Student Conference Proceedings
Vol 1, No 1, 2025, Article ID 1935, 4 Pages

Student Conference Lübeck

Proceedings Article

Auditory Attention Decoding – Constructing a Study –

Karina Kerstin Rohm[a,*] · René Pallenberg[b,*] · Alfred Mertins[b]

[a]Student of Medical Engineering Science, Universität zu Lübeck, Luebeck, Germany
[b]Institute for Signal Processing, Universität zu Lübeck, Luebeck, Germany
*Corresponding author, email: karina.rohm@student.uni-luebeck.de; r.pallenberg@uni-luebeck.de

Received 04 February 2025; Accepted 14 July 2025; Published online 18 July 2025

Abstract

Globally rising life expectations lead to increasing age-related conditions which need to be addressed by modern medicine. One of these is age-related hearing loss and the subsequent use of hearing aids (HA). A problem commonly associated with HA is the "cocktail party problem" occurring during social gatherings and potentially making the person with impaired hearing unable to focus on their chosen speaker due to too much background noise. To improve the living conditions of hearing-impaired people, a study was planned using Auditory Attention Decoding (AAD) to understand the processing of directional hearing in the brain during a listening task with different speakers and distracting noises. Out of three initial concepts, one idea was transformed into a program. AAD provided a possibility for the development of a study and its technical preparation. The results are intended to provide insight into the processing of incoming audio signals performed by the human brain.

I. Introduction

Hearing loss is considered to be a very common problem in the elderly community [1], [2]. An article by Mühlbauer et al. [3] indicates that a quarter of the population over the age of 60 years and half the population over the age of 70 years is suffering from age-related hearing loss. Age-related hearing loss does not only lead to decreased cognitive function, but also affects the general well-being, potentially leading to depression or self-isolation [1], [2]. Although hearing aids can improve the ability to process spoken language, loud and busy environments or multiple people speaking at once can significantly lessen their effectiveness. Oftentimes, hearing aid users cannot participate in social gatherings due to insufficiently differentiating hearing aids when listening to multiple speakers. This phenomenon is commonly referred to as the "cocktail party problem" [4].

Currently, hearing aids operate based on heuristic processes which select and enhance speakers e.g. based on the line of sight of the user or the volume of the speaker. The enhanced speaker is not necessarily the speaker the user is trying to focus on [4]. Furthermore, some hearing aids may perform poorly in noisy environments, especially if more than one speaker is actively talking, because, while the separation of noise from speech may function, separating different speakers can prove to be more of a challenge [2]. Thus, the current technology can possibly still be improved by a different approach.

A strategy is being explored in which the information about the source of sound a person focuses on is deduced directly from the brain. In a multi-speaker scenario, neural signals obtained by electroencephalography (EEG) can be used to decode the source of sound a hearing aid user focuses on. This strategy, which can potentially be used to develop neurologically steered hearing aids, is called Auditory Attention Decoding (AAD) [4], [2]. When the objective of an individual's auditory attention is known, it can be used to enhance the auditory stimulus the individual wants to focus on and suppress back-

ground noise [2].

The planned study aims to examine different possible implementations in a test environment in order to lay the foundations for further developments in the field of auditory enhancement.

II. Methods and Materials

For the research project, included tasks were the construction of a study and the technical implementation to make the execution thereof possible. The article by Orf et al. [5] proved to be a reasonable starting point.

Research on the topic of Auditory Attention Decoding (AAD) was a useful tool for a clear construction of the study, making the conduct more efficient in the future. The study aims to use AAD to gain information on the way auditory signals are perceived in the brain. To achieve this, a person is positioned in front of five evenly distributed speakers within a 180 ° radius as shown in Figure 1. Each of the speakers is set to an equal location and volume for all participants.

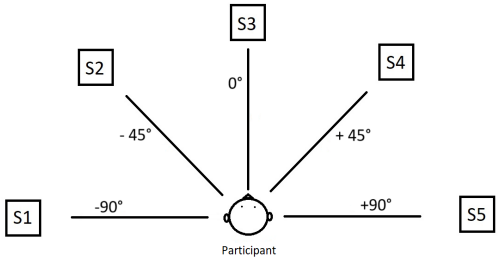

Figure 1: Setup of the Speakers. The participant is sitting in a semi-circle of equidistantly arranged audio speakers (S1-S5). The speakers are placed at 0 °, + 45 °, - 45 °, + 90 ° and - 90 ° relative to the participant's line of vision.

