Improving Speech Intelligibility by Hearing Aid Eye-Gaze Steering: Conditions With Head Fixated in a Multitalker Environment

Antoine Favre-Félix[1,2] ⓘ, Carina Graversen[1],
Renskje K. Hietkamp[1], Torsten Dau[2], and Thomas Lunner[1,2,3]

Abstract

The behavior of a person during a conversation typically involves both auditory and visual attention. Visual attention implies that the person directs his or her eye gaze toward the sound target of interest, and hence, detection of the gaze may provide a steering signal for future hearing aids. The steering could utilize a beamformer or the selection of a specific audio stream from a set of remote microphones. Previous studies have shown that eye gaze can be measured through electrooculography (EOG). To explore the precision and real-time feasibility of the methodology, seven hearing-impaired persons were tested, 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, a condition where eye gaze was estimated from EOG measures to select the desired audio stream, and an ideal condition with steering based on an eye-tracking camera. The "EOG-steering" improved the sentence correct score compared with the "no-steering" condition, although the performance was still significantly lower than the ideal condition with the eye-tracking camera. In conclusion, eye-gaze steering increases speech intelligibility, although real-time EOG-steering still requires improvements of the signal processing before it is feasible for implementation in a hearing aid.

Introduction

Hearing-impaired people are in general heavily challenged in listening scenarios that involve multiple speakers, often termed the *cocktail-party* problem (Arons, 2000). The most common hearing loss is caused by damage to the cochlea, leading to reduced neural input to the brain. However, the brain is also influenced by plasticity after long-term hearing loss, which results in an altered ability to discriminate multiple sound sources, as well as a reorganization of the neural networks (Cardin, 2016; Peelle, Troiani, Grossman, & Wingfield, 2011). The latter phenomenon causes the brain to process visual perception in the auditory cortex as early as 3 months after the onset of a profound hearing loss (Glick & Sharma, 2017). This is partly because the hearing-impaired person may search for nonauditory cues in the environment, where visual cues are often useful.

For hearing-aid applications, strategies where amplification of the sound source the listeners gaze at may be advantageous.

The idea of using eye gaze to steer a hearing aid has already been explored in several previous studies (Hart, Onceanu, Sohn, Wightman, & Vertegaal, 2009; Kidd, 2017; Kidd, Favrot, Desloge, Streeter, & Mason, 2013). Hart et al. compared eye-gaze selection of a target to

[1]Eriksholm Research Centre, Snekkersten, Denmark
[2]Hearing Systems Group, Department of Electrical Engineering, Danish Technical University, Lyngby, Denmark
[3]Linnaeus Centre HEAD, Swedish Institute for Disability Research, Linköping University, Sweden

Corresponding Author:
Antoine Favre-Félix, Eriksholm Research Centre, Rørtangvej 20, 3070 Snekkersten, Denmark.
Email: afav@eriksholm.com

manual selection (by pointing at a target or pressing a button) for the steering of audio. They found that the eye control was faster and allowed for a better recall. The participants also rated the eye control functionality as the *easiest, most natural*, and *best overall* compared with the manual selection functionality. Kidd et al. designed what they called a visually guided hearing aid that used eye-tracking glasses to steer audio coming from an acoustic beamforming microphone array. This device allowed participants to obtain near to or better than normal spatial release from masking. Expanding on these results, they found that both normal-hearing and hearing-impaired listeners could benefit from the steering offered by the visually guided hearing aid in speech-on-speech masking conditions. Nevertheless, these studies utilized eye-tracking systems, which are typically intrusive in the field of vision. To estimate the eye gaze, electrooculography (EOG) recorded from the temples is a well-established method, which is based on the positive potential at the cornea and the negative potential at the retina in the eye (Brown et al., 2006; Marmor & Zrenner, 1994).

While EOG is closely correlated to eye movements, the feasibility of real-time steering has not yet been explored. To evaluate the usefulness of the EOG signal alone, it is beneficial to keep the head fixated since eye movements are tightly coupled to head movements via the oculomotor system (Ackerley & Barnes, 2011). An ideal steering signal in such a head-fixated scenario would be a stable eye-gaze signal relative to the head (e.g., the relative angle vs. the frontal direction). Most eye-trackers give such stable relative eye-gaze signal. However, the case of EOG-steering is complicated by the fact that the skin-electrode junction creates a time-varying offset of the relative eye-gaze signal (Favre-Felix et al., 2017; Huigen, Peper, & Grimbergen, 2002) and thus the EOG signal cannot be used directly as a steering signal. Therefore, the gaze direction must be estimated by an algorithm. One example of such EOG gaze-estimation algorithm was recently suggested by (Hládek, Porr, & Brimijoin, 2018). In this study, an alternative algorithm is presented.

The idea with the current experiment was to investigate whether eye-gaze steering via an EOG gaze-estimation algorithm would work under relatively easy dynamic conditions, given that the EOG gaze-estimation algorithm might produce errors and thus classify wrongly eye-gaze directions. To achieve relatively easy dynamic conditions, the setup was using a target switching time of several seconds to assure that the test person had a stable gaze at the target, in a multitarget environment with several competing talkers.

The eye-gaze signal from an eye-tracker or the current EOG estimation algorithm was used to increase the amplification for the target talker, and a no-steering condition was used for reference.

It was hypothesized that when compared with a no-steering condition, the EOG-steering algorithm would improve speech-in-speech performance despite the errors that it may produce. Furthermore, it was hypothesized that the speech-in-speech performance with the EOG-steering would be inferior to a gaze-steering signal from an eye-tracker, since the eye-tracking errors plausibly would be lower.

