The Sensitive Ear: Towards Real-Time Monitoring of In-Ear Physiological and Audio Signals for Managing Decreased Sound Tolerance

by

Danielle BENESCH

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT OF A MASTER'S DEGREE
WITH THESIS
M.A.Sc.

MONTREAL, DECEMBER 21, 2022

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Prof. Jérémie Voix, Thesis supervisor
Department of Mechanical Engineering, École de technologie supérieure

Prof. Rachel Bouserhal, Thesis Co-Supervisor
Department of Electrical Engineering, École de technologie supérieure

Prof. Catherine Laporte, Chair, Board of Examiners
Department of Electrical Engineering, École de technologie supérieure

Prof. Alessandro Lameiras Koerich, Member of the Jury
Software and Information Technology Engineering, École de technologie supérieure

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON DECEMBER 16, 2022

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Jérémie Voix for giving me this opportunity to pursue research in an area that is personally meaningful to me. I am very grateful to him and my co-supervisor Rachel Bouserhal for their continued support and guidance throughout. Thank you both for teaching me about new fields, encouraging me to develop my ideas, being so patient and understanding, and making the time to help me get to this point. Thank you to everyone at CRITIAS who supported me in so many different ways. In particular, I would like to acknowledge the contributions of three former CRITIAS interns: Corentin Delain for his help with data acquisition, Nithin Raj for his work with the Tympan platform, and Léna Tallon for biosignal annotation. I would like to thank Sylvie Hébert, Alain Vinet and Bérangère Villatte for giving me the opportunity to be part of the AUDACE project, through an internship at Université de Montréal, and learn from their expertise. I would also like to thank everyone at EERS who helped me during my MITACS internship, in particular Ajin Tom and Philippe Chabot for providing input on the analysis, Chahinez Hocine for collecting data, and Shruti Kshirsagar for helping me with documentation. I would like to acknowledge the financial support of EERS, CRITIAS, DAAD, CIRMMT, and MITACS through the Accelerate and Globalink Graduate Fellowship programs. Lastly, I would like to thank my parents, sister, friends and partner for their support and encouragement, virtually and in person.

L’oreille sensible : Vers une surveillance en temps réel des signaux physiologiques et audio intra-auriculaires pour gérer la tolérance réduite au bruit

Danielle BENESCH

RÉSUMÉ

Les personnes dont la tolérance au bruit est réduite portent souvent des protections auditives ou des casques d’écoute isolants pour bloquer les sons auxquels elles sont sensibles. Cependant, ces protecteurs auditifs conventionnels bloquent tous les sons sans distinction, même lorsque ces sons ne causent pas de détresse et pourraient être utiles au porteur. En effet, l’isolation acoustique procurée par ces protecteurs peut empêcher son porteur de participer pleinement à son environnement et pourrait également avoir des conséquences à long terme, puisque la recherche suggère que l’évitement régulier du bruit peut exacerber la sensibilité au son. Une solution potentielle à l’évitement excessif du bruit pourrait être un appareil auditif spécifiquement conçu pour une tolérance sonore réduite, capable d’atténuer le son uniquement lorsque cela est nécessaire pour réduire le stress du porteur, en enregistrant les sons externes, en les traitant et en les transmettant sélectivement à l’oreille.

Cependant, le développement d’un tel appareil présente plusieurs défis : 1) Les dispositifs existants utilisés pour gérer la diminution de la tolérance aux sons ne sont pas facilement modifiables, et les ressources nécessaires pour développer indépendamment un nouveau dispositif peuvent constituer un obstacle au prototypage de nouvelles fonctionnalités ; 2) Il existe peu de recherches sur la meilleure façon de prendre en compte l’équilibre entre une performance d’utilisation accrue, telle l’efficacité du filtrage des sons pénibles. et les impacts potentiels de ce filtrage sur la compréhension de la parole. Ce pourrait être une préoccupation particulière pour les populations pour lesquelles les symptômes de tolérance sonore réduite coexistent avec des difficultés de communication ; 3) La forme revêtie par la tolérance réduite aux sons peut varier considérablement d’une personne à l’autre, ainsi qu’au fil du temps pour une même personne, et par conséquent, le traitement audio requis pourrait varier grandement en fonction des besoins actualisés de chaque porteur.

Cet mémoire présente les résultats des efforts de recherche visant à relever les défis énumérés ci-dessus : Premièrement, les plateforme matérielles ("hardware") et logicielles existants pour le prototypage d’algorithmes d’appareils auditifs ont été évaluées pour leur adéquation avec des activités de recherche sur la tolérance au son diminué. Un pipeline de traitement audionumérique reliant un ordinateur externe à une plate-forme équippée d’un microcontrôleur a ainsi été développé pour permettre aux algorithmes mis en œuvre sur l’ordinateur de contrôler le traitement audio embarqué en temps réel. Deuxièmement, une expérimentation en ligne sur 50 participants a été mis au point pour déterminer les exigences en terme de latence pour le traitement audionumérique en fonction de l’impact causé sur une tâche nécessitant la compréhension de la parole par des retards simulés. Troisièmement, un système de détection automatique du stress a été mis au point et validé à partir d’un ensemble de données provenant des battements cardiaques de 30 participants enregistrés dans le conduit auditif au cours de tâches induisant du stress. Ces résultats ouvrent ainsi la voie à un appareil auditif capable de s'adapter à l'utilisateur en surveillant simultanément le niveau de stress et les bruits à l'origine de ce stress pour des pertes souffrant de tolérance sonore réduite.

