Steering of audio input in hearing aids by eye gaze through electrooculography

ANTOINE FAVRE-FÉLIX[1,2,*], RENSKJE K. HIETKAMP[1], CARINA GRAVERSEN[1], TORSTEN DAU[2], AND THOMAS LUNNER[1,2,3]

[1] *Eriksholm Research Centre, Snekkersten, Denmark*

[2] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[3] *Swedish Institute for Disability Research, Linnaeus Centre, HEAD, Linköping University, Linköping, Sweden*

The behavior of a person during a conversation typically involves both auditory and visual attention. Visual attention implies that the person directs his/her eye gaze towards the sound target of interest, and hence the detection of the gaze may provide a steering signal for future hearing aids. Identification of the sound target of interest could be used to steer a beamformer or select a specific audio stream from a set of remote microphones. We have previously shown that in-ear electrodes can be used to identify eye gaze through electrooculography (EOG) in offline recordings. However, additional studies are needed to explore the precision and real-time feasibility of the methodology. To evaluate the methodology we performed a test with hearing-impaired subjects seated with their head fixed in front of three targets positioned at −30°, 0°, and +30° azimuth. Each target presented speech from the Danish DAT material, which was available for direct input to the hearing aid using head related transfer functions. Speech intelligibility was measured in three conditions: a reference condition without any steering, an ideal condition with steering based on an eye-tracking camera, and a condition where eye gaze was estimated from EarEOG measures to select the desired audio stream. The capabilities and limitations of the methods are discussed.

INTRODUCTION

Due to more advanced signal processing algorithms developed in recent years (Puder, 2009), the performance of hearing aids (HA) has greatly increased. Nevertheless, many HA users still report difficulties to communicate in acoustically challenging conditions with various sound sources and in the presence of reverberation. One of the reasons for the difficulties may be that the HAs are not sensitive to the user's attention and therefore cannot "react" accordingly while normal-hearing listeners typically are able to selectively attend to the target of interest. The European project COCOHA (COgnitive COntrol of a Hearing Aid) attempts to track the attention of the

*Corresponding author: afav@eriksholm.com

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
 ISBN: 978-87-990013-6-1.

HA user via electroencephalographic (EEG) brain activity as well as via electrooculography (EOG). EOG is strongly correlated with eye movements, such that the user's attention could be estimated by eye gaze, assuming that the user is looking at the target of interest. Manabe *et al*. (2013) and Favre-Félix *et al*. (2017) showed that it is possible to reliably measure EOG using in-ear electrodes. Favre-Félix *et al.*, (2017) even used a solution with electrodes integrated in the hearing aid mold designed by the Eriksholm research centre (Fiedler *et al.*, 2016, Pedersen *et al.*, 2014). Thus, advanced hearing devices may in the future be able to measure the eye gaze and steer the amplification of attended versus unattended audio input accordingly. However, even though EOG is closely correlated to eye movements, real-time steering from an EOG signal may be difficult to implement. Here, we designed an experiment to evaluate the potential of steering audio signals through EOG. Hearing-impaired participants were presented one target talker and two competing talkers. Different steering conditions were considered: a control condition where the eyes did not provide any steering; a condition where the eye gaze was estimated using the EOG signal and the audio was steered accordingly, and an ideal condition where the eye gaze was accurately detected through an eye tracker and the audio was steered accordingly.

METHODS

Participants

Eleven hearing-impaired participants took part in the study. Their average age was 75 years, with a standard deviation of 8.9 years. Their audiograms showed moderate to moderately-severe sensorineural, symmetrical hearing loss; the maximum difference between the left and right ear (averaged between 125 and 8000 Hz) was 10 dB and the frequency pure-tone average of thresholds at 500, 1000, 2000, and 4000 Hz ranged from 45 to 69 dB HL (average 55 dB HL). The participants were wearing state-of-the-art behind-the-ear devices fitted with the NAL-NL2 rationale with directionality and noise reduction features turned off.

Stimuli and experimental setup

The participants were presented speech from the Danish DAT material (Nielsen and Dau, 2013), an open-set speech corpus with target words embedded in a carrier sentence. The material consists of sentences in the form of "Dagmar/Asta/Tine tænkte på en *skjorte* og en *mus* i går" ("Dagmar/Asta/Tine thought of a *shirt* and a *mouse* yesterday"). "Skjorte" and "mus" are two target words that change between each sentence and between each talker. All sounds were presented diotically via direct audio input. The participants were asked to direct their gaze at the talker indicated by a light-emitting diode (LED) and to repeat the two target words after the sentence was presented.