For the duration of the measurement, participants are asked to wear a semi-dry / saline-based EEG-cap containing 23 channels coupled to a Smarting mobi amplifier. Each speaker is set to play one sound, so there can be up to five different sounds at the same time. At all times, two audio books are played simultaneously, one narrated by a male voice, one by a female voice. The participant is asked to concentrate on one voice at a time and ignore the other, as well as any other distracting sounds coming from other speakers at the same time. In total, one trial consists of thirty 60 s long sound fragments, so the attention may be redirected up to 29 times.

II.I. Participants

Participants must be of normal hearing and understand German on a native level, as the audio books are narrated in German. Additionally, participants must be able to focus on an audio track and switch their attention to a different audio track on command. While the first two attributes can be confirmed verbally by participants, the third can be tested in the program. Participants must be able to press a button on command or answer a short question with little response time. Furthermore, they should be between 18 and 30 years old to represent a realistic, healthy adult's hearing capacity [5]. Preferably, a group of at least 20 individuals, both male and female, should participate in the study. Lastly, participants are asked to keep their eyes open during measurements in case a visual instruction is required to indicate which audio book they should direct their focus on.

II.II. Audio Books

Generally, the selected audio books should be easy to focus on. To ensure that the measurement is free of gender-specific artifacts, an equal amount of male- and female-narrated books must be chosen. Stuttering during the narration is unacceptable, as it interferes with the control mechanism for the attention. As briefly mentioned, the study had been outlined before it was constructed more clearly. In the course of the outline, two audio books were provided that are called: "Nero Corleone kehrt zurück: es ist immer genug Liebe da" written by Elke Heidenreich and Quint Buchholz [6] and "Sapiens: eine kurze Geschichte der Menschheit" written by Yuval Noah Harari [7]. The two audio books that were chosen during the project to broaden the range of genres are: "Rubinrot: Liebe geht durch alle Zeiten", the first book in a trilogy by Kerstin Gier [8] and shorter stories called "Ein Traum", "Der Kübelreiter", "Der Geier", "Die Sorge des Hausvaters", "Ein Brudermord" and "Ein altes Blatt" by Franz Kafka, to be found in a collection of stories by Kafka called "Sämtliche Erzählungen" [9]. The audio narrations for Gier and Kafka are openly available [10], [11].

II.III. Implementation

The audio books were subdivided into shorter, 60 s long sound fragments in order to be able to play them sequentially. Afterwards, the basic main program was constructed using PsychoPy [12]. Further adaptation of the program to the specific needs of the study was necessary. For example, the entirety of the audio fragments and the order they are played in had to be inserted into the code. Additionally, the necessary components for enabling the control mechanism of the first version explained in the following needed to be set up. A synchronisation with the EEG device and speakers has not been added to the program yet, but will be necessary before any measurements can be made. Solutions for this issue are being explored and include the use of Lab Streaming Layer (LSL), an open-source software. Whereas the exact results of the

10.18416/SCP.2025.1935

study cannot currently be foreseen, correlations between the direction of the source of the sound and the resulting brain activity are expected.

Different approaches to ensure participants' attention have been considered. In the first version, a short repetition of a sound fragment within the audio book the participant is supposed to focus on has to be acknowledged by pressing a button. Each repetition was assigned a duration of 0.5 s. In each minute, repetitions are possible at six equidistantly spaced moments. It was decided, that a total of three repetitions should occur within a minute in unpredictable intervals to make manipulation of the study less likely. Another version is very similar, but instead of a repetition, a new verbal cue is introduced. The screen shows four different shapes in different colours and one of those shapes is cued, e.g. "blue triangle". The colours are supposed to shorten search times, but in case of colour-blindness, a participant can still rely on the shapes. Participants are then asked to indicate the named shape. The last version uses a questionnaire to confirm the focus was correctly directed. As the sounds the participants are supposed to focus on are played sequentially for 1 min each, questions need to be found for each 60 s long soundtrack. An extensive table has been created, containing at least two to three questions correlated to each minute of each audio book. At least one question is correlated to the first 30 s and at least one more to the other 30 s of a minute. In total, 278 questions were considered suitable to ensure the participant's attention was directed at the correct source. The questions were developed to be easily answered, e.g. asking for the main character's name or asking about an important statement made in that minute. During the time it takes the participant to answer the question, the audio track needs to be paused.

