length of each video is between 2 to 4 minutes. They represent a large variety of television content as shown in Table 1. For each type, 2 videos are taken from the same source with equivalent criteria. The selection criteria are based on the motion level they contain (high or low according to human perception) and their moderate to high caption rate (100 to 250 words per minute). For each video, a test is determined to measure the information retention level on the visual content and on the caption.

Table 1. Description of the video stimulus

Id.	Type	Motion level	Caption rate	Total nb. shots	Length (frame)	Mean (frame/shot)	Nb. Part.	Impaired	Hearing
video 1	01- Culture	Low	High	21	4037	191,8	5	2	3
video 2	03- Film	High	Moderate	43	5645	143,6	5	1	4
video 3	04- Film	High	Moderate	73	6789	106,8	8	5	3
video 4	05- News	Low	High	32	4019	131,0	8	4	4
video 5	07- Docum.	Low	Moderate	11	4732	442,3	7	3	4
video 6	09- Sport	High	High	10	4950	515,8	6	3	3
				190	**30172**	**158,8**	**39**	**18**	**21**

3.2. Apparatus

The following hardware and software is used to conduct the experiment:

- ViewPoint EyeTracking hardware and software with a head stabilizer
- Video stimulus projected on a wall with a resolution of 1024x768 pixels (viewer at 3.25 meters form the wall
- 1 PC hosting a homemade C++ software (VDPlayer) to start, control and synchronize eye-tracking, data acquisitions, external equipment and display stimulus.

3.3. Participants

The reported experiment here is part of a wider study conducted on 18 people (9 hearing and 9 hearing-impaired). Among them, 7 hearing-impaired and 8 hearing participated to the eye tracking analysis. We excluded deaf people communicating mainly with sign languages since eye-tracking has to be done with the lights off and we could not communicate with them in the dark. We also excluded people with bifocal glasses since their gaze could not be accurately registered by the apparatus. In all our analysis, the hearing-impaired (IMP) viewers are considered the experienced viewers from whom best strategies can be derived. The results from the hearing viewers (HEA) are used to discriminate against a potential behavior that would not be part of the best strategy. In such case, any diverging behavior from the IMP compared to the HEA viewers will be interpreted as a guideline for the best strategy.

3.4. Procedure

The experiment was conducted in two parts. Part one: all participants viewed 5 videos and are questioned about the visual and caption content in order to assess information retention. Questions are designed so that reading the caption could not give the answer to visual content questions and vice versa. Part two: when participants are wearing the eye tracking glasses, calibration is done using a 30 points calibration grid. Then, all participants viewed 5 new videos. The viewing is also combined with attention overload detection. In this part, no questions are asked between viewing to avoid disturb participants and disrupt calibration.

4. RESULTS

Results presented here are from a first level of analysis which aims at answering those following questions

- Is caption reading favored over viewing and what is different for IMP viewers?
- Do selected a priori ROI really draw visual attention of viewers and are some of them neglected?
- Do viewers look at ROI that we did not anticipated?
- Are the scanpaths of the impaired viewers similar among themselves and do they differ from the scanpaths of the hearing viewers?

4.1. On caption reading

Did participants spend more time reading than viewing as observed by Jensema [10] ? We found that viewers spent 10% to 31.8% of their time reading captions (Figure 2). Participants spent less time in reading captions on video with lower caption rate (e.g. video 2 and 3) but not for the video 5. This could be partially explained by the fact that since motion is low, the participants may have preferred spending time on caption.

