International Association of Scientific Innovation and Research (IASIR)
(An Association Unifying the Sciences, Engineering, and Applied Research)

ISSN (Print): 2279-0063
ISSN (Online): 2279-0071

International Journal of Software and Web Sciences (IJSWS)
www.iasir.net

Investigation of Regions of Interest for Deaf in Czech Sign Language Content

Zdeněk Švachula, Martin Bernas, Petr Zatloukal, Lucie Klabanová[1]
Dept. of Radioelectronics, Faculty of electrical engineering
Czech Technical University in Prague
Technicka 2, 166 27 Prague 6
Czech Republic
[1]Institute of Czech Language and Theory of Communication, The Faculty of Arts
Charles University in Prague
Nam. Jana Palacha 2, 116 38 Praha 1
Czech Republic

Abstract: *Tasks such as displaying optional television sign language interpreter or synthesis of sign language assume fine resolution while maintaining low bit rates. This paper describes the determination of important regions of interest for deaf and hard of hearing when watching a sign language speaker or interpreter in television content. Proposed results are useful for selecting an appropriate compression scheme for coding both real and virtual sign language interpreters, and for virtual sign language interpreter modeling as well.*

Keywords: *Region of interest; Czech sign language; eye tracking; fixation; finger alphabet*

I. Introduction

The Coding of image sequences with sign language (SL) interpreter, as well as modelling and animating of SL interpreter avatar, requires good knowledge of how important different body parts of interpreter are for understanding and congenial perception. SL is visual language and each sign consists of four manual components (hand configuration, place of articulation, hand movement, and hand orientation) and non-manual components like facial expression and body posture [1]. Likewise as eye contact is important during speech, also when using SL, people do not watch only hands and also perceive face. If we proceeded from the hypothesis that the face is perceived the most of viewing time and the articulation of Czech sign language is seen only peripherally, it is also interesting to find out whether this blurred vision negatively affect the understanding of the watched expression.
A British research group worked on a task of efficient video coding with the use of visual regions of interest (ROI) in British Sign Language [2, 3]. They found the head as the primary ROI observed most of the time. Emmorey et al. [4] investigated fixations while comprehending a short narrative and a spatial description in American Sign Language and found that deaf native signers tended to focus on or near the signer's eyes, whereas beginning signers fixated more on the mouth. Considering Czech Sign Language (CSL), there are no previously published results in this area. We managed two experiments with CSL users to investigate their ROIs, when watching TV content with a CSL interpreter or speaker. The difference compared to previous researches is in covering of all utilized methods, where signer can be situated on the screen – size and position, and in investigating of whether inclusion of Czech finger alphabet in signed content can influence eye movements. Further we use the word speaker for the content primary prepared in CSL and interpreter when translating from Czech.
The aim of this study is to find out, how is the visual attention of deaf divided between signer's face and hands, while play the role of CSL TV content observers, and how is affected by using of finger alphabet in this content.
.

II. Eye Movement Analysis

A. Eye Tracking System

We used the eye tracking system ViewPoint Eye-Tracker® to capture eye movements. This system is binocular and belongs to the video-oculographical techniques extracting eye parameters from video recording. Participants must therefore wear special glasses with infrared light source (IR diode) and a small camera for both eyes. Infrared light is necessary to ensure clear distinguishable pupil in the captured picture. This system does not provide a point of regard (POR) measurement. It measures eye position relative to the head and requires the head to be fixed to keep the detected position of the eyes and POR identical [3]. To prevent head movement observers were asked to rest their elbows on the table and then put their head in their hands.
The cameras' resolution was set to 640x480 pixels with frame rate 30 fps. In each frame, dark pupil is detected as a rotated ellipse and x- and y- coordinates of its centre is mapped according to a calibration grid to the point of regard and saved into a text file together with time mark and other features such as pupil size. Coordinates are relative to upper left monitor corner as (0, 0). Bottom right corner acts as (1, 1). We processed only coordinates and total time marks. Calibration is performed as a number of predefined calibration points shown to the participant to cover the whole monitor area. We used 16 points to capture the calibration grid of each observer.

B. Method of experiment

In the first part of the experiment 19 participants (aged from 20 to 84) watched a special video sequence in standard definition resolution (720x576 16:9) designed to cover all basic scene compositions with signer on TV. We selected a viewing distance of about 60 cm to ensure the appropriate conditions for accurate measurement at 20-inch screen. As shown in Fig. 1, four short clips (about 30 seconds each) were comprised to include interpreter in front of neutral background, speaker in front of a graphical background with picture and text, framed interpreter in bottom right corner in Sama doma magazine and non-framed interpreter appearing only during speech in fairy tale. The magazine and fairy tale served as dynamic clips with cuts and changes in scene composition or camera angle. Observers were asked to watch the video sequence in a common way, as if they were watching television at home.