The experimental setup is shown in Fig. 1. The participants' head was fixed by a chin-rest. In front of the participant, at a distance of 72 cm, the voices of three talkers (one target talker, two interferers) were presented from the locations −30°, 0° and +30°

azimuth relative to the chin-rest. The sounds were generated via generic head-related transfer functions (HRTFs) corresponding to the three directions. The level of the target talker was initially 6 dB higher than the level of each of the interfering maskers. This was done since hearing-impaired listeners typically have a speech reception threshold (corresponding to 50% correct speech intelligibility) at a target-to-masker ratio (TMR) of +6 dB (Nielsen and Dau, 2013).

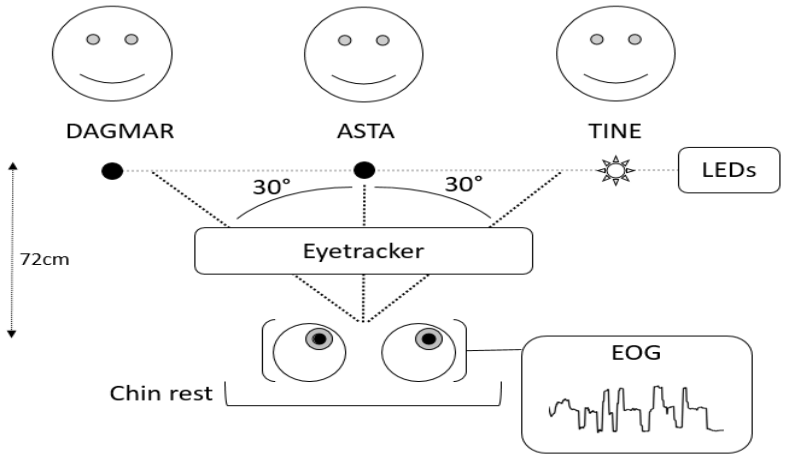

Fig. 1: Representation of the experimental setup. There were three talkers (one target talker (Tine in this example), indicated by an active LED, and two interfering talkers) in front of the participant. The head was fixed with a chin-rest and the eye gaze was measured with an eye tracker and estimated via EOG.

In the control condition (without steering), the behavior of the participant had no impact on the presentation of the audio signal. In the "EOG steering" condition, the EOG signal was used to estimate the eye gaze and to amplify the audio coming from the estimated target talker. In the "eye-tracking" steering condition, the eye gaze of the participant was detected through an eye tracker. In the "EOG steering" and the "eye-tracking" conditions, the audio signal coming from the visually attended talker at was amplified by additional 12 dB to ensure that the participant could clearly identify the target source while still perceiving the interferers (McShefferty *et al.*, 2016).One training list of 20 sentences and three test lists of 20 sentences were used for each condition. The target switched randomly between each sentence such that each talker was presented at least six times per list and each possible transition occurred at least twice.

In the eye-tracking condition, the gaze was estimated at a rate of 30 Hz using an Eyetribe eye tracker (The Eye Tribe ApS, Copenhagen, Denmark). For practical reasons the calibration of the eye tracker was set once and was not adjusted to each individual participant. For the EOG signal, the bioelectric potentials were measured

with a g.tec biosignal amplifier (medical engineering GmbH, Schiedlberg, Austria) sampling at 256 Hz, using three in-ear electrodes in each ear, an electrode on each temple and a reference and a ground electrode on the arm. The EarEOG signal studied was from the cleanest electrode in the right ear re-referenced to the cleanest electrode in the left ear, the EOG signal studied was from the electrode on the right temple re-referenced to the electrode on the left temple. Originally, the goal was to use in-ear electrodes (EarEOG) to steer the amplification instead of surface EOG on the temples. However, during testing the EarEOG signal was considered to be too noisy to be reliably used for steering at this stage. Therefore, the EOG signal, which is less affected by noise interference, was considered instead.