Methods

Participants

Seven hearing-impaired participants (two males) were enrolled in the study. The average age was 77 years, with a standard deviation (*SD*) of 4.7 years. Their audiograms showed moderate to moderately severe sensorineural, symmetrical hearing losses. The maximum difference between the left and right ear's audiometric thresholds (averaged between 125 and 8000 Hz) was 10 dB and the thresholds at 500, 1000, 2000, and 4000 Hz ranged from 45 to 59 dB HL (average 54 dB HL). The average audiogram is shown in Figure 1. The participants wore state-of-the-art behind-the-ear devices fitted with the NAL-NL2 rationale with directionality and noise reduction features turned off.

The study was approved by the ethics committee for the capital region of Denmark (Journal number H-1-2011-033). The study was conducted according to the Declaration of Helsinki, and all participants signed a written consent prior to the experiment.
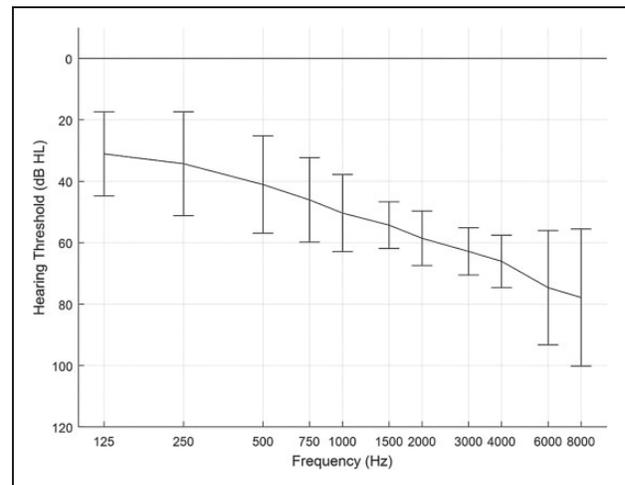
Figure 1. Average audiogram for both ears for the seven participants, including error bars (±*SD*).

Stimuli and Experimental Setup

The paradigm consisted of four steps. First, the temple EOG electrodes were mounted, and the eye-tracking camera was adjusted to be able to capture the gaze from the test subject. Next, a calibration session was conducted, where the test subject was instructed to follow a light-emitting diode (LED) without any auditory input to estimate the EOG thresholds reflecting changes in the attended source. Next, a training session was conducted to acquaint the participants to the speech-in-speech test. The training session consisted of one list of 20 sentences for each of the three test conditions. Finally, the actual experiment was conducted.

The experiment consisted of three conditions: (a) *no-steering*, (b) eye-gaze steering obtained from EOG, and (c) eye-gaze steering obtained from an eye-tracking camera. The conditions were presented in a double-blinded randomized block design, with each block consisting of 20 stimuli. A total of 180 stimuli were presented to each test subject. The three conditions were chosen to be able to compare the proposed solution with *EOG-*

steering to the worst scenario of *no-steering* and to the optimal scenario with a highly robust eye-tracking camera.

During the recordings, the participants' head was fixed using a chinrest, as illustrated in Figure 2. 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 chinrest. The audio streams were generated via generic head-related transfer functions (HRTFs) corresponding to the three directions. The generic HRTFs were obtained from the CIPIC HRTF database (Algazi, Duda, Thompson, & Avendano, 2001). The level of the target talker was initially 6 dB higher than the level of each of the interfering maskers, that is, a target to masker ratio (TMR) of $+6\,dB$. This was done since hearing-impaired listeners typically have a speech reception threshold (corresponding to 50% correct speech intelligibility) at a TMR of $+6\,dB$ (Bo Nielsen, Dau, & Neher, 2014).

The participants were presented speech from the Danish DAT material (Bo Nielsen et al., 2014), an
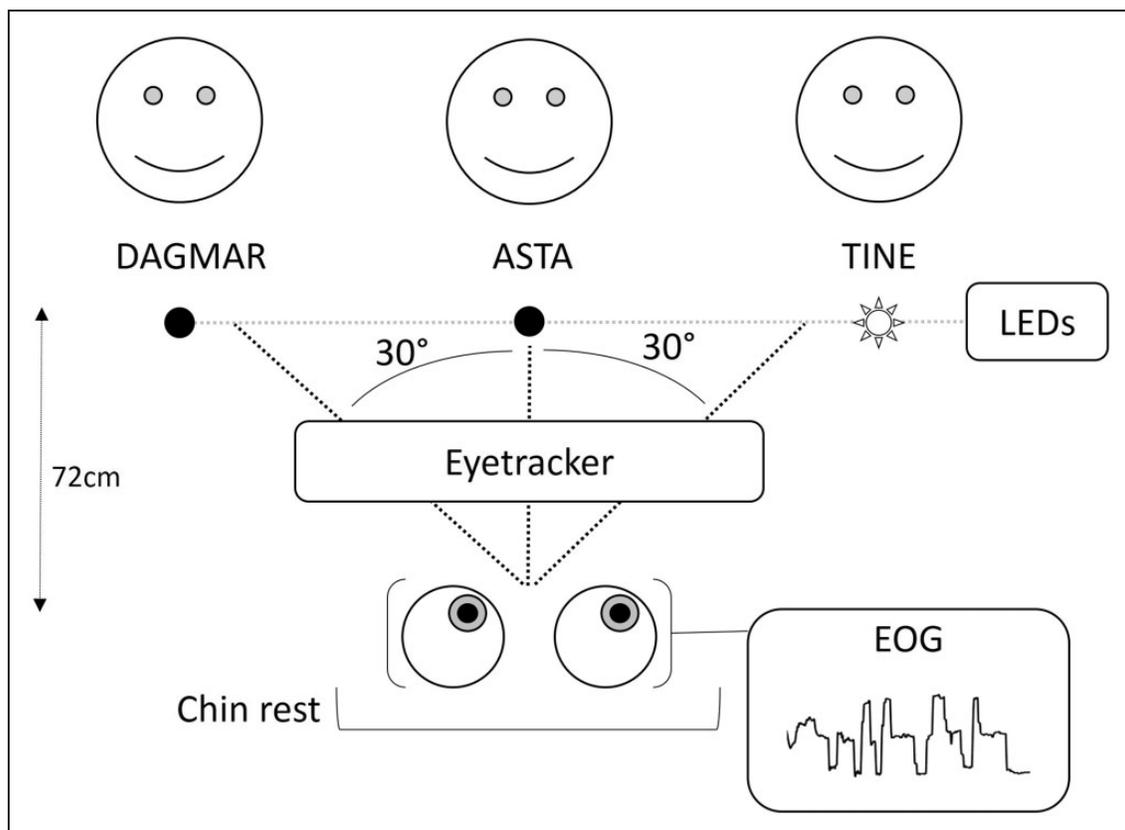