Mots-clés: tolérance sonore réduite, appareils auditifs, reconnaissance automatique du stress

The Sensitive Ear: Towards Real-Time Monitoring of In-Ear Physiological and Audio Signals for Managing Decreased Sound Tolerance

Danielle BENESCH

ABSTRACT

People with decreased sound tolerance often wear hearing protection or headphones to block out the sounds they are sensitive to. However, conventional hearing devices indiscriminately block out all sounds, even when these sounds do not cause distress and may be useful for the wearer to hear. The resulting acoustic isolation can prevent the wearer from fully participating in their environment and could also have long term consequences, as research suggests that regular noise avoidance may exacerbate sensitivity to sound. A potential solution to over-avoidance of noise could be a hearing device specifically designed for decreased sound tolerance to attenuate sound only when necessary to reduce the wearer's stress, by recording external sounds, processing them, and selectively transmitting them inside the earcanal.

However, there are several challenges associated with developing a hearing device that selectively attenuates distressing sounds: 1) Existing devices used to manage decreased sound tolerance are not readily modifiable, and the resources required to independently develop a new device can be a barrier to prototyping new features; 2) There is limited research on how to balance usability considerations such as how effectively distressing sounds are filtered out and potential impacts on speech comprehension, which could be of particular concern for populations where decreased sound tolerance symptoms co-occur with communication challenges; 3) The presentation of decreased sound tolerance can vary greatly, for different individuals as well as over time for the same individual, and therefore, the audio processing required could depend on each wearer's current needs.

This thesis presents the findings of research efforts aiming to address the above challenges: First, existing hardware and software tools to prototype hearing device algorithms were evaluated for their suitability for decreased sound tolerance research. A processing pipeline interfacing an external personal computer with a microcontroller-based platform was developed to allow for algorithms implemented on the computer to control real-time audio processing. Second, an experimental protocol, piloted online with 50 participants, was developed to assess the latency requirements for real time audio processing based on how simulated delays affected a task necessitating speech comprehension. Third, with a dataset of 30 participants' heartbeat sounds recorded from a hearing device during a stress-induction protocol, an automatic stress recognition system was developed and validated. These results open the possibility for a hearing device that can adapt to the wearer by simultaneously monitoring stress and controlling triggering noise for the benefit of users with decreased sound tolerances.

Keywords: decreased sound tolerance, hearing devices, automatic stress recognition

TABLE OF CONTENTS

Page

INTRODUCTION 1

0.1 Context 1

0.2 Motivation 1

0.2.1 Decreased sound tolerance and over-avoidance 1

0.2.2 Barriers to development of hearing technology for decreased sound tolerance 2

0.3 Objectives 3

0.4 Structure of the thesis 4

0.4.1 Chapter 1 - Literature review 4

0.4.2 Chapter 2 - Interfacing the Tympan open-source hearing aid with an external computer for research on decreased sound tolerance 4

0.4.3 Chapter 3 - Evaluating the effects of audiovisual delays on speech understanding with hearables: A pilot study 5

0.4.4 Chapter 4 - Automatic stress recognition with in-ear heartbeat sounds 5

0.4.5 Chapter 5 - Conclusions and Recommendations 5

CHAPTER 1 LITERATURE REVIEW 7

1.1 Decreased sound tolerance 7

1.1.1 Characteristics 7

1.1.2 Assessment 8

1.1.3 Neurophysiology 8

1.1.4 Interventions 9

1.2 Automatic stress recognition 11

1.3 Summary 12

CHAPTER 2 INTERFACING THE TYMPAN OPEN-SOURCE HEARING AID WITH AN EXTERNAL COMPUTER FOR RESEARCH ON DECREASED SOUND TOLERANCE 13

2.1 Abstract 13

2.2 Introduction 13

2.3 Background 15

2.4 Approaches 16

2.4.1 ARP 3.1 hardware connected to a PC 16

2.4.1.1 Latency test setup 17

2.4.1.2 Latency test results 21

2.4.2 Standalone Tympan hardware 22

2.4.2.1 CPU test setup 23

2.4.2.2 CPU test results 24

2.4.3 Tympan hardware connected to a PC 24

2.4.3.1 Proof of concept of processing pipeline 25
2.5 Discussion 27
2.6 Acknowledgments 28

CHAPTER 3 EVALUATING THE EFFECTS OF AUDIOVISUAL DELAYS ON SPEECH UNDERSTANDING WITH HEARABLES: A PILOT STUDY 31
3.1 Abstract 31
3.2 Introduction 32
3.3 Methods 36
3.3.1 Materials 37
3.3.2 Participants 44
3.3.3 Procedure 44
3.3.4 Analysis 46
3.4 Results 47
3.4.1 Reaction times 47
3.4.2 Response accuracy 50
3.4.2.1 Referent choice task 50
3.4.2.2 Simultaneity judgments 51
3.4.2.3 Distortion judgments 52
3.5 Discussion 53
3.6 Conclusions 57
3.7 Acknowledgements 57

CHAPTER 4 AUTOMATIC STRESS RECOGNITION WITH IN-EAR HEART-BEAT SOUNDS 59
4.1 Abstract 59
4.2 Introduction 60
4.3 Background 61
4.4 Methods 63
4.4.1 Dataset 63
4.4.1.1 Participants 63
4.4.1.2 Materials 63
4.4.1.3 Procedure 65
4.4.2 Preprocessing 66
4.4.2.1 Task segmentation 66
4.4.2.2 Synchronization 68
4.4.2.3 Heartbeat annotations 68
4.4.3 Prediction 69
4.4.3.1 Feature extraction 69
4.4.3.2 Error-balanced data synthesis 69
4.4.3.3 Feature normalization 71
4.4.3.4 Classification algorithms 72
4.4.3.5 Evaluation 73

4.5 Results 74
4.5.1 Baseline normalization 74
4.5.2 Comparison of ECG and in-ear audio 75
4.5.3 Error-balanced data synthesis 81
4.6 Discussion 85
4.6.1 Baseline normalization 85
4.6.2 Comparison of ECG and in-ear audio 86
4.6.3 Error-balanced data synthesis 87
4.7 Conclusions 89
4.8 Acknowledgments 90

CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS 91
5.1 Potential applications 91
5.2 Limitations, considerations and recommendations 92
5.3 Contributions 94
5.3.1 Industrial 94
5.3.2 Scientific 94

APPENDIX I LIST OF IMAGE STIMULI FROM CHAPTER 3 97

APPENDIX II LIST OF SENTENCE STIMULI FROM CHAPTER 3 103

APPENDIX III SUPPLEMENTARY STATISTICAL ANALYSES FROM CHAPTER 3 111

APPENDIX IV DESCRIPTION OF FEATURES FROM CHAPTER 4 119

APPENDIX V SUPPLEMENTARY EXPERIMENTS FROM CHAPTER 4 129

APPENDIX VI INSTITUTIONAL REVIEW BOARD DOCUMENTATION 133

LIST OF REFERENCES 143

LIST OF TABLES

Page

Table 2.1 Estimates of the latency of audio pass-thru 22

Table 2.2 CPU Usage of wide dynamic range compression on Tympan Rev D 24

Table 3.1 The names and descriptions of the delay conditions 37

Table 3.2 Model coefficients for referent choice task responses: comparison against baseline 49

Table 3.3 Model coefficients for referent choice task responses: contrast only AV vs. AV & RA 50

Table 3.4 Model coefficients for the simultaneity judgment task responses 53

Table 4.1 Accuracy of stress classifiers trained on ECG and tested on ECG with baselining 75

Table 4.2 Accuracy of stress classifiers trained on ECG and tested on ECG without baselining 75

Table 4.3 Accuracy of stress classifiers trained on IEM and tested on IEM without baselining 76

Table 4.4 Accuracy of stress classifiers trained on IEM and tested on ECG without baselining 77

Table 4.5 Accuracy of stress classifiers trained on IEM and tested on SYN without baselining 83

Table 4.6 Accuracy of stress classifiers trained on ECG and tested on IEM without baselining 84

Table 4.7 Accuracy of stress classifiers trained on ECG & SYN using all HRV features without baselining 84

LIST OF FIGURES

Page

Figure 2.1 Auditory Research Platform 3.1 17

Figure 2.2 Diagram of the setup used to estimate the latency of audio pass-thru 19

Figure 2.3 5 ms input signal used to measure audio pass-thru latency 20

Figure 2.4 Example recording used to estimate latency of audio pass-thru 21

Figure 2.5 Tympan Rev D with earpieces 23

Figure 2.6 Diagram of Tympan-PC pipeline 25

Figure 2.7 Diagram of white noise toggle 26

Figure 2.8 Recording of white noise toggle 27

Figure 3.1 Overview of a hearable 33

Figure 3.2 Representation of the delay conditions 39

Figure 3.3 Transfer function of the earpiece 40

Figure 3.4 Example screens from the referent choice and SJ tasks 43

Figure 3.5 Reaction times on the referent choice task 48

Figure 3.6 Accuracy on the referent choice task 51

Figure 3.7 Responses for the simultaneity judgment task 53

Figure 4.1 Auditory Research Platform earpiece 64

Figure 4.2 Diagram of synthetic data generated with error 69

Figure 4.3 Diagram of prediction pipeline 74

Figure 4.4 Box plot of feature coefficients 78

Figure 4.5 Example IEM recording and automatically detected heartbeats 79

Figure 4.6 Bland-Altman plots assessing agreement between ECG and IEM 80

Figure 4.7 Bland-Altman plots assessing agreement between ECG and SYN 82

LIST OF ALGORITHMS

Page

Algorithm 4.1 NumPy-style pseudocode for error-balanced data synthesis 71

LIST OF ABBREVIATIONS

ARP	Auditory Research Platform
AV	Audiovisual
BPM	Beats per minute
CERC	Comité d’éthique pour la recherche clinique
CPU	Core Processor Unit
DSP	Digital Signal Processor
ECG	Electrocardiography
ÉTS	École de technologie supérieure
GUI	Graphical User Interface
HATS	Head and Torso Simulator
HRV	Heart Rate Variability
IEM	In-Ear Microphone
LFHF	Ratio of Low Frequency to High Frequency power
MedianNN	Median heart period
OEM	Outer-Ear Microphone
PC	Personal Computer
PTA	Pure Tone Average
RA	Residual Audio
RMSSD	Root-Mean-Square of Successive Differences

RT	Reaction Time
SD	Standard Deviation
SDANN2	Standard Deviation of the Average of Normal-to-Normal intervals over two minutes
SE	Standard Error
SPK	Loudspeaker
SJ	Simultaneity Judgment
SYN	Synthetic data
TSST	Trier Social Stress Test
WDRC	Wide Dynamic Range Compression
XGBoost	Extreme gradient boosting

LIST OF SYMBOLS AND UNITS OF MEASUREMENTS

dB	decibel
dBA	A-weighted decibel
dB HL	decibel for Hearing Level
dB SPL	decibel for Sound Pressure Level

INTRODUCTION

0.1 Context

For certain individuals, everyday sounds can be intolerable to hear. Decreased tolerance to sounds can heavily impair quality of life, interfering with education, work and interpersonal relationships (Fackrell *et al.*, 2019; Jager, de Koning, Bost, Denys & Vulink, 2020; Cavanna & Seri, 2015; Brout *et al.*, 2018; Frank & McKay, 2019). Currently, there is limited research on decreased sound tolerance and no well-established treatment protocol (Fackrell *et al.*, 2019; Raj-Koziak, Gos, Kutyba, Skarzynski & Skarzynski, 2021; Brout *et al.*, 2018).