In the case of video 6 with high caption rate and high motion level, participants have the lowest percentage on caption (19.1%). This could indicate that for sports, viewing is preferred to reading even if caption rate is high. The high results of video 1 and 4 are aligned to Jensema findings, higher caption rate tends to increase fixations on caption and increase time spent on reading.
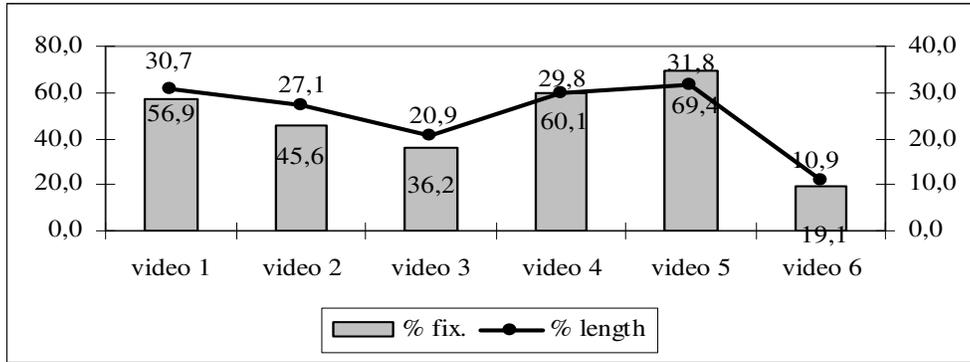
Figure 2: Percentages of fixations on caption and duration

Comparing the results (Figure 3) among the hearing impaired (IMP) and hearing (HEA) groups show that percentage and time of fixations of caption is significantly less for IMP than HEA (except for video 3). This suggests that IMP do fewer fixations but that they are highly informative since they score the best in the information retention test. In video 3, both spent the same amount of time on caption. Looking at the scanpaths, we found that, in black and white shots, some of the IMP viewers did not read caption or even scan the image. They simply stared at the center of screen. The effect of black and white visual content would require further investigation since we also found that *a priori* ROI in black and white shots are often ignored.
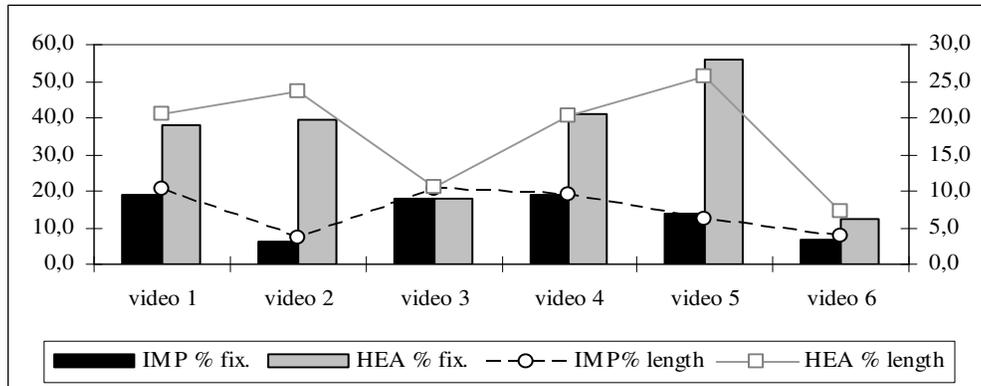
Figure 3: Fixations on caption - Comparison between Impaired (IMP) and Hearing (HEA)

4.2. On validating *a priori* ROI

The stimulus data set included a total of 297 a priori dynamic ROI manually identified. The total of potential hits (a hit is at least one fixation in ROI) is obtained by multiplying the number of ROI in each shot by the number of participants who viewed it. A total of 790 actual hits (AH) are found on the 1,975 potential hits (PH) (Table 2).

Table 2. Potential and actual hits per video.

	Nb. ROI	PH	AH	%
video 1	55	275	100	36,4%
video 2	64	320	102	31,9%
video 3	80	640	341	53,3%
video 4	70	560	136	24,3%
video 5	12	84	47	56,0%
video 6	16	96	64	66,7%
Total	**297**	**1975**	**790**	**40,0%**