Figure 2. Representation of the experimental setup. There were three talkers, one target talker indicated by an active LED (Tine in this example), and two interfering talkers in front of the participant. The head was fixed with a chinrest, and the eye gaze was measured with an eye-tracker and estimated via EOG.
EOG = electrooculography; LED = light-emitting diode.

open-set speech corpus with two target words embedded in a carrier sentence, similar to the English TVM corpus (Helfer & Freyman, 2009). 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. By measuring the time when the first word is spoken in 20 sentences for each talker, it was estimated that the first target word is presented roughly 750 ms after the start of the sentence. For a given participant, "Dagmar sentences" came all from the same direction and was marked in the scene with *Dagmar* (see Figure 2). And the same for *Asta* and *Tine* having their respective directions for the given participant. However, the positions of *Dagmar, Asta*, and *Tine* were randomized between participants. To give a natural spatial impression, left ear HRTFs were applied on the sound files for each of the three talker directions and added together to a single output signal presented to the left ear and transmitted to a left behind-the-ear hearing device by direct audio input. Similarly, a right ear signal was created via HRTFs. Thus, the participants received a dichotic signal with a spatial impression. The participants were asked to direct their gaze at the talker indicated by an LED and to repeat the two target words after the sentence was presented. The LED was activated 2 s before the start of the sentence to give the participant and the steering algorithm enough time to get ready for the new sentence, estimating less than 500 ms for reaction time (Gezeck, Fisher, & Timmer, 1997) and 500 ms for the algorithm to make a decision. It remained activated until it changed to another target.

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-tracker-steering" condition, the eye gaze of the participant was detected through an eye-tracking camera. In the EOG-steering and the eye-tracker-steering conditions, the audio signal coming from the visually estimated attended talker was amplified by an additional 12 dB to ensure that the participant could clearly identify the target source while still perceiving the interferers (McShefferty, Whitmer, & Akeroyd, 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; each talker was presented at least six times per list and each possible transition (no change, one position to the right, one position to the left, two positions to the right, and two positions to the left) 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 USBamp biosignal amplifier (Medical Engineering GmbH, Schiedlberg, Austria) sampling at 256 Hz, using an electrode on each temple and a reference and ground electrode on the arm. The EOG signal studied was from the electrode on the right temple rereferenced to the electrode on the left temple.

EOG-Steering Algorithm

The main challenge of using EOG-steering in real time is a direct current drift that is created by the interface between the skin and the electrodes (Favre-Felix et al., 2017; Huigen et al., 2002). Figure 3 illustrates the difference in stability between the signal from the eye-tracker and the EOG measured. Therefore, it is not straightforward to accurately determine the eye-gaze position relative to the nose from these measurements, whereas eye movements indicative of an attentional switch can be detected. To extract meaningful information, a bandpass filter with cutoff frequencies of 0.1 and 4 Hz was applied to the EOG signal. This filtering is effective when the eyes move rapidly, that is, when the eyes stay less than 2 s on a target, but not when the eyes are fixated on a target (Favre-Felix et al., 2017). When the eyes are fixated, low-frequency components appear in the EOG signal, 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, that is, to estimate when the eyes switched 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 were three possible targets, five patterns of potential movements could 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, which was higher than the first, differentiated between switching to one or two targets as illustrated in Figure 4. 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, thus allowing 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
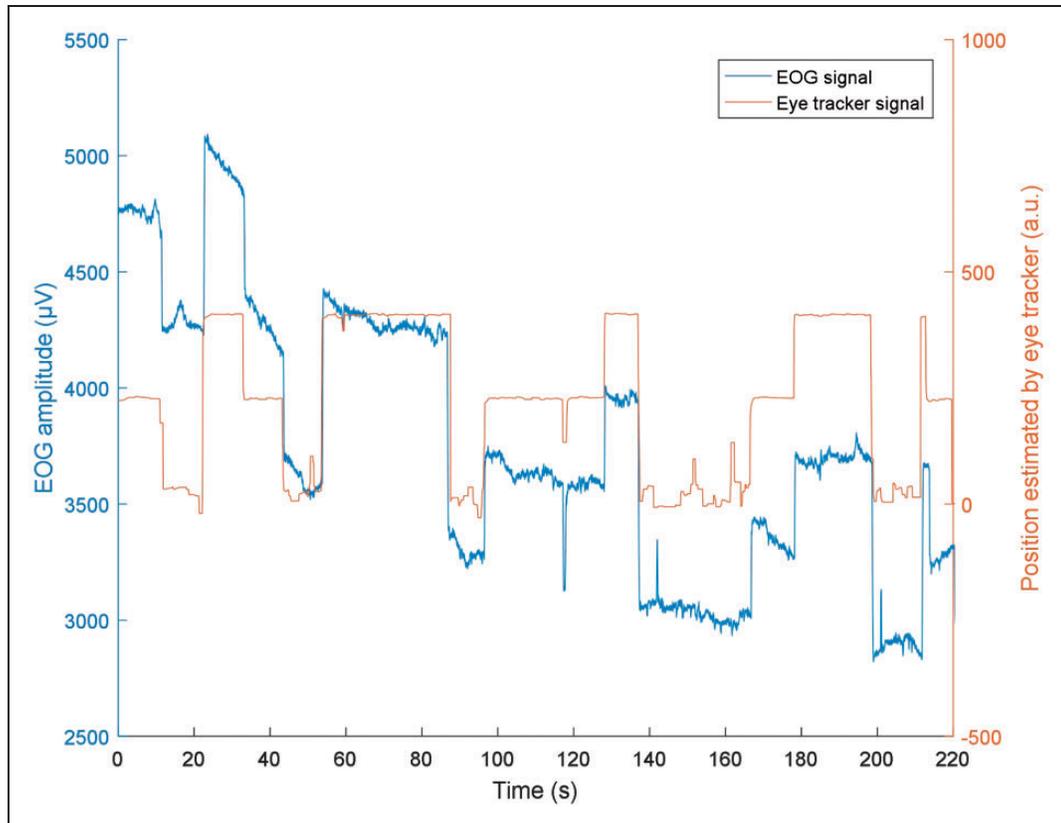
Figure 3. Individual trace of EOG measured (blue) compared with eye-tracking data recorded during the same period (red). EOG = electrooculography.