0.2 Motivation

0.2.1 Decreased sound tolerance and over-avoidance

Hearing protection or noise-cancelling headphones are frequently worn by those with decreased sound tolerance (Cavanna & Seri, 2015; Neave-DiToro, Fuse & Bergen, 2021; Jastreboff & Jastreboff, 2014; Frank & McKay, 2019; Sammeth, Preves & Brandy, 2000). Conventional passive hearing protection devices and noise-cancelling headphones, however, block out all background noise without differentiating between intolerable and useful sounds such as speech. Therefore, conventional hearing devices like earplugs and headphones may attenuate the sound too much for the wearer, which can make it difficult for those with decreased sound tolerance to completely engage with their surroundings (Pfeiffer, Erb & Slugg, 2019a; Sammeth *et al.*, 2000; Neave-DiToro *et al.*, 2021). Moreover, excessive noise avoidance may be counterproductive in the long term, potentially exacerbating sound sensitivity over time (Pienkowski *et al.*, 2014; Jastreboff & Jastreboff, 2014).

0.2.2 Barriers to development of hearing technology for decreased sound tolerance

A possible solution to over-avoidance of noise could be the use of hearing devices with variable attenuation (Traynor, 2021). An electronic hearing device containing microphones outside the ear and speakers inside the ear could achieve selective attenuation, by blocking out intolerable sounds and transmitting other sounds inside the ear. However, there are several challenges impeding the development of novel hearing technologies for decreased sound tolerance, including:

1. **Lack of tools for prototyping.** Prior research on managing decreased sound tolerance with hearing devices has mostly concentrated on proprietary, closed-source devices (Pfeiffer *et al.*, 2019a; Pfeiffer, Stein Duker, Murphy & Shui, 2019b; Valente, Goebel, Duddy, Sinks & Peterin, 2000). Therefore, the resources required to translate research to a usable prototype can be a barrier for individual researchers (Miller & Donahue, 2014), as incorporating new algorithms into a hearing device would require independently redesigning all the hardware and software necessary for high-quality audio processing.
2. **Limited data to inform device requirements.** There are multiple factors determining whether a hearing device would be usable in daily life, and these factors may need to be separately assessed in different contexts. For example, the latency requirements for hearing aid algorithms have been defined based on perceptual studies suggesting the threshold at which delaying the sound played by the hearing aid would be detectable or annoying (Goehring, Chapman, Bleeck & Monaghan, 2018; Stone & Moore, 2003; Dillon, 2012). It is unclear whether the requirements defined for hearing aids would extend to other hearing devices used by different populations.
3. **Individual and dynamic nature of decreased sound tolerance.** The presentation of decreased sound tolerance is highly heterogeneous: individuals can differ in the sets of sounds they are intolerant towards (Hansen, Leber & Saygin, 2021; Stiegler & Davis, 2010; Cavanna & Seri, 2015; Brout *et al.*, 2018), as well as their reactions to these sounds (Brout *et al.*, 2018; Cavanna & Seri, 2015). Furthermore, the experience of decreased sound

tolerance is not static within individuals; momentary mental states such as stress can affect how an individual reacts to sounds (Hasson, Theorell, Bergquist & Canlon, 2013). This means that to effectively reduce sound-induced distress, there needs to be a way to tailor devices to each individual's in-the-moment needs.

0.3 Objectives

The general objective of this work is to adapt a hearing device for decreased sound tolerance, with the following sub-objectives:

1. **Evaluate the suitability of existing tools for prototyping a hearing device for decreased sound tolerance.** Recently, research platforms for prototyping hearing aids and advanced hearing protection have been developed (Alexander, Clavier & Audette, 2018; Nadon, Bonnet, Bouserhal, Bernier & Voix, 2021). It is of interest to assess whether these existing tools meet the hardware and software requirements to implement the envisioned applications for managing decreased sound tolerance.
2. **Explore the effects of a hearing device's audio processing latency on language understanding.** Detectability and annoyance of delays may not be the only factors relevant to determining a hearing device's latency requirements. In certain usage contexts, such as continuous use in the classroom by students on the autism spectrum (Pfeiffer *et al.*, 2019b), the impact of a hearing device's delays on language understanding could affect learning outcomes.
3. **Assess the feasibility of stress monitoring with in-ear heartbeat sounds.** As stress plays a key role in decreased sound tolerance, one way to address its individual and dynamic nature could be to monitor an individual's stress to determine how to adapt the audio processing. Heartbeat signals have been widely used in automatic stress recognition (Giannakakis *et al.*, 2019a) and can be recorded with a hearing device from the occluded earcanal (Martin & Voix, 2018).

0.4 Structure of the thesis

The thesis presents the main advancements achieved in the effort to address the research objective. It is structured around three papers, included in full in chapters 2 to 4. In addition to these three chapters, there is a literature review (chapter 1) and a concluding chapter (chapter 5). The content of the different chapters is as follows:

0.4.1 Chapter 1 - Literature review

The literature review provides background on decreased sound tolerance and automatic stress recognition. The first section characterizes decreased sound tolerance and its subtypes, outlines how decreased sound tolerance is currently identified and measured, presents research on neurophysiological correlates, and lists the common interventions used to address decreased sound tolerance. The second section presents the state-of-the-art on automatic stress recognition, concentrating on systems trained on heartbeat signals. Finally, the literature review is summarized, with a focus on the novel contributions of the current project.

0.4.2 Chapter 2 - Interfacing the Tympan open-source hearing aid with an external computer for research on decreased sound tolerance

Chapter 2 is a conference proceedings paper published in September 2022 in *Proceedings of Meetings on Acoustics* titled “Interfacing the Tympan open-source hearing aid with an external computer for research on decreased sound tolerance”. This work aimed to address the first sub-objective of this thesis: to evaluate the suitability of existing tools for prototyping a hearing device for decreased sound tolerance. The limitations of two existing research platforms were characterized, and a processing pipeline was developed to overcome these limitations, allowing for a wide variety of algorithms with different latency and computational constraints to be tested for managing decreased sound tolerance.