In some cases our *a priori* ROI are good predictors of visual attention. For instance, in video 2, 3, 6 ROI received more than 50% of the AH. Nevertheless, we had a poor ROI selection in some instances. For example, moving face or objects blurred by speed are ignored by most participants. In the case of video 4, only 24,3% of AH is observed, a more detailed analysis of the data revealed that when many faces are presents they are ignored by most participants (mainly HEA) who keep reading the caption (Figure 4). Then, we compared results for IMP and HEA to detect if this could be a strategic choice for IMP.
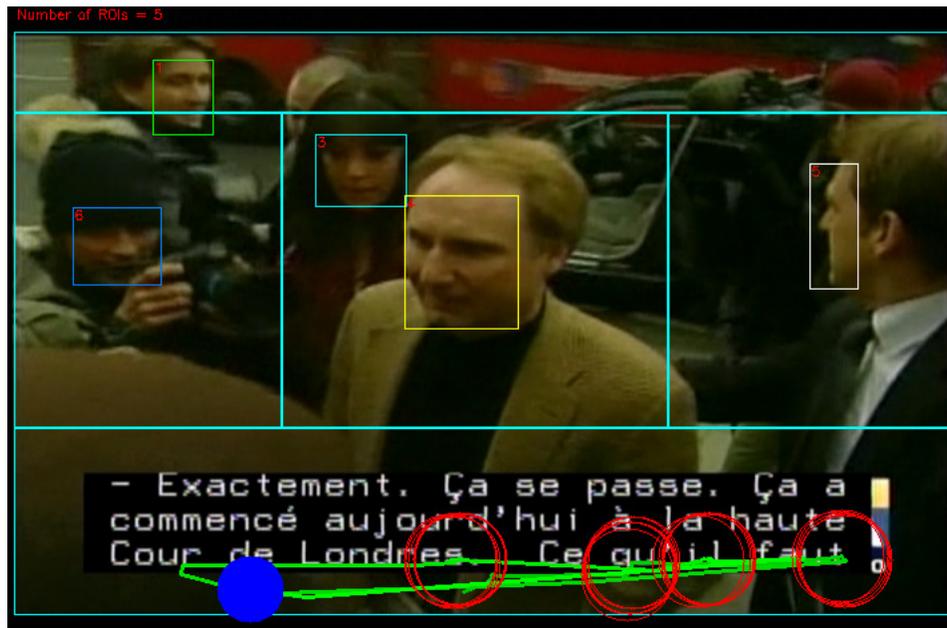
a) Discarded faces by all HEA
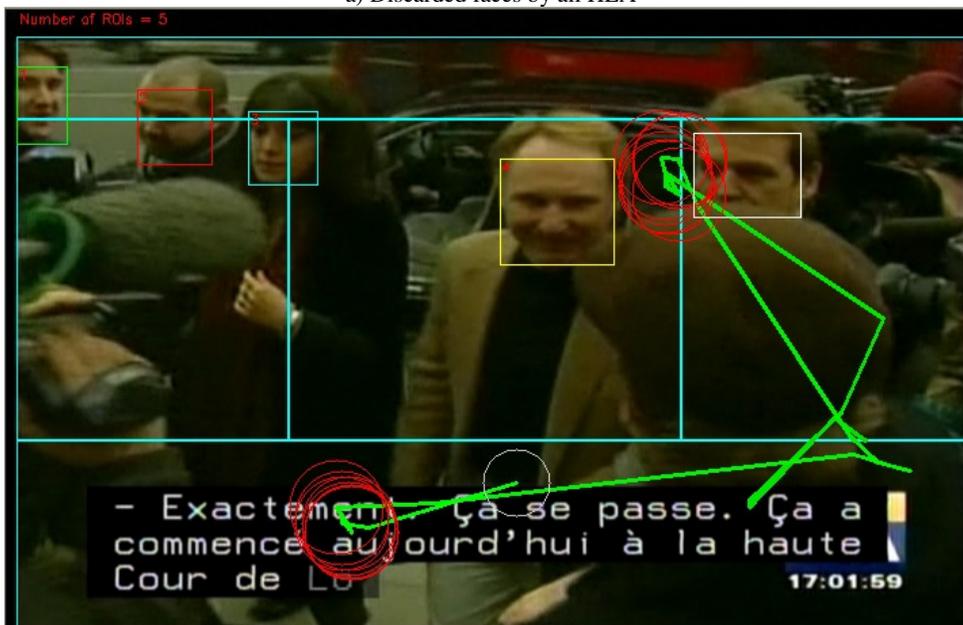
b) Actual hits (AH) by IMP

From video 4: ROI are the color rectangles, the green line is the scanpath for the shot, the red circles are past fixation in the shot and blue circle is the actual fixation in the frame.