III. Results and Discussion

Although all approaches could work, it became apparent that one idea is more uniform for all participants than the other two. After discarding versions two and three of the program for different reasons explained hereinafter, version one has been implemented in large parts.

Initially thought to be a viable option, the second program was discarded due to complexity. For successful implementation, it would have been necessary to give specific verbal cues in the voice of the current narrator. Using a different voice to utter the cues would potentially cause the participant to focus on a change of voice, rather than paying attention to the audio book, possibly falsifying data. As the audio books do not naturally contain the cues anywhere in the written text, they would have needed to be generated in the audio book narrators' voices. Furthermore, the audio track would either have needed to be paused or the pictured shapes changed without the program receiving any input if the participant missed the cue. This would, in turn, provide feedback about a missed cue to the participant. Possibly resulting in negative emotions such as frustration, this might have impacted the study.

The reasons for discarding the third version include the possibility of the existence of prior knowledge about one or more of the audio books as a result of participants having read or listened to them before. Consequentially, the participant would not necessarily need to pay attention to the audio book in order to give correct answers, potentially causing major distortions of the measurements. Furthermore, the audio track would constantly have to be paused for the participant to answer questions concerning a specific sound fragment. As this would cause unnecessary interruptions, measurements could be distorted further. Additionally, the individual times needed to conduct the study would vary between participants, depending on their reading speed and ability to answer the questions. Generally, this version is more unpredictable than the other versions. Nevertheless, questions that could be put on a questionnaire have been collected in a table, sorted by minutes, because it was initially unclear which program would be implemented in the end.

As the advantages of the first version became apparent, it was chosen to be implemented as visualised in Figure 2. Using an extra button, the participant can acknowledge perceived clues without immediately receiving feedback on their performance. Thus, negative emotions cannot be triggered and, consequentially, cannot distort the measurement. Although the 0.5 s long audio fragments can only be repeated at six given times within a minute, it is possible to randomise the intervals between their occurrences. The method is not ideal, but sufficient for the purpose of the study. All repetitions were chosen individually to ensure that they could actually be perceived.

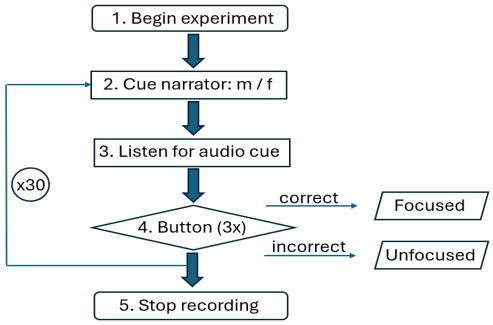

Figure 2: Visualisation of the first version of implementation. Steps 2-4 represent 1 minute each. One entire trial takes 30 minutes, thus steps 2-4 must be repeated 30 times in total. The attention may be redirected towards male (m) or female (f) narration at the beginning of each minute.

In order to guarantee an extraction of the genuine information on the direction of the attention and not the detection of a pattern, an extensive table was created in which each audio speaker was assigned an ordered list of sounds ranging from the actual audio books to possible distractions that may be played at a certain time. With the structure in the table, it is possible to permute the direction the sound is played from, making the randomisation of the direction possible before beginning the measurements. Randomising the directions in the beginning makes the measurements less likely to be influenced by repeating patterns. Note that a decision regarding the mechanism of cuing the audio book that should be focused on has not been made yet. Possible options are verbal cuing by the conductor of the study or visual cuing on a screen.

IV. Conclusion

The project's aim was the technical development of a study based on the concept of Auditory Attention Decoding to research the way the brain processes the direction of incoming audio signals.