0.4.3 Chapter 3 - Evaluating the effects of audiovisual delays on speech understanding with hearables: A pilot study

Chapter 3 is a journal article submitted in October 2022 to *Speech Communication* titled "Evaluating the effects of audiovisual delays on speech understanding with hearables: A pilot study". This work aimed to address the second sub-objective of this thesis: to explore the effects of a hearing device's audio processing latency on language understanding. Sentence-length audiovisual speech stimuli were modified to simulate processing delays and evaluated in a comprehension task with 50 participants, where it was found that the influence of the delay condition changed towards the end of the sentence, underlining the need for ecologically valid stimuli to fully characterize the effects that delays can have on comprehension of continuous speech.

0.4.4 Chapter 4 - Automatic stress recognition with in-ear heartbeat sounds

Chapter 4 is a journal article submitted in December 2022 to *IEEE Transactions on Affective Computing* titled "Automatic stress recognition with in-ear heartbeat sounds". This work aimed to address the third sub-objective of this thesis: to assess the feasibility of stress monitoring with in-ear heartbeat sounds. Stress recognition systems were developed and evaluated using simultaneous electrocardiography and in-ear audio signals recorded from 30 participants as they performed stressful tasks, with the final system based on in-ear audio achieving comparable performance to the systems trained on "gold-standard" electrocardiography signals.

0.4.5 Chapter 5 - Conclusions and Recommendations

Chapter 5 first presents the potential applications of the methods and tools developed in this thesis. Then, the limitations of the advancements made are discussed, along with considerations and recommendations for future research. Finally, the chapter lists the industrial and scientific contributions of this project.

CHAPTER 1

LITERATURE REVIEW

1.1 Decreased sound tolerance

1.1.1 Characteristics

Decreased sound tolerance is a term used to describe a cluster of conditions where an intolerance to sounds that do not bother the average listener is a common feature (Jastreboff & Jastreboff, 2014). Decreased sound tolerance conditions identified in recent literature include misophonia, characterized by decreased tolerance to specific sounds or stimuli associated with such sounds (Swedo *et al.*, 2021), phonophobia, characterized by unusual fear of specific sounds (Williams, He, Cascio & Woynaroski, 2021; Fackrell *et al.*, 2019; Jastreboff & Jastreboff, 2014), and hyperacusis, characterized by the perception of sounds as excessively loud or painful (Williams *et al.*, 2021; Goodson & Hull, 2015), although how exactly these individual conditions are defined varies among researchers (Williams *et al.*, 2021).

Decreased sound tolerance can occur as a standalone symptom (Fackrell *et al.*, 2019) but is also associated with other conditions that have additional symptoms; for example, it is thought that decreased sound tolerance has affected 50-70% of children and adults on the autism spectrum (Fackrell *et al.*, 2019; Williams *et al.*, 2021) and at least 90% of children and young adults with Williams' syndrome (Baguley & McFerran, 2011). Even within a single condition, the presentation of decreased sound tolerance can vary between individuals, e.g. people on the autism spectrum have reported distress associated with loud stimuli in general as well as specific quiet sounds (Williams *et al.*, 2021). Severity of reactions can also vary within a single individual depending on their mental state (Hasson *et al.*, 2013). Severe decreased sound tolerance can heavily impair quality of life: it can interfere with education and the ability to work, as well as relationships with family and friends (Fackrell *et al.*, 2019; Jager *et al.*, 2020; Cavanna & Seri, 2015; Brout *et al.*, 2018; Frank & McKay, 2019).

1.1.2 Assessment

There is currently no single standard measure for decreased sound tolerance, in part because its presentation is so diverse. Decreased sound tolerance has been assessed in interviews and case descriptions, with people experiencing decreased sound tolerance themselves (Edelstein, Brang, Rouw & Ramachandran, 2013; Brout *et al.*, 2018) or parents/teachers (Pfeiffer *et al.*, 2019a). A variety of questionnaires have been proposed for different subtypes of decreased sound tolerance (Talay-Ongan & Wood, 2000; Wu, Lewin, Murphy & Storch, 2014; Schröder, Vulink & Denys, 2013; Khalfa *et al.*, 2002).

Hyperacusis can also be assessed with loudness discomfort levels (LDLs), whereby the patient indicates the sound level at which sounds cause discomfort (Sheldrake, Diehl & Schaette, 2015; Sammeth *et al.*, 2000). However, as the ecological validity of the artificial sounds generally used in measuring LDLs has been challenged (Anari, Axelsson, Eliasson & Magnusson, 1999), psychoacoustic tests for misophonia (Enzler, Loriot, Fournier & Noreña, 2021b) and hyperacusis (Enzler, Fournier & Noreña, 2021a) using real-world sounds have been developed.

1.1.3 Neurophysiology

Numerous studies have investigated the neurophysiological responses associated with decreased sound tolerance (Neacsiu *et al.*, 2022; Goodson & Hull, 2015). While the exact neurophysiological mechanism is inconclusive and may differ depending on the subtype of decreased sound tolerance (Jastreboff & Jastreboff, 2015), a pattern that has emerged across most studies has been differences in biosignals related to autonomic nervous system activity, which can reflect psychological stress (Kemeny, 2003). Differences have been found in electrodermal activity (Edelstein *et al.*, 2013; Kumar *et al.*, 2017; Keith, Jamieson & Bennetto, 2019; Kuiper, Verhoeven & Geurts, 2019; Pfeiffer *et al.*, 2019b) and heart activity (Kumar *et al.*, 2017; Keith *et al.*, 2019; Aazh *et al.*; Schröder *et al.*, 2019; Fodstad, Kerswill, Kirsch, Lagges & Schmidt, 2021), both in comparison to control groups as well as within individuals, as a function of their response to sound stimuli.