classification system, a mistake could potentially propagate over several sentences. Therefore, the algorithm was reset to the middle target at the beginning of each list of 20 sentences. Moreover, when the participant repeated the words they heard, the algorithm was locked to avoid interference from jaw movements. For implementation of the EOG-steering algorithm, Simulink, implemented with MATLAB R2014a, was used (Mathworks Inc., Natick, MA, USA).

Analysis of the Behavioral Data

The scoring of the correctly repeated words per sentence from the DAT material was measured. Two aspects of this 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. The scores were obtained by averaging the performance for each list per participant (hence, a total number of 21 measurements per condition were used for analysis). For a clearer representation of the distribution of the scores for all participants, a histogram of the responses depending on the steering condition was generated.

Afterwards, a series of *t* tests was performed, in a fixed condition and between scores, and at a fixed score and between conditions, to highlight the significant parameters involved.

The accuracy of the EOG eye-gaze detection algorithm was estimated throughout the 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 with the target to which the participant was supposed to attend. Two types of errors were obtained: 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 to which degree the algorithm deviated from the attended target (one or two targets to the left or right).

Statistical Analysis

Descriptive statistics are reported as the mean ($\pm SD$) unless otherwise indicated. The *p* values presented in this article have been Bonferroni corrected (Cleophas & Zwinderman, 2016), that is, the *p* values for the
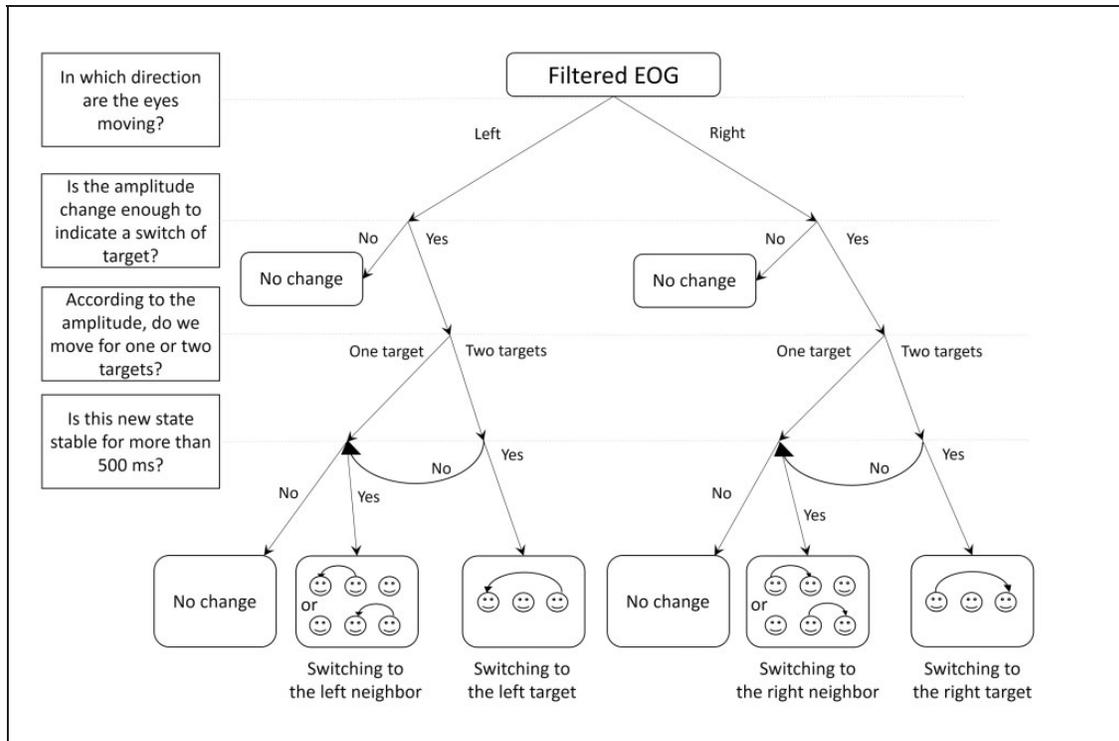
Figure 4. Decision tree representing the decisions taken by the algorithm to estimate the attention shift for the EOG-steering system. First, the algorithm evaluates the sign of the filtered EOG to determine the direction that the eyes are moving. Then, the signal is compared with threshold values to decide if the estimated eye movement is large enough to change the target source and, if so, to decide which target to switch to. Finally, the algorithm includes a control that assures that the signal change is not caused by a transient noise. EOG = electrooculography.

t tests comparing the overall word scoring and sentence scoring between conditions have been multiplied by 3; and the *p* values for the series of *t* tests comparing the score distribution for the fixed condition and the conditions at the fixed score have been multiplied by 9. After these corrections, a value of $p < .05$ was considered as an indication of statistical significance. All statistical analysis was performed with the MATLAB R2016a software (Mathworks Inc., Natick, MA, USA).