When developing a concept for a study, it is important to consider many options and choose the most reasonable one. In this case, in order to minimise the amount of data distortion, the options that were most prone to corruption were eliminated. The first version was believed to provide the most uniform measurements.

The second option proved to be much more complex than the first, while providing no further advantages to balance out the complexity. In addition, it was possible to give correct answers without actually concentrating on the indicated track, which was a reason not to implement the idea. The third option was eliminated due to the possibility of too much prior knowledge of the audio books and the flow of the audio constantly being interrupted, thus interrupting the measurement itself. The unpredictability of this option was considered a disadvantage. In conclusion, the disadvantages of both the second and third version outweighed the advantages compared to the first version of implementation which is very similar in structure, but provides more uniform measurements.

With an aging society, the problem of age-related hearing loss should be taken seriously. Auditory attention decoding (AAD) provides a possibility for the development of neurologically steered hearing aids in comparison to the heuristically steered hearing aids currently provided to users. Thus, the study described in this report may provide an initial step towards technological progress in the field of hearing aids.

Based on the internship's work, a master thesis will cover the technical preparation and test runs for the study described above, thus, hopefully, contributing to the enhancement of future generations of hearing aids. Furthermore, the first of the three programs has been implemented in large parts and can most probably be used to conduct the study once it is finished. The setup of the speakers and usage of the EEG-system will be the main part of the research for the master thesis.

Acknowledgments

The work has been carried out and supervised by the Institute for Signal Processing, Universität zu Lübeck.

Karina Rohm would like to thank Mr J. Schillhorn for enabling her to work on this project by providing her with a laptop, when her own broke. This was essential to the completion of this project.

Author's statement

Conflict of interest: Authors state no conflict of interest.

References

[1] S. Moser, W. Luxenberger, and W. Freidl. Wahrnehmung von Hörproblemen bei älteren Menschen. *HNO*, 65(8):671–679, 2017.

[2] G. Ciccarelli, M. Nolan, J. Perricone, P. T. Calamia, S. Haro, J. O'Sullivan, N. Mesgarani, T. F. Quatieri, and C. J. Smalt. Comparison of Two-Talker Attention Decoding from EEG with Nonlinear Neural Networks and Linear Methods. *Scientific Reports*, 9(1):11538, 2019.

[3] S. Mühlbauer. Hörverlust bei älteren erschwert die Kommunikation. *Heilberufe*, 69(12):21–21, 2017.

[4] S. Geirnaert, S. Vandecappelle, E. Alickovic, A. De Cheveigne, E. Lalor, B. T. Meyer, S. Miran, T. Francart, and A. Bertrand. Electroencephalography-Based Auditory Attention Decoding: Toward Neurosteered Hearing Devices. *IEEE Signal Processing Magazine*, 38(4):89–102, 2021.

[5] M. Orf, M. Wöstmann, R. Hannemann, and J. Obleser. Target enhancement but not distractor suppression in auditory neural tracking during continuous speech. *iScience*, 26(6):2022–06, 2023.

[6] E. Heidenreich and Q. Buchholz, Nero Corleone kehrt zurück: es ist immer genug Liebe da. München: Dt. Taschenbuch-Verl, 2013,

[7] Y. N. Harari, Sapiens: eine kurze Geschichte der Menschheit. München: Penguin Verlag, 2024,

[8] K. Gier, Rubinrot: Liebe geht durch alle Zeiten. Würzburg: Arena, 2009,

[9] F. Kafka, Sämtliche Erzählungen. München: Anaconda, 2024,

[10] (2017). "Edelsteintrilogie - Rubinrot - Liebe geht durch alle Zeiten - Hörbuch — YouTube", URL: https://www.youtube.com/watch?v=pKLsA3LqkO8&list=PLS1Dih4YvMkuzm-WEYXOG6IMFmCuBQKpJ (visited on 03/13/2024).

[11] (2021). "Franz Kafka: Die besten Kurzgeschichten | Hörbuch Compilation zum Einschlafen (ohne Werbepausen) — YouTube", URL: https://www.youtube.com/watch?v=S7qvbOzg-pc&list=PLS1Dih4YvMkuzm-WEYXOG6IMFmCuBQKpJ&index=2 (visited on 03/13/2024).

[12] J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195–203, 2019.