The majority of these studies have been conducted in controlled experimental settings; however, wearable biosensors have offered the opportunity to assess autonomic reactions in real-world settings. Ambulatory monitoring may have more ecological validity, as context can play a role in how individuals respond to sound and controlled data collection can be stressful in itself (e.g. sitting still, scanner noise). Therefore, there have recently been efforts to collect data in naturalistic settings (Fodstad *et al.*, 2021; Pfeiffer *et al.*, 2019b); for example, in an ambulatory psychophysiological study conducted with students on the autism spectrum while they were wearing headphones (Pfeiffer *et al.*, 2019b). Although this study demonstrated the feasibility of ambulatory biosignal monitoring to assess headphone use, there were a couple of limitations to be noted: there was a high drop out rate among participants, likely due to the burden of physiological data collection, and no data collected on the use of features such as "aware" mode (i.e. "hear-through" or acoustical transparency), so the relationship between psychophysiological responses and how the headphones were used could not be investigated. Hearing devices with integrated stress monitoring could offer new opportunities for data collection in naturalistic settings by reducing the burden of data collection, as the device used for ambulatory biosignal monitoring already being used to manage decreased sound tolerance, as well as by simultaneously tracking usage of the device.

1.1.4 Interventions

Treatments for decreased sound tolerance include psychotherapies (Potgieter *et al.*, 2019; Aazh, Landgrebe, Danesh & Moore, 2019), medications (Potgieter *et al.*, 2019; Pienkowski *et al.*, 2014), biofeedback (Goodson & Hull, 2015; Claiborn, Dozier, Hart & Lee, 2020; Hoffman), and sound therapies (Pienkowski *et al.*, 2014; Noreña & Chery-Croze, 2007; Jastreboff & Jastreboff, 2014; Schröder, Vulink, van Loon & Denys, 2017; Frank & McKay, 2019). However, there is currently limited evidence on the effectiveness of these treatments (Fackrell *et al.*, 2019; Raj-Koziak *et al.*, 2021; Brout *et al.*, 2018; Smith *et al.*, 2022).

Lifestyle changes and accommodations can allow for people with decreased sound tolerance to avoid situations in which distressing sounds are present (Poore-Pariseau, 2019; Smith *et al.*, 2022). When complete avoidance of the situation is not possible, hearing protection and headphones are widely used to attenuate (Edelstein *et al.*, 2013; Neave-DiToro *et al.*, 2021; Sammeth *et al.*, 2000) or mask (Edelstein *et al.*, 2013) intolerable sounds. Research on hearing protection and headphones worn in the classroom has suggested that hearing devices can effectively reduce sound-induced distress and improve attention on tasks, particularly if these tasks do not require engagement with the surrounding acoustic environment, such as written tests (Smith & Riccomini, 2013; Rowe, Candler & Neville, 2011; Pfeiffer *et al.*, 2019b,a). However, regular noise avoidance may have negative consequences: sound isolation can hinder interactions with others and prevent full participation in everyday activities (Pfeiffer *et al.*, 2019a; Neave-DiToro *et al.*, 2021; Frank & McKay, 2019; Schaaf, Klofat & Hesse, 2003; Schröder *et al.*, 2017), and in the long-term, over-avoidance may exacerbate sensitivity to sound (Pienkowski *et al.*, 2014; Jastreboff & Jastreboff, 2014; Schaaf *et al.*, 2003).

To prevent the consequences of over-avoidance, interventions using hearing devices to expose the wearer to some noise have been proposed (Traynor, 2021; Jastreboff & Jastreboff, 2014; Sammeth *et al.*, 2000; Eddins, Formby & Armstrong, 2020). Such devices include earplugs with variable passive attenuation (Traynor, 2021), "hear-through" headphones that can toggle between acoustic isolation and actively transmitting external sounds inside the ear (Pfeiffer *et al.*, 2019a), and loudness suppression devices that pass through lower-level sounds but reduce the level of sounds above the wearer's loudness discomfort levels (Sammeth *et al.*, 2000; Eddins *et al.*, 2020). Hearing devices can also be used to support long-term desensitization strategies, by gradually reducing reliance on masking noise or attenuation (Jastreboff & Jastreboff, 2014; Traynor, 2021; Eddins *et al.*, 2020); for example, by slowly increasing the threshold at which external sounds are attenuated (Eddins *et al.*, 2020). However, these approaches require manual adjustment to ensure that the noise exposure remains tolerable for the patient (Jastreboff & Jastreboff, 2015; Eddins *et al.*, 2020).

1.2 Automatic stress recognition

With information on a wearer's stress level, a hearing device could potentially adapt to the idiosyncratic sensitivities of each individual without the need for manual adjustments. Advances in wearable biosensing and machine learning have offered the possibility to track an individual's stress level in real time by automatically processing biosensor data to detect stress-related changes in autonomic nervous system activity (Giannakakis *et al.*, 2019a).

In previous work, automatic stress recognition has been implemented with a wide variety of hardware configurations and machine learning methods (Giannakakis *et al.*, 2019a). Multimodal systems have been developed relying on numerous biosignal inputs e.g. electrodermal activity and heartbeat signals combined (Mishra *et al.*, 2018), while others are based on data from just one type of sensor e.g. heartbeat signals alone (Cho, Park, Dong & Youn, 2019). A wide range of machine learning techniques varying in complexity and interpretability have been applied to classify stress, ranging from end-to-end deep learning models directly trained on biosensor data to linear models trained on handcrafted features (Prajod & André, 2022).

Multiple features found to be related to psychological stress can be extracted from heartbeat signals, as both the sympathetic "fight or flight" and parasympathetic "rest and digest" responses of the autonomic nervous system affect heart activity at different time scales (Kim, Cheon, Bai, Lee & Koo, 2018). Given the precise number of milliseconds between the timestamps of successive heartbeats, numerous heart rate variability (HRV) indices can be computed. These indices can be divided into time-domain, frequency-domain, and nonlinear features depending on the signal processing technique used (Shaffer & Ginsberg, 2017).