Results

All seven participants completed the study and since no outliers were detected, the statistical analysis includes all participants.

Behavioral Performance Results

In terms of word scoring, in the no-steering condition, the participants repeated each word correctly 58.5% ($\pm$19.6%) of the time, on average. In the EOG-steering condition, the percentage of correct responses was 63.5% ($\pm$24.2%). In the eye-tracker-steering condition, the percentage of correct responses was 85.1% ($\pm$12.2%). There was a significant difference between the no-steering and

eye-tracker-steering conditions ($p < .001$) but there was no significant difference between the EOG-steering and eye-tracker-steering conditions, nor between the EOG-steering and the no-steering conditions, as illustrated in the left panel in Figure 5.

In terms of sentence scoring, in the no-steering condition on average, the participants repeated the whole sentence correctly 39.1% ($\pm$22.2%) of the time. In the EOG-steering condition, the corresponding percentage correct was 59.3% ($\pm$25.9%). In the eye-tracker-steering condition, the percentage correct amounted to 79.1% ($\pm$16.4%). There was a significant difference between the no-steering and eye-tracker-steering conditions ($p < .001$), between the EOG-steering and eye-tracker-steering conditions ($p<.05$), and between the EOG-steering and the no-steering conditions ($p < .01$), as illustrated in the right panel in Figure 5.

Distribution of the Scoring for the Different Conditions

In the no-steering condition, for 22.1% ($\pm$17.9%) of the sentences, none of the words were repeated correctly; in 38.8% ($\pm$9.6%) of the sentences, only one word was repeated correctly; and in 39.1% ($\pm$22.3%) of the sentences, both words were repeated correctly. In the EOG-
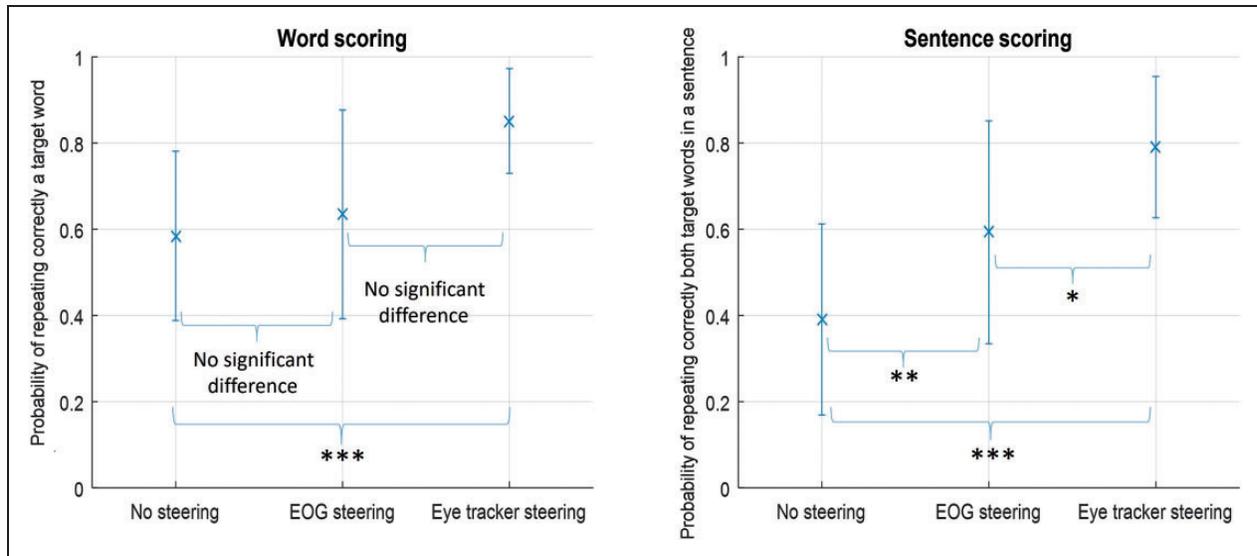
Figure 5. Average probability of repeating correctly a target word (left panel) per condition and average probability of repeating correctly both target words in a sentence (right panel) per condition (*$p<.05$; **$p<.01$; ***$p<.001$) in the three conditions of no-steering, EOG-steering, and Eye-tracker-steering, including error bars ($\pm SD$).
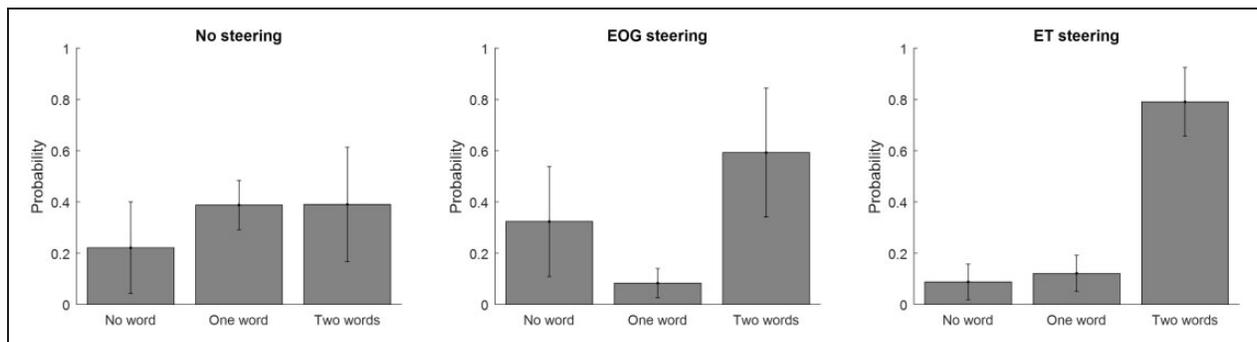EOG = electrooculography.