The comparison indicates that most of the time, ROI are hit less often by IMP than by HEA. However, when IMP has more hits, it is significantly more as we observed in video 2 and 4 (Figure 5). Detailed analysis of tracking data for video 4 reveals that multiple faces are given attention by IMP and not by HEA. Thus, this confirms that multiple faces detection is a valid approach for a dynamic ROI detection algorithm.

In addition, a detailed analysis of video 1 shows that the faces of the news anchors, which are seen several times in prior shots, are ignored by IMP in latter shots. The same evidence is also found on other video are close-up images are more ignored by IMP than by HEA. This suggests that IMP rapidly discriminate against repetitive images that would not bring further information.
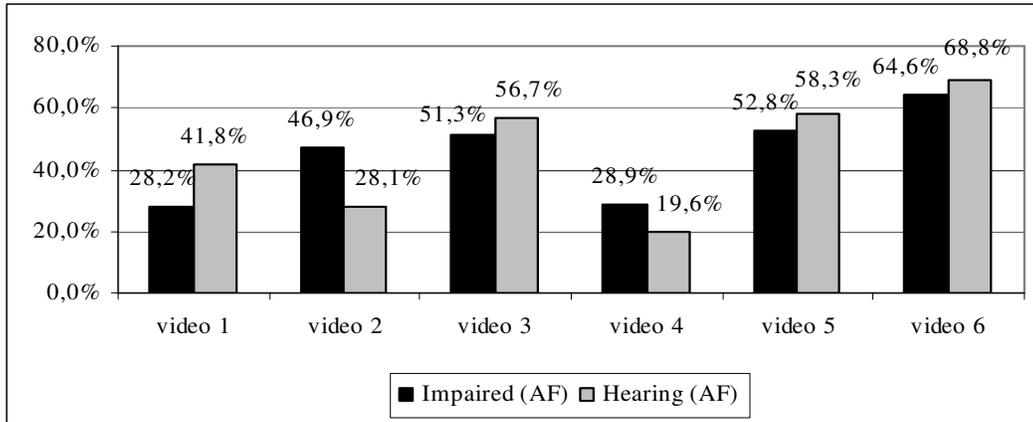
Figure 5: Percentage of hits on a priori ROI - Comparison between Impaired (IMP) and Hearing (HEA)

4.3. On finding *a posteriori* ROI

This analysis is based on the percentage of fixations per region on a screen divided in 10x10 regions that fall outside the dynamic ROI and caption area. The total number of regions hit among the 100 is also used to calculate a coverage ratio.

We observe (Figure 6) that 27% of fixations is done on *a posteriori* ROI with a coverage ratio of 29,9% of the region hit. As shown in Figure 6, *a posteriori* ROI are the foremost focus of visual attention (more than 50%) in video 6 (hockey). We had identified mostly the hockey disk as a dynamic ROI, but in fact, participants look more at the players. This suggests that ROI in sports may not be the moving object but the players (not always moving) who can become active in the game.
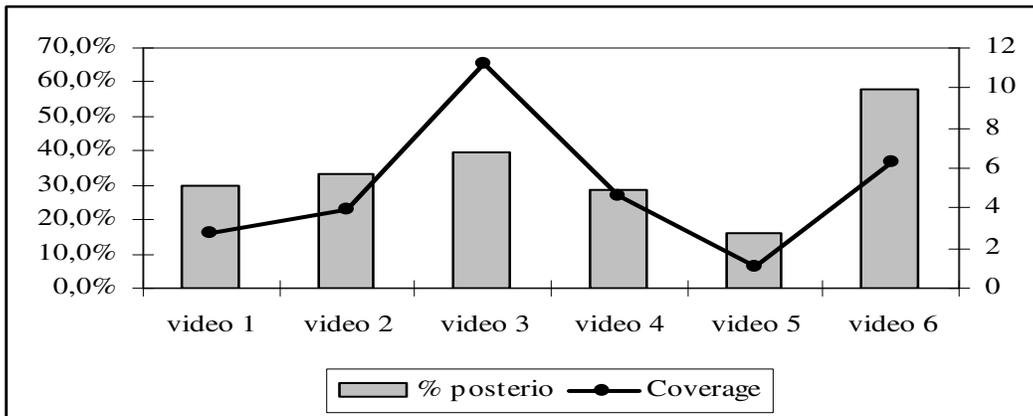
Figure 6: Percentage of fixations in posteriori ROI and coverage ratio