Heartbeat sounds captured with a hearing device from inside the occluded earcanal have been previously used to measure the mean heart rate (Martin & Voix, 2018) and may also be suitable for computing HRV indices to train a stress recognition system. However, given that ECG recordings are less likely to be corrupted by artifacts and that the time between successive beats can be more precisely calculated from the difference between the peaks in the ECG signal, ECG is regarded as the gold standard heartbeat signal for HRV analysis (Laborde, Mosley & Thayer, 2017). As a result, a large portion of the research on automatic stress recognition up to this point has concentrated on HRV features extracted from ECG signals (Giannakakis *et al.*, 2019a). There is a need to assess whether a stress recognition system trained on HRV features extracted from in-ear audio signals can achieve similar performance as those trained on HRV features extracted from gold standard ECG signals.

1.3 Summary

Decreased sound tolerance is a diverse cluster of conditions with symptoms varying across different individuals. However, one pattern that has emerged from research on various types of decreased sound tolerance is differences in biosignals dependent on the autonomic nervous system, which can indicate increased psychological stress. Hearing devices capable of monitoring stress-related changes in biosignals could potentially improve existing hearing device-based interventions for decreased sound tolerance, which currently require manual changes in settings to adjust to the wearer's needs, such as the tolerable level of noise exposure. Heartbeat signals, among the biosignals that can be captured by a hearing device from inside the earcanal, have previously shown promise for automatic stress recognition. However, the processing applied to the more conventional heartbeat signals used in previously-validated stress recognition systems may not be applicable to in-ear heartbeat sounds, and it still needs to be determined whether automatic stress recognition would work in practice with audio signals recorded from inside the ear with a hearing device.

CHAPTER 2

INTERFACING THE TYMPAN OPEN-SOURCE HEARING AID WITH AN EXTERNAL COMPUTER FOR RESEARCH ON DECREASED SOUND TOLERANCE

Danielle Benesch[1] , Kocherla Nithin Raj[1] , Rachel Bouserhal[1] , Jérémie Voix[1]

[1] Department of Mechanical Engineering, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Article published in “Proceedings of Meetings on Acoustics” in September 2022.

2.1 Abstract

While headphones and hearing protectors are often used to relieve the distress associated with decreased sound tolerance, unnecessary attenuation of surrounding sounds may have negative consequences. With an advanced hearing device that can attenuate according to the wearer’s surroundings or mental state, it may be possible to reduce distress while still mitigating possible negative effects of over-attenuation. To investigate the utility of these features in practice, there is a need for a hearing device with which these features can be implemented. Two existing platforms, the Auditory Research Platform 3.1 and the Tympan Open-Source Hearing Aid Platform RevD, were evaluated for their suitability for decreased sound tolerance research. Due to the high latency of the Auditory Research Platform and the limited computational resources of the Tympan, an open-source software pipeline was built such that some of the processing, for which latency is not as crucial, could run on an external computer.

2.2 Introduction

To cope with decreased sound tolerance, intolerance to sounds that do not bother the average listener (Jastreboff & Jastreboff, 2014), hearing protection or noise-attenuating headphones are often worn (Cavanna & Seri, 2015; Neave-DiToro *et al.*, 2021; Jastreboff & Jastreboff, 2014; Frank & McKay, 2019; Sammeth *et al.*, 2000). However, conventional passive hearing protectors and noise-cancelling headphones attenuate all surrounding sounds and do not distinguish between sounds causing negative reactions and those conveying information. Conventional hearing protectors and headphones used to block out certain sounds may therefore attenuate too much sound for the wearer, which can prevent people with decreased sound tolerance from being fully engaged with their environment (Pfeiffer *et al.*, 2019a; Sammeth *et al.*, 2000; Neave-DiToro *et al.*, 2021). Furthermore, with time, over-avoidance of noise may exacerbate intolerance to sounds (Pienkowski *et al.*, 2014; Jastreboff & Jastreboff, 2014).

A possible solution to over-attenuation of surrounding sounds is a hearing device that can attenuate surrounding sounds only when necessary, by capturing external sounds, processing them, and transmitting them to the wearer selectively. Recent advances in sound event detection (Bhat, Shankar & Panahi, 2020; Wang *et al.*, 2022) and biosignal monitoring (Martin & Voix, 2018; Shu *et al.*, 2018; Giannakakis *et al.*, 2019a) have made adaptation to the acoustic environment or an individual's psychophysiological response increasingly realizable. However, a key barrier to developing novel technologies to manage decreased sound tolerance is the lack of tools to test these ideas in practice by integrating them into a hearing device. Previous studies have mainly focused on commercial hearing devices that are not readily modifiable by researchers (Pfeiffer *et al.*, 2019a,b; Valente *et al.*, 2000). The need to independently redevelop all the hardware and software required for high-quality audio processing often poses a challenge for individual research groups (Miller & Donahue, 2014).

Recognizing the potential of dedicated software and hardware to facilitate translation of academic research to real-world applications, several platforms have been recently developed for research on hearing aids and protection (Hansen *et al.*, 2019; Alexander *et al.*, 2018; Nadon *et al.*, 2021; Pisha *et al.*, 2019; Herzke, Kayser, Loshaj, Grimm & Hohmann). It is possible that these existing devices could be adapted for research on decreased sound tolerance, but some techniques that could be used to manage decreased sound tolerance, such as sound event and stress detection, could require more time and processing power than most algorithms presently used in hearing assistive technologies. Therefore, in this work, several hardware and software configurations were assessed for their suitability for decreased sound tolerance research.

2.3 Background

There are several ways that a hearing device could be used to manage decreased sound tolerance, beyond attenuation. One possible method to prevent over-attenuation is acoustical transparency, also known as "aware mode" or "transparency mode" in some commercially available headphones. These devices include a microphone to pick up external sounds and play them back inside the occluded earcanal, allowing for the wearer to hear the sounds around them without removing the device. In a study with students on the autism spectrum using headphones to manage decreased sound tolerance, headphones that constantly attenuated posed a barrier to participation in the classroom, however, a different model with acoustical transparency did not have this concern (Pfeiffer *et al.*, 2019a). Another technique that has been proposed for preventing over-attenuation is dynamic range limiting (Eddins *et al.*, 2020), which reduces the level of louder sounds without completely removing them from the played back signal. These techniques are based on processing the audio stream played back to the wearer.