EOG steering algorithm

The main challenge of using EOG in real-time is a direct current (DC) drift that is created by the interface between the skin and the electrodes (Huigen *et al*., 2002; Favre-Félix *et al.*, 2017). Therefore, it is not straightforward to accurately determine the eye gaze relative to the nose from these measurements, whereas eye movements indicative of attention switch can easily be detected. In order to extract meaningful information, a bandpass filter with cut-off frequencies of 0.1 and 4 Hz was applied to the EOG signal. This filtering is effective when the eyes move rapidly (i.e. when the eyes stay less than two seconds on a target), but not when the eyes are fixated on a target. When the eyes are fixated, low-frequency components appear in the EOG, which are then filtered out such that the signal approaches zero. The algorithm used in this study was designed to detect the changes in eye gaze, to estimate when the eyes switch from one target to another and to anticipate this modification of the EOG signal caused by the filtering. According to the positioning of the electrodes that were used to measure the EOG, the filtered EOG signal was positive when the eyes moved to the right and the filtered EOG signal was negative when the eyes moved to the left Since there are three possible targets, five patterns of potential movements can occur: no movement, switching to a target on the right, switching to a target on the left, switching to two targets on the right and switching to two targets on the left.

For this continuous classification, two thresholds were set. The first threshold differentiated between a movement and no movement. The second threshold, higher than the first one, differentiated between switching by one or two targets as illustrated in Fig. 2. The sign of the EOG signal indicated whether the eyes were moving to the left or to the right. A target change was detected when the signal remained above the threshold for 500 ms. This allowed the system to be robust against brief noises, such as eye blinks and jaw movements. Once a target change was detected, the EOG signal was reset to zero to anticipate the modification caused by the filtering. Using this classification system, a mistake could potentially propagate over several sentences Therefore, the algorithm was reset to the middle target in the beginning of each list of 20 sentences. When the participant repeated the words they heard, the algorithm was locked because jaw movements are known to have a strong influence on the in-ear electrode signal.

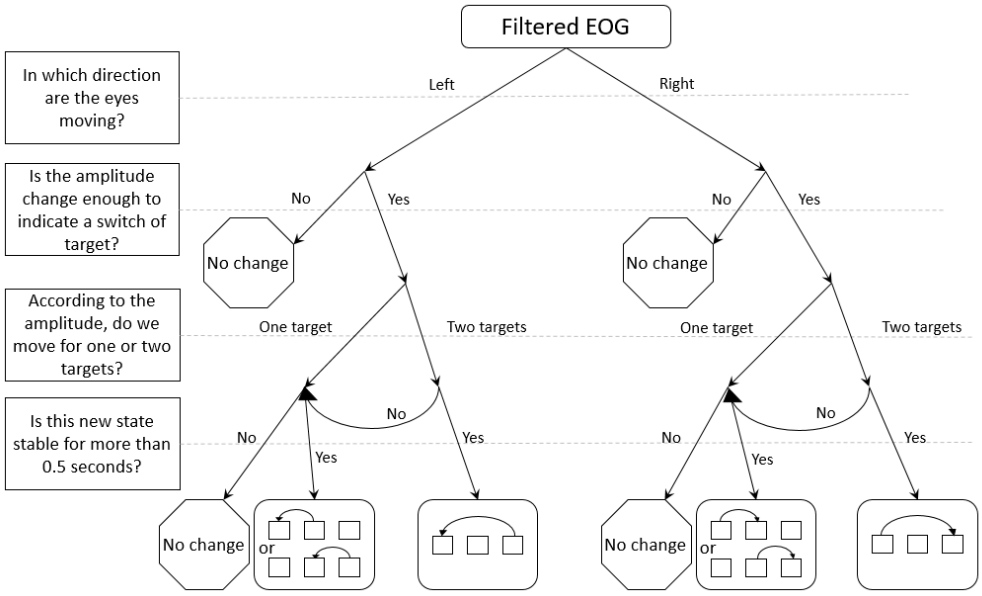

Fig. 2: Decision tree representing the decisions taken by the algorithm to estimate attention shift for the EOG steering system. First, the algorithm evaluates the sign of the filtered EOG to determine the direction the eyes are moving. Then the signal is compared to thresholds values to decide if the estimated eye movement is large enough to change the target source and, if so, to decide to switch to a target. Finally, the algorithm controls that the signal change is not caused by a transient noise.

Analysis

The analysis of the data focused on the results obtained from the seven participants for whom EOG data were measured. For the other four participants, for whom only EarEOG signals were recorded, the data were too noisy for the algorithm to detect the attended target reliably. The scoring of the correctly repeated words per sentence from the DAT material was measured. Two aspects of that score were considered: the score of individual words that were correctly repeated, and the score of full sentences that were correctly repeated. A t-test analysis was applied to compare these scores between conditions, averaged across participants.

The accuracy of the eye-gaze detection algorithm was estimated throughout the whole duration of the experiment, including during the “no steering” and the “eye-tracker steering” conditions. For the duration of each sentence, the estimated target was compared to the target the participant was supposed to attend. Analysis showed two types of errors: when the algorithm changed the target while the sentence was playing, representing a “switch error”, and when the algorithm was fixed on the wrong target, in which case it was possible to estimate how much the algorithm deviated from the attended target.