Comparing fixations (Figure 7) on *a posteriori* ROI between IMP and HEA viewers, we observe that IMP tends to look more outside *a priori* ROI than HEA. Since most of the time the coverage area is below 4 %, this suggests that they do fixations on specific area. A detailed analysis enables the identification of several ROI that are missed as *a priori* but that we should include in our detection algorithm. For instance, the gaze of IMP viewers is attracted to a lot more moving objects than expected. Also, they seem more proactive in video viewing then HEA, as if they are always seeking for potential source of information. For example, 1) in sport videos, they look at more players than HEA viewers almost as they are anticipating the action, 2) in shot with no caption, they look at more objects in the image than HEA who tend to stare at one point and 3) in shot with no action, the same behavior is observed, IMP scan the image while HEA stares.
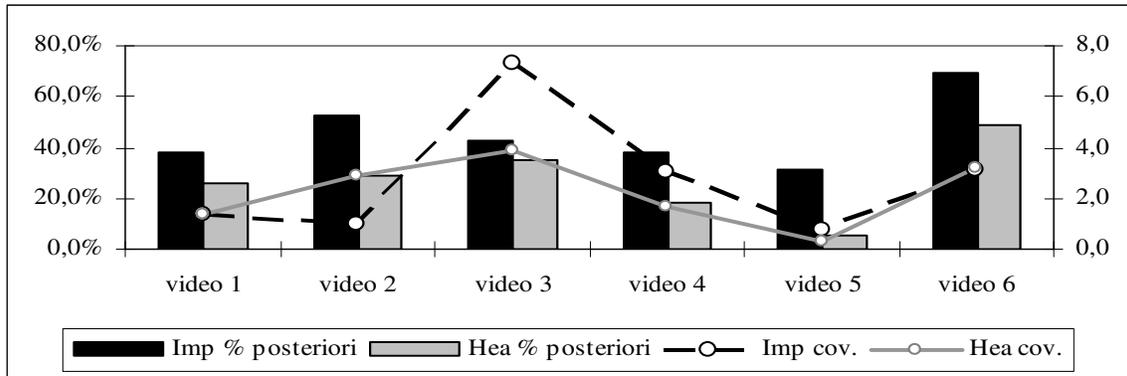

Figure 7: Fixations on posterior ROI - Comparison between Impaired (IMP) and Hearing (HEA)

4.4. On scanpaths comparison

As mentioned above, we postulate that better retention results from better viewing strategies. The scanpaths of the best and the worst viewer are taken as an arbitrary reference against which to compare all the scanpaths for differentiating between IMP and HEA viewing strategies.

Each scanpath is ranked against the best and the worst scanpath for each video. In video 1 (Figure 8) IMP are ranking closer to the best scanpath 70% of the times while they score closer to the worst in only 30% of the times. HEA often score closer to the worst (60%). This suggests that for this video, viewing strategies of IMP and HEA form separable classes based on their similarity to the reference participants. The same conclusion can be reached for video 2, 4 and 6.

