Evaluation of Machine Learning Algorithms for Speech Prioritisation in Noisy Environments

Nikodem Bieniek

MInf Project (Part 1) Report
Master of Informatics
School of Informatics
University of Edinburgh

2024

Abstract

This skeleton demonstrates how to use the `infthesis` style for undergraduate dissertations in the School of Informatics. It also emphasises the page limit, and that you must not deviate from the required style. The file `skeleton.tex` generates this document and should be used as a starting point for your thesis. Replace this abstract text with a concise summary of your report.

Research Ethics Approval

Instructions: *Agree with your supervisor which statement you need to include. Then delete the statement that you are not using, and the instructions in italics.*
Either complete and include this statement:
This project obtained approval from the Informatics Research Ethics committee.
Ethics application number: ???
Date when approval was obtained: YYYY-MM-DD
[If the project required human participants, edit as appropriate, otherwise delete:]
The participants' information sheet and a consent form are included in the appendix.
Or include this statement:
This project was planned in accordance with the Informatics Research Ethics policy. It did not involve any aspects that required approval from the Informatics Research Ethics committee.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Nikodem Bieniek*)

Acknowledgements

Any acknowledgements go here.

Table of Contents

Chapter 1

Introduction

1.1 Motivations

Hearing loss is a prevalent condition affecting as much as 430 million people - or 1 in 18 people. This is expected to rise to 1 in 10 by 2050 [5]. The most common treatment for hearing loss is the provision of hearing technology - such as Hearing Aid (HA) or cochlear implants. There are many types of HAs, but the most common type is the behind-the-ear (BTE) hearing aid [3]. However, HA users often report that they struggle to hear speech in noisy environments. For example, in a study by Kochkin [3], 42% of HA users reported that wind noise was a significant issue for them. This project aims to evaluate the effectiveness of machine learning algorithms in prioritising the speech in various environments (such as windy environments).

Modern hearing aids now apply a wide range of techniques to achieve better speech prioritisation. For wind noise reduction, this can be achieved from mechanical solutions - product design to covers that reduce wind noise - to signal processing techniques to compensate for mechanical limitations. However, current techniques are still not perfect as shown by the study from Kochkin.

Wind noise reduction and indeed, noise reduction in general, is a challenging problem when paired with speech. This is because you have to strike a balance between reducing background noise and preserving speech. Korhen's paper [4] outlines various techniques that could be used to reduce the wind noise in hearing aids - from modulation-based noise reduction algorithms (Wiener filtering), adaptive filtering algorithms, to machine learning techniques. The paper mentions that the the proposed ML technique: Long Short-Term Memory (LSTM) neural networks provided modest improvements in wind noise reduction, however, it did highlight that ML techniques may still have utility through further research and careful algorithmic choices.

This project aims to pair the proposed machine learning techniques with signal processing techniques to evaluate the effectiveness of speech prioritisation in noisy environments. The idea is to first perform acoustic scene analysis (ASA) to classify the environment based on the audio signal. Afterwards, speech enhancement techniques will be applied to the signal to prioritise speech. ASA can be done using a dataset of varying environments, and using machine learning techniques to classify the environment.

In this project, we will be using a novel dataset proposed by Huwel et al. [2] This dataset (called HEAR-DS) is unique because it is specially tailored for HA signal processing and contains various environments. Normally, voice activity detection (VAD) is used to detect speech, however, the dataset helpfully contains labels which samples contain speech. This can be used to implicitly train the machine learning model to classify the environment and whether speech is present. Additionally, the paper presents an elementary example of how the dataset can be used: to classify the environment - it showcases the use of a convolutional neural network (CNN) to classify the environment. This project will be extending the paper by actually using the dataset and comparing various machine learning techniques to evaluate the effectiveness of the proposed machine learning techniques.