RESULTS

In terms of word scoring, in the "no steering" condition the participants repeated the word correctly 58.1% of the time, on average, with a standard deviation of 19.3%. In the "EOG steering" condition, the percentage of correct responses was 66.2%, with a standard deviation of 21.7%. In the "eye-tracker steering" condition, the percentage correct was 84.9% with a standard deviation of 11.9%. There was a significant difference between the "no steering" and "eye-tracker steering" conditions (p<0.00001) and between the "eye-tracker steering" and "EOG steering" conditions (p<0,001), but no significant difference between the "no steering" and the "EOG steering" conditions as illustrated in the left panel in Fig. 3.

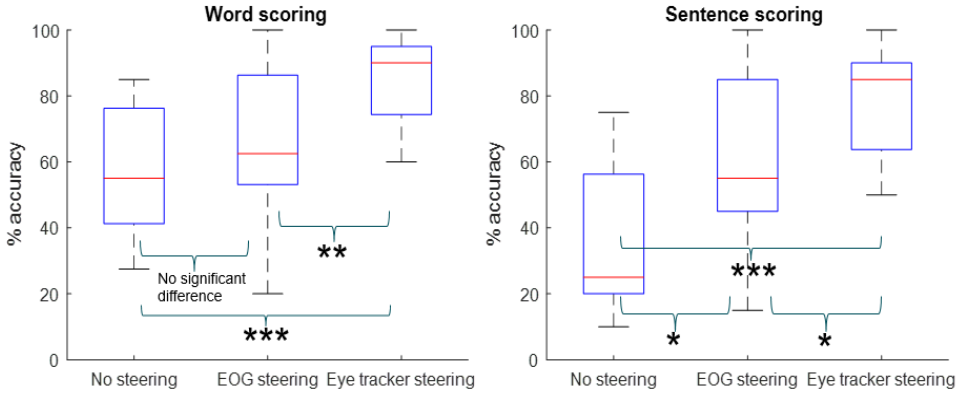

Fig. 3: Average word scoring (left panel) per condition and average sentence scoring (right panel) per condition (* p<0.01; ** p<0.001; *** p<0.00001) in the three conditions with "no steering", "EOG steering" and "Eye-tracker steering".

In terms of sentence scoring, in the "no steering" condition the participants, on average, repeated the whole sentence correctly 38.8% of the time, with a standard deviation of 22%. In the "EOG steering" condition, the corresponding percentage correct was 61.7%, with a standard deviation of 23.4%. In the "eye-tracker steering" condition, the percentage amounted to 78.1% (standard deviation 15.8%). There was a significant difference between the "no steering" and "eye-tracker steering" conditions (p<0.00001), between the "EOG steering" and "eye-tracker steering" conditions (p<0,01) and between the "no steering" and the "EOG steering" conditions (p<0,01) as illustrated in the right panel in Fig. 3.

The algorithm to estimate the attended target through EOG had an accuracy of 65%. The algorithm erroneously detected a change in the middle of a sentence 8.5% of the time and selected the wrong target 26.5% of the time. Specifically, 13.5% of the time the left neighbour was selected, 7% the right neighbour, 3% the left target when it actually was the one to the right, and 3% the right target when it actually was the one to the left. This error distribution is illustrated in Fig. 4. Since there are three targets, a random selection of the target would result in 33% accuracy, or less if the change

during the sentence was taken into account. Therefore, the target estimation algorithm used in this experiment was considered effective.

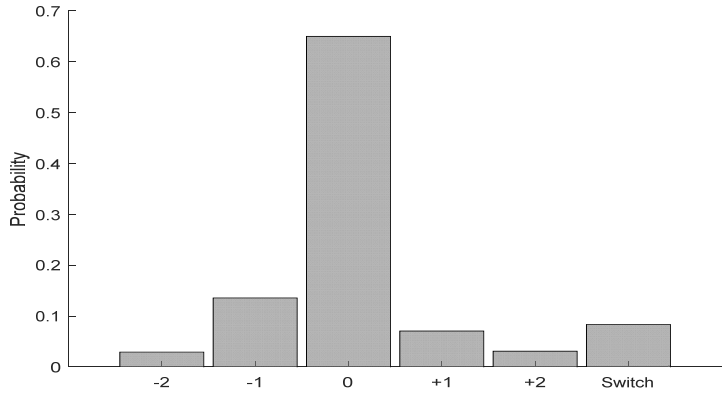

Fig. 4: Histogram representing the accuracy of the algorithm representing the distribution of correct ("0") and incorrect decisions ("+/−1", "+/−2", switch).