Figure 6. Histogram representing the distribution of scoring depending on the condition presented, including error bars ($\pm SD$).
EOG = electrooculography; ET = eye-tracker.

steering condition, the participants were unable to repeat either word correctly in 32.4% (± 21.5%) of the sentences, they repeated only one word in 8.3% (± 5.7%) of the sentences, and they repeated both words correctly in 59.3% (± 25.1%) of the sentences. For the eye-tracker-steering condition, in only 8.8% (± 6.9%) of the sentences, none of the words were repeated correctly; in 12.1% (± 7.1%) of the sentences, only one word was correct; and in 79.1% (± 12.9%) of the sentences, both words were repeated correctly by the test subject. The results are shown in Figure 6.

When comparing the word scoring within one condition, t tests determined that there was a significant difference ($p < .05$) in the EOG-steering condition between

one and two words correctly repeated; in the eye-tracker-steering condition, there was a significant difference between the sentences where neither word was correctly repeated and the sentences where both words were correctly repeated ($p < .001$), as well as a significant difference between the sentences with one and two words correctly repeated ($p < .01$). These t test results can be found in Table 1.

For the fixed scoring, t tests determined that when only one word was repeated correctly, there was a significant difference between the no-steering condition and the EOG-steering condition ($p < .001$), as well as between the no-steering and the eye-tracker-steering condition ($p < .05$). These t test results can be found in Table 2.

Table 1. Results of *t* Tests Comparing the Different Scores at Fixed Condition.

	No-steering	EOG-steering	Eye-tracker-steering
0 vs. 1	>0.05	>0.05	>0.05
1 vs. 2	>0.05	*	**
0 vs. 2	>0.05	>0.05	***

Note. The *p* values have been Bonferroni corrected (*p < .05. **p < .01. ***p < .001).
EOG = electrooculography.

Table 2. Results of *t* Tests Comparing the Different Conditions for a Fixed Score.

|	No-steering vs. EOG-steering	EOG-steering vs. eye-tracker-steering	No-steering vs. eye-tracker-steering
0	>0.05	>0.05	>0.05
1	***	>0.05	*
2	>0.05	>0.05	>0.05

Note. The *p* values have been Bonferroni corrected (*p < .05. ***p < .001).
EOG = electrooculography.

Evaluation of the EOG-Steering Algorithm

The algorithm used to estimate the attended target through EOG had an accuracy of 62.5% (±18.4%). Since there were three targets, a random selection of the target would result in an accuracy of 33%, or less if the change during the sentence was taken into account. The algorithm erroneously detected a change in the middle of a sentence 7.8% (±6%) of the time and selected the wrong target 29.7% of the time. Specifically, the left neighbor was selected 12.9% (±6.9%) of the time, the right neighbor was selected 10.1% (±8.4%) of the time, the left target was selected 2.3% (±2.3%) of the time when it was actually the one to the right, and the right target was selected 4.4% (±4.9%) of the time when it was actually the one to the left. This error distribution is illustrated in Figure 7. Taken together, the target estimation algorithm used in this experiment was considered to be effective.

Discussion

This study evaluated the effect of eye-gaze steering of a hearing aid on speech intelligibility in hearing-impaired subjects. The results demonstrated that eye-gaze steering, achieved in real time via EOG measures, improved speech intelligibility compared with a no-steering condition.
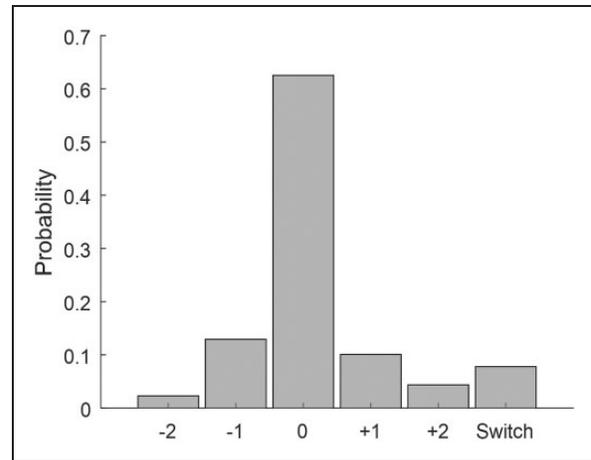
Figure 7. Histogram representing the accuracy of the EOG-steering algorithm detailing the distribution of correct (0) and incorrect decisions (±1, ±2, switch), including error bars (±SD).

Experimental Setup

The experimental setup was designed to show the benefit of amplification of a selected audio stream. Several considerations were taken into account regarding the level of amplification. Previous studies showed that the within-subject variance for scoring is 1.4 dB for the DAT material (Bo Nielsen et al., 2014). For hearing-impaired subjects, a TMR of +6 dB was found to correspond to the speech recognition threshold at 50% (Bo Nielsen et al., 2014). Therefore, when designing the experiment, a TMR of +6 dB was chosen for the control condition. An additional amplification of 12 dB was chosen here to substantially improve speech intelligibility. Even though the setup in the present study was slightly different from the one used in Bo Nielsen et al., the results obtained in the control condition with no-steering were reasonably close (58.5% ± 19.6%) to what was expected. It is important to note that the setup used in this experiment was different than the one used in the testing of the DAT material (Bo Nielsen et al., 2014). First, the target was separated by 50° while here they are separated by 30°. But more importantly, in Bo Nielsen's paper, the task was different in that the participants were not aware of where the target would be positioned and therefore had to make additional cognitive effort to answer correctly. For those reasons, it is understandable that the results obtained in the control condition were not exactly 50% correct responses. Nevertheless, this point of the psychometric function was considered to be "comfortable" for the subjects while still providing a large dynamic range to explore higher performance levels through the steering of the audio was available.