This project will investigate how recurrent neural networks (RNN) and their variants, such as LSTMs, can be used to classify the environment. Based on the environment classified, the system will then apply signal processing techniques to enhance the speech. The project however has to be mindful in its algorithmic choices - as the computational power required in HA is limited. It is difficult to pinpoint the exact computational power of a HA due to the proprietary nature of the devices. In August 2024, Phonak (Sonova Holding AG) released a new HA which is their first AI equipped HA. The device is said to be capable of handling 7,700 Million Operations Per Second to accommmodate its neural network with 4.5 million parameters. Contrast this with a paper from 2021 investigating techniques in VAD for hearing aids quotes that it ‘rarely exceeds 5 million instructions per second (MIPS)‘ [1] Moreover, Apple’s release of a FDA approved hearing aid in which was previously a mainstream earphone wearable, Apple Airpods, also shows that there’s more interest in this area. So suffice to say, the computational power of hearing aids is accelerating and is most likely going to continue to grow.

Delay constraints are also important in HA, as the user needs to hear the speech in real-time. For example, so long as the speech production is no higher than 30 milliseconds (ms), and ideally less than 20ms, the user is unaffected by the delay [6].

The project is predominantly aimed at the hearing aid industry. If successful, the project could advance the state-of-the-art techniques in hearing aids. Which in turn could improve the quality of life for hearing aid users. A secondary goal is to make hearing aid research more accessible to the computer science community. The project also hopes to benefit other fields that deal with audio signal processing, such as mainstream wearables like headphones or microphones.

Chapter 2

Background

Chapter 3

Conclusions

3.1 Final Reminder

The body of your dissertation, before the references and any appendices, *must* finish by page 40. The introduction, after preliminary material, should have started on page 1.

You may not change the dissertation format (e.g., reduce the font size, change the margins, or reduce the line spacing from the default single spacing). Be careful if you copy-paste packages into your document preamble from elsewhere. Some LaTeX packages, such as `fullpage` or `savetrees`, change the margins of your document. Do not include them!

Over-length or incorrectly-formatted dissertations will not be accepted and you would have to modify your dissertation and resubmit. You cannot assume we will check your submission before the final deadline and if it requires resubmission after the deadline to conform to the page and style requirements you will be subject to the usual late penalties based on your final submission time.

Bibliography

[1] Joaquín García-Gómez, Roberto Gil-Pita, Miguel Aguilar-Ortega, Manuel Utrilla-Manso, Manuel Rosa-Zurera, and Inma Mohino-Herranz. Linear detector and neural networks in cascade for voice activity detection in hearing aids. *Applied Acoustics*, 175, 2021. doi: 10.1016/j.apacoust.2020.107832.

[2] Andreas Hüwel, Kamil Adiloğlu, and Jörg-Hendrik Bach. Hearing aid research data set for acoustic environment recognition. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020. doi: 10.1109/ICASSP40776.2020.9053611.

[3] Sergei Kochkin. Marketrak VIII: Consumer satisfaction with hearing aids is slowly increasing. *The Hearing Journal*, 63(1), 2010. doi: 10.1097/01.HJ.0000366912.40173.76.

[4] Petri Korhonen. Wind Noise Management in Hearing Aids. *Seminars in Hearing*, 42(3), 2021. doi: 10.1055/s-0041-1735133.

[5] World Health Organization. Deafness and hearing loss, 2024. URL `https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss`.

[6] Michael A. Stone and Brian C. J. Moore. Tolerable Hearing Aid Delays. II. estimation of limits imposed during speech production. *Ear and Hearing*, 23(4), 2002. doi: 10.1097/00003446-200208000-00008.

Appendix A

First appendix

A.1 First section

Any appendices, including any required ethics information, should be included after the references.

Markers do not have to consider appendices. Make sure that your contributions are made clear in the main body of the dissertation (within the page limit).

Appendix B

Participants' information sheet

If you had human participants, include key information that they were given in an appendix, and point to it from the ethics declaration.

Appendix C

Participants' consent form

If you had human participants, include information about how consent was gathered in an appendix, and point to it from the ethics declaration. This information is often a copy of a consent form.