DISCUSSION

In this study, we evaluated the potential of steering audio signals through EOG. When estimating EOG from surface electrodes, a significant improvement was seen in sentence performance but not for word scoring compared to the no steering condition. However, the performance of the EOG was in both cases significantly less compared to the ideal condition using an eye tracker to estimate gaze direction.

The EOG estimates were based on surface electrodes on the temples, although the original goal was to use EarEOG. Unfortunately, the EarEOG were too noisy to run the algorithm. This phenomenon was an unexpected challenge, as previous recordings of EarEOG did not display such noise issues (Favre-Félix *et al.*, 2017). Furthermore, the setup in this experiment was tested in a pilot experiment before the actual study presented in this paper was conducted. It is possible that the participants of the present study felt less comfortable with the experimental setup than those involved during pilot testing. Further studies both with normal-hearing and hearing-impaired listeners will clarify the origin of the noise.

The results obtained with both word and sentence scoring using the eye-tracker steering demonstrated the potential of a device that is steered via eye gaze. There were still some errors in this condition, mostly resulting from the calibration of the eye tracker that was not adjusted to the individual participant. Based on the results obtained in this study, a future technology to separate voices in a "cocktail-party" like situation may be based on gaze steering. This could for example be utilized by using a remote microphone for each talker. These results demonstrate that, in such a scenario, a steering device controlled by the eyes may greatly benefit the user.

When the EOG steering algorithm selects the correct target, the whole sentence is amplified by 12 dB. Therefore, if one of the target words is repeated correctly it is likely that the other target word will also be correct. This is not the case in the "no steering" condition. This may explain the significant difference between the two conditions in the case of sentence scoring. The EOG steering algorithm is helpful and increases performance. However, the error rate needs to be minimized before the algorithm can be considered in a realistic implementation.

Moreover, in the current system, classification errors may propagate over several sentences. Since only EOG was used for that steering condition, no additional information was provided for a better error estimation, e.g. via Kalman filtering. Information provided by e.g., head movements and an eye-gaze behavior model that takes head movements into account may allow a better estimation of the visual attention and, thus, may reduce the number of errors to a satisfying degree.

Indeed, in the tests of the present study, the participants had their head fixed, which is unnatural during a conversation. Further studies will take head movement into account. Head movement can be estimated using an accelerometer, a gyroscope and a magnetometer, which can easily be implemented inside a hearing aid.

In conclusion, this study has demonstrated the advantage of applying a steering interface to hearing impaired persons to increase speech intelligibility. However, the study also pointed out challenges of noise reduction from in-ear sensors and the need for additional studies allowing free movements of the head.

REFERENCES

Favre-Felix, A., Graversen, C., Dau, T., and Lunner, T. (**2017**). "Real-time estimation of eye gaze by in-ear electrodes," IEEE EMBC.

Fiedler, L., Obleser, J., Lunner, T., and Graversen, C. (**2016**). "Ear-EEG allows extraction of neural responses in challenging listening scenarios – a future technology for hearing aids?," IEEE EMBC.

Huigen, E., Peper, A., and Grimbergen, C.A. (**2002**). "Investigation into the origin of the noise of surface electrodes," Med. Biol. Eng. Comput., **40**, 332-338. doi: 10.1007/BF02344216

McShefferty, D., Whitmer, W.M., and Akeroyd, M.A. (**2016**). "The just-meaningful difference in speech-to-noise ratio", Trends Hear. doi: 10.1177/2331216515626570

Manabe, H., Fukumoto, M., and Yagi, T. (**2013**). "Conductive rubber electrodes for earphone-based eye gesture input interface," ISWC. doi: 10.1145/2493988.2494329

Nielsen, J.B., and Dau T. (**2013**). "A Danish open-set speech corpus for competing-speech studies," J. Acoust. Soc. Am., **135**, 407-420. doi: 10.1121/1.4835935

Pedersen, E.B., and Lunner, T. (**2014**) "Cognitive hearing aids? Insights and Possibilities", Mechanics of Hearing. doi: 10.1063/1.4939399

Puder, H. (**2009**), "Hearing aids: An overview of the state-of-the-art, challenges, and future trends of an interesting audio signal processing applications", IEEE ISPA. doi: 10.1109/ISPA.2009.5297793