However for video 3 and 5, the scanpaths of IMP are scored farther from the best and closer to the worst. This could mean that some IMP individuals are scoring closer to the worst or that some HEA are scoring closer to the best. This could also imply that our choice of reference scanpaths is not efficient for these video. Further analysis of scanpaths would be needed.
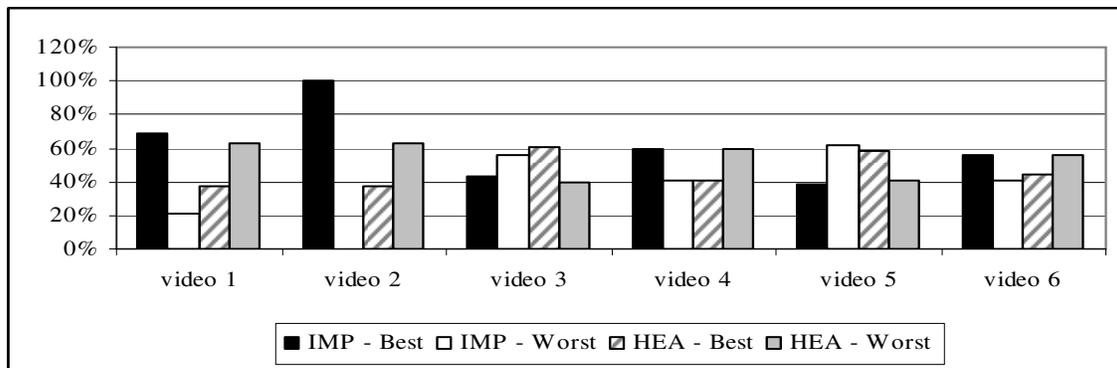

Figure 8: Comparison between Impaired (IMP) and Hearing (HEA)

5. CONCLUSION

We investigated on caption video viewing strategies of deaf and hearing impaired using eye tracking analysis.. Our study enables to answer our research questions and elaborate guidelines for better captioning.

- **On caption reading**: Impaired viewers had different strategies not only for reading caption but also for watching the visual content. We found that they spent significantly less time reading caption than hearing viewers and time allocated would vary not only upon caption rate but also motion level in images. So any assessment made by hearing human on caption rate while captioning may be inaccurate if based on reading speed only. The solution would be to build a deaf and hearing-impaired reading model coupled with motion level detection.

- **On validating *a priori* ROI**: We expected faces and moving objects to be ROI. Our results confirm that most of the time they draw visual attention. Actually, our preliminary results indicate that impaired viewers will look more at faces than hearing viewers. However, this is not the case if the faces have been shown before or in close-up. This suggests that face recognition should be added to face detection. For example, caption should be placed beneath the person talking but not for a face recognized in a previous shot. In that case, caption should remain in place to support fast caption reading while no image viewing is necessary. Further study on close-up on faces and moving objects would be needed to determine when they attract less interest. In the case of sports, motion detection may not be the best way to determine potential ROI of viewers. Further study on sports contents would be required to assess if smarter caption is possible in this case. We also noted that even though deaf and hearing-impaired people had described hockey as full of action and more challenging to watch when we apply motion detection on video 6, the obtained motion level is lower than the ones for video 2 and 3. The low level can be explained by the fact that shots were mostly long camera panning views making changes from frame to frame not very distinctive. So it seems the perception of action in this case could not be measure by this technique of motion detection.

- **On finding *a posteriori* ROI**: Since impaired viewers seem to continuously scan the screen for potential information, much more regions of interest could be defined. This suggests that positioning caption closer to ROI could facilitate reading but as number of ROI increases in a scene, there would be a need for clustering the ROI to find the best location.

- **On scanpaths comparison**: The scanpaths of all impaired viewers share similarities with the proposed best scanpath as opposed to hearing viewers. This reinforces the notion that impaired viewers have distinctive visual strategies and further study would most certainly foster smarter captioning guidelines.