Figure 1 Still frames from clips in the testing video sequence.

Figure 2 Still frame from the video sequence with finger alphabet

The second part of the experiment involved 7 deaf volunteers. Presented video sequence of approximately 1 minute was visually the same as the CSL interpreter in front of neutral background in the first part of the experiment (Fig. 2). The sequence was downloaded from the website of Czech sign language centre Pevnost, as a content already prepared for another purpose, and contained four examples of Czech finger alphabet: WFD (2x), OSN and Frontrunners.

All clips were spaced out with a 5s neutral grey background with a white cross mark in the middle. The mark provided the initial visual point for all participants of the experiment, as well as a way to detect the head movement. Observers tended to move slightly their head either horizontally or vertically or both. Rotational motion was rare. Assuming translational motion, it can be compensated using error vector between centroid of filtered captured coordinates (k-means clustering with points detected as fixations as described below) and middle of cross mark with coordinates (0.5, 0.5). This can be applied for small deviations, because conversion of pupil position to POR is not a linear operation due to the curvature of the eye.

C. Fixation detection

Eye movement signal is recorded at a uniform sampling rate, so the velocity detection method can be applied [6, 7]. Successive samples are investigated to estimate eye movement velocity *v* according to the equation

$$v = \frac{\sqrt{\left(x_{t+1} - x_t\right)^2 + \left(y_{t+1} - y_t\right)^2}}{dt}, \qquad (1)$$

where *x* and *y* are coordinates of point of regard in time *t*.

At first, all values out of ‹0, 1› interval were removed from the captured data. Then velocity criteria was calculated and compared with the velocity threshold set to 30 pixels per frame, which is in our case approximately 1.5° of visual angle as the size of foveola, the region with the highest visual acuity in the eye. All miniature movements appearing during fixation (microsaccades, drift or tremor) are considered to be within this angle. We used the statistical minimum duration of the fixation 150 *ms*, as the temporal threshold to detect a single fixation [5]. Only fixations were used for the further investigation of ROIs.

D. Processing of results

It was found difficult for some observers to trace the centers of calibration marks. Some participants moved their head excessively. Despite proper lighting and camera settings, the system was not able to detect the pupil correctly. In some cases rapid oscillation between the pupil and other dark part of observer's eye appeared in the signal (usually the corner of eye), depending on the shape of the eye, the length of eyelashes, and other aspects. For this reason, only the results of twelve observers were used to draw the eye positions during fixation.

III. Results

In Fig. 3–7 we propose horizontal (x) and vertical (y) coordinates of the captured points of regard within the detected fixations in time in the first experiment. In Fig. 6 we propose vertical coordinates of the captured points of regard in clip with finger alphabet. The zone between the dotted lines represents an area where the head of the interpreter is located during entire clip. Grey fields in Fig. 6 represent parts of the fairy tale with the interpreter hidden. Grey fields in Fig. 7 illustrate parts where finger alphabet is used. Regarding the horizontal position of point of regard, coordinate zero represents left side, while value one the right side of the monitor. Regarding the vertical position, coordinate zero represents upper area, while value one the bottom area.

Figure 3 Position of points of regard for clip 4 (CSL interpreter in front of a neutral background).

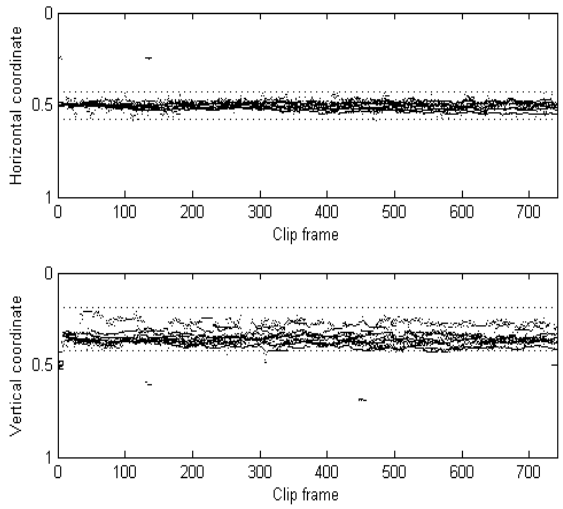


Figure 4 Position of points of regard for clip 1 (CSL speaker in front of a graphical background).