Improvement of Speech Intelligibility by Eye-Gaze Steering of Audio Input

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, which primarily resulted from the calibration of the eye-tracker, as it was not adjusted to the individual test subject. Based on the results shown in Figure 6, the eye-gaze steering led not only to a higher average word score but also to an increased sentence intelligibility. The results obtained in this study confirm earlier findings suggesting that a future technology to separate voices in a "cocktail-party"-like situation may be based on eye-gaze steering (Hart et al., 2009; Kidd, 2017; Kidd et al., 2013). In these previous studies, it was assumed that the different sources could be isolated with ideal beamformers. In a hearing-aid application, a viable separation of the sources could, for example, be achieved by using a remote microphone for each talker.

Potential and Limitations of the EOG Algorithm

The first hypothesis stated that the EOG signals could provide a feasible solution to extract the steering signal. When estimating the eye gaze from surface EOG electrodes, a significant improvement was observed for sentence scoring but not for word scoring, compared with the no-steering condition. However, the performance of the EOG-steering was significantly lower than in the ideal condition using an eye-tracker to estimate gaze direction when comparing the sentence scoring. This lower performance was caused by the limitation of the algorithm used which committed a significant number of errors. This difference between word scoring and sentence scoring is represented by those sentences where only one word was repeated correctly. A statistical analysis revealed that a score of one correctly repeated word was more likely to happen in the no-steering condition than in the EOG-steering or in the eye-tracker-steering conditions due to the steering process. When the system selected an audio stream to be amplified in the two steering conditions, the whole sentence coming from that stream was amplified by 12 dB. Therefore, when one of the target words was repeated correctly, it was more likely that the other target word would also be repeated correctly. This was not the case in the no-steering condition.

In contrast, in the EOG-steering condition, no significant difference was found between the sentences where no word was repeated correctly and the sentences where both words were correctly repeated. This higher number of errors compared with the eye-tracker-steering condition was caused by errors in the EOG-steering algorithm, where the wrong target was detected and amplified. Different types of issues may have caused errors. An unexpected low-frequency noise in the signal may have been detected as a change in gaze. Moreover, although the thresholds for the algorithms were set individually, the calibration procedure to estimate the threshold values was empirical and did not estimate very accurate values. This could result in errors in the selection of the target by the algorithm. Finally, if an error had occurred, the detection of the next target started from an erroneous position, resulting in an error that could propagate over several sentences. Thus, the error rate still needs to be minimized before the algorithm can be considered in applications. The algorithm used in this study was estimated heuristically, based on bandpass filtering followed by threshold detection. Several other studies have used EOG for brain computer interfaces (Behrens, MacKeben, & Schröder-Preikschat, 2010; Hládek et al., 2018), focusing on saccade detection. However, there is a disadvantage in connection to using only saccades, since other natural eye movements can then not be considered.

This study showed that it was possible to improve speech-in-speech performance with EOG-steering for the current setup. However, the setup lacked realism in the sense that target switches are typically faster than the 2-s delay before sentence start simulated in this experiment. Best, et al., 2017 showed that much faster steering would be expected and that the EOG-steering algorithm thus should do precise classifications on a shorter time scale, rather below 500 ms.

Furthermore, the current setup did not allow head movements, which is a limitation. Future EOG gaze-estimation algorithms should also include estimates of head rotations for a natural steering.

There is nothing fundamental preventing a highly improved EOG-steering algorithm.

Future Perspectives for Hearing-Aid Technologies

The objective behind this study was to explore the possibility of a hearing-aid device that could be steered via eye gaze. To apply this system in a hearing aid, EOG measured at the temples, with the head position fixed, is not a feasible solution. However, previous studies have explored the possibility to use electrodes inside the ear canal to record electrophysiological responses (Favre-Felix et al., 2017; Fiedler et al., 2016; Kidmose et al., 2013; Manabe, Fukumoto, & Yagi, 2015; Petersen & Lunner, 2015). These in-ear electrodes could be combined with a hearing aid to estimate eye movements. Head movements could be estimated using an accelerometer, a gyroscope, and a magnetometer, which can also fit into a hearing aid. The combination of those two signals (EarEOG and head tracking), along with more advanced processing (e.g., a better error estimation via Kalman

filtering [Roth & Gustafsson, 2011]), would offer the possibility to steer audio in future hearing devices. Moreover, the combined information provided by eye gaze and head movements may enable a behavioral model that can predict an attended talker in a conversation and, thus, may reduce the number of errors.

Conclusion

In a dynamic competing talker scenario, eye-gaze steering was evaluated using an eye-tracking system and an EOG-based algorithm. The results showed that the EOG-steering improved speech intelligibility compared with the no-steering condition. Although the algorithm used in this study contains inaccuracies and does not take head movement into account, it may be interesting for future hearing-aid applications.

Declaration of Conflicting Interests

Funding

ORCID iD

Antoine Favre-Félix ![ORCID] http://orcid.org/0000-0002-8291-6252

References

Ackerley, R., & Barnes, G. R. (2011). The interaction of visual, vestibular and extra-retinal mechanisms in the control of head and gaze during head-free pursuit. *The Journal of Physiology*, 7, 1627–1642. doi:10.1113/jphysiol.2010.199471

Algazi, V. R., Duda, R. O., Thompson, D. M., & Avendano, C. (2001). The CIPIC HRTF database. *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics* (Cat. No.01TH8575; pp. 99–102). Washington, DC: IEEE.