These findings are being implemented in a computer-assisted captioning software in order to test an automatic presentation technique for real-time captioning. This smart captioning software uses a motion and face detection techniques. This implementation is now undergoing a user evaluation with hearing impaired participants to validate if viewing and reading is easier and more efficient. This will be reported in a near future.

ACKNOWLEDGEMENTS

REFERENCES

1. T.T. Blackmon, Y.F. Ho, D.A. Chernyak, M. Azzariti, L.W. Stark, "Dynamic scanpaths: eye movement analysis methods", Proc. Human Vision and Electronic Imaging IV, Bernice E. Rogowitz; Thrasyvoulos N. Pappas; Eds. (SPIE- Vol. 3644), pp. 511-519, 1999

2. R.G. Bosworth, K.R. Dobkins, "The effects of spatial attention on motion processing in deaf signers, hearing signers and hearing non signers", Brain and Cognition, 49, pp 152-169, 2002

3. C. Chapdelaine, M. Beaulieu, L. Gagnon, "Vers une synchronisation du sous-titrage en direct français avec les mouvements de l'image", Proc. of the 18th international conference on Association Francophone d'Interaction Homme-Machine, pp.277-280, 2006

4. G. D'Ydewalle, I. Gielen, "Attention allocation with overlapping sound, image and text", In Eyes movements and visual cognition", Springer-Verlag, pp 415-427, 1992

5. A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*, Springer-Verlag, London, UK. 252 p., 2003

6. M. Fayol, J.E. Gombert, "L'apprentissage de la lecture et de l'écriture", Dans J.A. Rondal et E. Espéret (Eds.), Manuel de Psychologie de l'enfant.Bruxelles: Mardaga. 1999, pp. 565-594

7. J. Henderson, P.A. Weeks A. Hollingworth, "The effects of semantic consistency on eye movements during complex scene viewing", Journal of Experimental Psychology: Human perception and Performance, Vol 25, No 1, pp 210-228, 1999

8. C. Jensema,"Viewer Reaction to Different Television Captioning Speed*"*, American Annals of the Deaf, 143 (4), pp. 318-324, 1998

9. C. J. Jensema, R. D. Danturthi, R. Burch, "Time spent viewing captions on television programs" American Annals of the Deaf, 145(5), pp 464–468, 2000

10. C. J. Jensema, S. Sharkawy, R. S. Danturthi, "Eye-movement patterns of captioned-television viewers", American Annals of the Deaf, 145(3), pp. 275-285, 2000

11. S. Josephson, "A Summary of Eye-movement Methodologies", http://www.factone.com/article_2.html, 2004.

12. S. Josephson and M. E. Holmes, "Clutter or content? How on-screen enhancements affect how TV viewers scan and what they learn", Proc. of the 2006 symposium on Eye tracking research & applications, pp. 155-162, 2006.

13. B. Lucas, T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", Proc. of 7th International Joint Conference on Artificial Intelligence, pp. 674-679, 1981

14. K. A. Peker, "Framework for measurement of the intensity of motion activity of video segments", Journal of Visual Communications and Image Representation, Vol. 14, Issue 4, December 2003

15. A. Poole, L. J. Ball, "Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects", Chapter in C. Ghaoui (Ed.): Encyclopedia of Human-Computer Interaction. Pennsylvania: Idea Group, Inc., 780 pages, 2005.

16. J. Proksch, D. Bavelier, "Changes in the spatial distribution of visual attention after early deafness", Journal of Cognitive Neuroscience, 14:5, pp 687-701, 2002

17. K. Rayner, "Eye movements in reading and information processing: 20 years of research", Psychological Bulletin, volume 124, pp 372-422, 1998

18. D.D. Salvucci, J.H. Golberg, "Identifying fixations and saccades in eye-tracking protocols", In Proc. of Eye Tracking Research and Applications Symposium, ACM Press, New York, pp. 71-78, 2000

19. R.C. Verma, C. Schmid and K. Mikolajczyk, "Face Detection and Tracking in a Video by Propagating Detection Probabilities", IEEE Trans. On PAMI, Vol. 25, No. 10, 2003