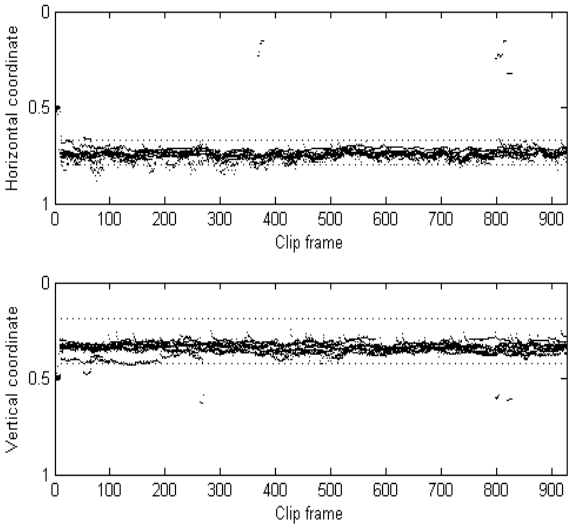


Figure 5 Position of points of regard for clip 2 (framed CSL interpreter in bottom right corner in magazine).

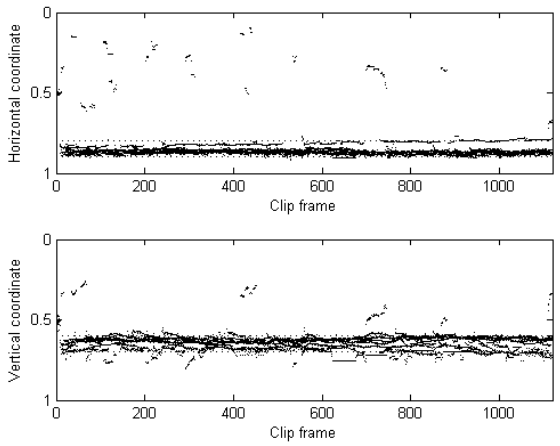


Figure 6 Position of points of regard for clip 3 (non-framed interpreter appearing only during speech in fairy tale). Grey fields represent parts without interpreter.

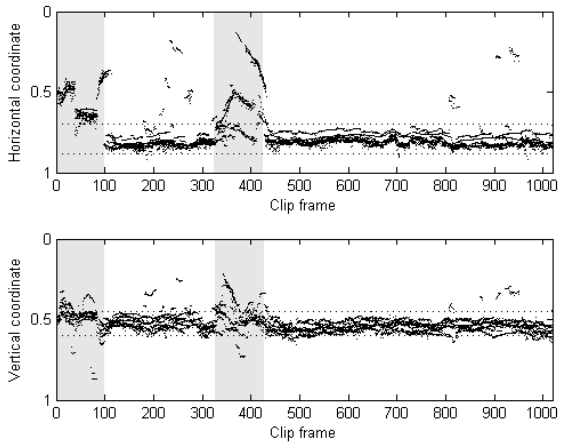
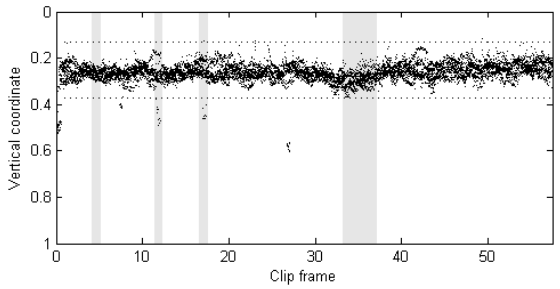


Figure 7 Position of points of regard in the finger alphabet experiment. Grey fields represent parts where Czech finger alphabet was used.



IV. Discussion

The first finding of our study is the confirmation of head as the most important visual ROI. In all cases participants spend the most of the time watching head of the signer instead of hands or even the rest of picture in the magazine or tale case. Considering the interpreter as the only object in the presented scene (Fig. 1, clip 1), observers tended to watch the area defined with dotted lines where the face of the interpreter was situated. Only in two cases participants moved their focus from the head to the hands (Fig. 3). Fixations are centred on or slightly beneath the signer's nose. 97.6 % of points of regard (POR) during fixations in average are in the area where the head of the interpreter is located (see Table I). Hands are perceived with parafoveal and peripheral vision turning attention while keeping the POR on the head (as illustrated in upper picture in Fig. 8).

Table I Number of points of regard during fixations, which are directed at the head of the interpreter.

Clip 1	Clip 2	Clip 3	Clip 4
97.6 %	94.6 %	80.8 %	72.0 %

Similar results can be seen when the signer is located in front of a graphical background (Fig. 1, clip 2). Only two observers moved their POR to parts of the background where a picture with text were placed (Fig. 4). Now 94.6 % of POR during fixations are in the area where the head of speaker is located with the centre on or slightly beneath the signer's nose as well.