Arons, B. (2000). A review of the cocktail party effect. The separation of speech channels early work. *Transition*, p. 16.

Behrens, F., MacKeben, M., & Schröder-Preikschat, W. (2010). An improved algorithm for automatic detection of saccades in eye movement data and for calculating saccade parameters. *Behavior Research Methods*, 42(3), 701–708. doi:10.3758/BRM.42.3.701

Best, V., Roverud, E., Streeter, T., Mason, C. R., & Kidd, G. (2017). The benefit of a visually guided beamformer in a dynamic speech task. *Trends in Hearing*, 21, 233121651772230. doi:10.1177/2331216517722304

Bo Nielsen, J., Dau, T., & Neher, T. (2014). A Danish open-set speech corpus for competing-speech studies. *The Journal of the Acoustical Society of America*, 135(1), 407–420. doi:10.1121/1.4835935

Brown, M., Marmor, M., Vaegan, Zrenner, E., Brigell, M., & Bach, M. (2006). ISCEV standard for clinical electrooculography (EOG) 2006. *Documenta Ophthalmologica*, 113(3), 205–212. doi:10.1007/s10633-006-9030-0

Cardin, V. (2016). Effects of aging and adult-onset hearing loss on cortical auditory regions. *Frontiers in Neuroscience*, 10, 1–12. doi:10.3389/fnins.2016.00199

Cleophas, T. J., & Zwinderman, A. H. (2016). *Clinical data analysis on a pocket calculator* (pp. 99–102). Springer International Publishing. doi:10.1007/978-3-319-27104-0

Favre-Felix, A., Graversen, C., Dau, T., & Lunner, T. (2017). Real-time estimation of eye gaze by in-ear electrodes. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4086–4089). Washington, DC: IEEE.

Fiedler, L., Obleser, J., Lunner, T., & Graversen, C. (2016). Ear-EEG allows extraction of neural responses in challenging listening scenarios—A future technology for hearing aids? *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 5697–5700). Washington, DC: IEEE.

Gezeck, S., Fisher, B., & Timmer, J. (1997). Saccadic reaction times: A statistical analysis of multimodal distributions. *Vision Research*, 37(15), 2119–2131. doi:10.1016/S0042-6989(97)00022-9

Glick, H., & Sharma, A. (2017). Cross-modal plasticity in developmental and age-related hearing loss: Clinical implications. *Hearing Research*, 343, 191–201. doi:10.1016/j.heares.2016.08.012

Hart, J., Onceanu, D., Sohn, C., Wightman, D., & Vertegaal, R. (2009). The attentive hearing aid: Eye selection of auditory sources for hearing impaired users. In T. Gross, et al (Eds), *Human-computer interaction—INTERACT 2009. Lecture Notes in Computer Science* (pp. 19–35). Berlin, Germany: Springer-Verlag.

Helfer, K. S., & Freyman, R. L. (2009). Lexical and indexical cues in masking by competing speech. *The Journal of the Acoustical Society of America*, 125(1), 447–456. doi:10.1121/1.3035837

Hládek, Ľ., Porr, B., & Brimijoin, W. O. (2018). Real-time estimation of horizontal gaze angle by saccade integration using in-ear electrooculography. *PLoS One*, 13(1), 1–24. doi:10.1371/journal.pone.0190420

Huigen, E., Peper, A., & Grimbergen, C. A. (2002). Investigation into the origin of the noise of surface electrodes. *Medical & Biological Engineering & Computing*, 40(3), 332–338. doi:10.1007/BF02344216

Kidd, G. (2017). Enhancing auditory selective attention using a visually guided hearing aid. *Journal of Speech Language and Hearing Research*, 60(10), 3027–3038. doi:10.1044/2017_JSLHR-H-17-0071

Kidd, G., Favrot, S., Desloge, J. G., Streeter, T. M., & Mason, C. R. (2013). Design and preliminary testing of a visually

guided hearing aid. *The Journal of the Acoustical Society of America*, *133*(3), EL202–EL207. doi:10.1121/1.4791710

Kidmose, P., Looney, D., Ungstrup, M., Rank, M. L., & Mandic, D. P. (2013). A study of evoked potentials from ear-EEG. *IEEE Transactions on Biomedical Engineering*, *60*(10), 2824–2830. doi:10.1109/TBME.2013.2264956

Manabe, H., Fukumoto, M., & Yagi, T. (2015). Conductive rubber electrodes for earphone-based eye gesture input interface. *Personal and Ubiquitous Computing*, *19*(1), 143–154. doi:10.1007/s00779-014-0818-8

Marmor, M. F., & Zrenner, E. (1994). Standard for clinical electro-oculography. *Documenta Ophthalmologica*, *85*(2), 115–124. doi:10.1007/BF01371127

McShefferty, D., Whitmer, W. M., & Akeroyd, M. A. (2016). The just-meaningful difference in speech-to-noise ratio.

Trends in Hearing, *20*, 1–11. doi:10.1177/2331216515626570

Peelle, J. E., Troiani, V., Grossman, M., & Wingfield, A. (2011). Hearing loss in older adults affects neural systems supporting speech comprehension. *Journal of Neuroscience*, *31*(35), 12638–12643. doi:10.1523/JNEUROSCI.2559-11.2011

Petersen, E. B., & Lunner, T. (2015). Cognitive hearing AIDS? Insights and possibilities. *AIP Conference Proceedings*, *1703*, 090001. doi:10.1063/1.4939399

Roth, M., & Gustafsson, F. (2011). An efficient implementation of the second order extended Kalman filter. In *Proceedings of the 14th International Conference on Information Fusion* (pp. 1–6). IEEE.