Even if the interpreter is small and located in the right bottom corner of the television image (Fig. l, clip 3), still 80.8 % of POR during fixations are on the facial area. Some observers briefly looked at the moderator's face after a visual cut (Fig. 5). The resolution of cameras and system considering the head of the signer is not as precise as in the previous cases, but we can still claim, that visual attention is focused on the nose area.

More interesting results were obtained from clip where the non-framed interpreter was appearing only during speech. Grey areas in Fig. 6 represent parts of a clip with a hidden interpreter. PORs are distributed almost uniformly towards important objects, mostly the faces of actors. Once the interpreter is shown on the screen, the focus of all observers immediately fixed his face (still the nose area) and changed occasionally when some more fundamental action occurred in the television image.

Differences between the perception of the interpreter in the corner of the screen or in the full view show pictures in Fig. 8. Foveal (direct) and macula (parafoveal) perception areas are calculated according to the equation

$$S = \tan\left(\frac{A}{2}\right) 2D, \qquad (2)$$

where S is the real subjects' size, A is the spatial angle and D = 0,6 *m* is the viewing distance from the screen. For the 5° fovea spatial angle and 16.7° macula spatial angle we get the real diameters on the screen 5.2 *cm* and 17.6 *cm* respectively (considering screen size of 40.9x30.9 *cm*).

Figure 8 Pictures from a clip with CSL interpreter in front of a neutral background (a) and Sama doma magazine (b), both with areas of foveal and macula visions with the centre according to average results.

(a)

(b)

CSL interpreter in the corner of the screen seems to be perceived better because more information can be processed with the observer's direct vision. However size and resolution of the interpreter's body is lower. Participants in the experiment spent most of the time watching the facial area, while hands were perceived only with peripheral vision. Practically identical results emerged from all clips, no matter what the size or location of the interpreter in the scene was. In further developing a proper compression methods or sign language synthesis, the facial area of interpreter (as the most important visual ROI) must be compressed slightly or modeled and animated precisely enough to avoid the annoyed perception.

In the clip with Czech finger alphabet only one observer moved his attention to hands as seen in Fig. 7. The rest focused on face all the time, even when the face is at the moment when finger alphabet is used practically motionless. Finger alphabet is articulated relatively high, approximately in the area around the neck, which makes the perception with peripheral vision easier.

V. Conclusion

Common TV CSL content includes except CSL also Czech finger alphabet for words that do not have an exact equivalent as a sign. In both cases are obserever's eyes focused on the nose area. Comparing the attributes as the size or position of the signer, we can assume that their importance is less than necessary for visible differences in the results. In the compression case we would now like to compare how different quality may be in the hands and face areas still for the pleasant experience, because it is obvious that the hands are seen peripherally with a lower

resolution, less sharp. In the case of the synthesis of sign language in addition to the quality of the model's face enough to ensure his most authentic and natural animation, eg. using motion capture data.

References

[1] W. Stokoe, D. Casterline, and C. Croneberg. "A Dictionary of American Sign Language on Linguistic Principles". Gallaudet College Press, Washington D.C., USA, 1965.

[2] D. Agrafiotis, et al. "Perceptually optimised sign language video coding based on eye tracking analysis". Electronics Letters. 2003, vol. 39, no. 24, p. 1703–1705.

[3] D. Agrafiotis, N. Canagarajah, D. R. Bull. "Perceptually optimised sign language video coding". In Electronics, Circuits and Systems, 2003. ICECS 2003. Proceedings of the 2003 10th IEEE International Conference on. 2003, vol. 2, p. 623–626.

[4] K. Emmorey, R. Thompson, R. Colvin. "Eye gaze during comprehension of American Sign Language by native and beginning signers". Journal of Deaf Studies and Deaf Education, 2009, 14(2), p. 237-243.

[5] D. Agrafiotis, et al. "A perceptually optimised video coding system for sign language communication at low bit rates". Signal Processing: Image Communication, Volume 21, Issue 7, August 2006, Pages 531-549.

[6] A. T. Duchowski. Eye Tracking Methodology: Theory and Practice. 2nd ed. London: Springer, 2007.

[7] O. V. Komogortsev. "Standardization of automated analyses of oculomotor fixation and saccadic behaviors". IEEE Transactions on Biomedical Engineering, 2010, vol. 57, p. 2635–3645.

[8] D. D. Salvucci, J. H. Goldberg. "Identifying fixations and saccades in eye-tracking protocols". In Proceedings of the Eye Tracking Research and Applications Symposium. 2000, p. 71–78.

Acknowledgments

Research described in the paper was supported by the Grants Agency of the Czech Technical University in Prague, grant No. SGS12/078/OHK3/1T/13.