Other techniques that may be helpful to people with decreased sound tolerance would aim to control the audio processing as a function of a wearer's physiology or environment. For example, an event activity detector of useful signals, such as speech (Lezzoum, Gagnon & Voix, 2013), emergency alarms (Bhat *et al.*, 2020), or school bells, could trigger the device to toggle between attenuation and acoustical transparency. Furthermore, because individuals with decreased sound tolerance may become distressed by different sets of sounds (Stiegler & Davis, 2010; Brout *et al.*, 2018) and momentary mental states may also influence how an individual responds to sounds (Hasson *et al.*, 2013), control of the input stream based on inter- and intra-individual differences would be favorable. More specifically, mitigation of the triggering external sounds as a function of the individual's stress level could offer an individualized strategy that can reduce the risk of over-attenuation. An advanced hearing device could conceivably adapt to the idiosyncratic sensitivities of individual wearers by incorporating biosignal processing to detect distress. Hearing devices that acoustically seal the ear create an occlusion effect, amplifying bone-conducted sounds within the earcanal (Berger & Voix, 2019). These amplified sounds can be recorded with an in-ear microphone and classified automatically (Chabot, Bouserhal, Cardinal & Voix, 2021). Heartbeat and respiratory signals, widely used in emotion recognition (Shu *et al.*, 2018; Giannakakis *et al.*, 2019a), are among the audio events that can be extracted from in-ear microphone data (Martin & Voix, 2018).

Processing power and latency requirements could be different for these separate classes of algorithms. Stress and audio event detection algorithms are often resource-heavy and require more time than audio processing algorithms intended to be used in "real time", that is, with sufficiently low latency. The latency of algorithms used in hearing devices is often kept below 10 ms to avoid undesirable perceptual effects (Alexander, 2016; Goehring *et al.*, 2018). For example, if the audio playback has a long delay, the synchronization of audio and visual information could be disrupted, in particular lip synchronization with speech (Summerfield, 1992). Therefore, a research platform for decreased sound tolerance intended to accommodate a variety of algorithms would need to be both computationally powerful and fast.

2.4 Approaches

Three hardware configurations were considered for the purpose of real-time implementation of the envisioned algorithms for managing decreased sound tolerance: the Auditory Research Platform (ARP) hardware, the Tympan Hardware, and the Tympan connected to an external PC. These hardware configurations are presented respectively in sections 2.4.1, section 2.4.2 and 2.4.3, together with the relevant tests that were conducted to assess their suitability for real-time use.

2.4.1 ARP 3.1 hardware connected to a PC

The Auditory Research Platform (ARP) is a portable device for hearing research developed developed over the years within the NSERC-EERS Industrial Research Chair in In-Ear Technologies (CRITIAS), currently in its third generation. The ARP 3.1 earpieces, illustrated in Fig. 2.1, each feature an outer-ear microphone, an in-ear microphone, as well as an internal miniature loudspeaker. The ARP 3.1 includes a processing unit that can emulate a sound card, with which audio signals captured by the earpieces' microphones can be recorded on a PC. In addition to recording data from the ARP's earpieces, it was of interest whether a PC could be used to process and transmit audio back to the earpieces for real-time applications. This setup would enable any algorithm that can run on the PC to be used for audio processing, even if it is implemented in a high-level language and requires substantial computational power (e.g. a machine learning algorithm implemented in MATLAB). However, the PC would need to process and transmit the audio at an acceptably low latency in order to avoid perceptual disturbances. This aspect is key for the envisioned application, and the relevant tests are presented in the following section.

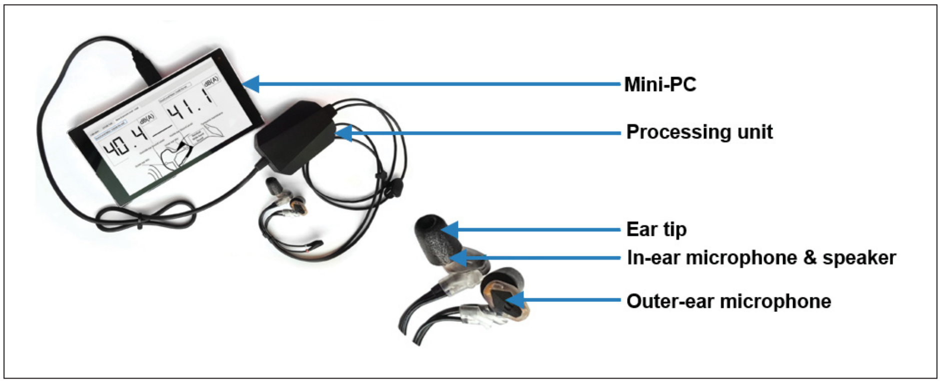


Figure 2.1 The Auditory Research Platform (ARP) 3.1. The ARP earpieces each feature an outer-ear microphone, a high-attenuation ear tip, and an in-ear speaker while the ARP software framework is hosted on an external personal computer (PC)

2.4.1.1 Latency test setup

To test the minimum possible latency of audio processing on the PC, the ARP processing unit was used to send audio directly to and from the PC without applying any processing algorithms (i.e. audio pass-thru). Four channels of audio, corresponding to the two outer-ear microphones and the two in-ear microphones, were transmitted via USB to a PC running Windows (Dell Latitude 5500 with 32 GB RAM & Intel(R) Core i7-8665U CPU). ASIO4ALL (2.14) [1], a hardware-independent driver, was used to directly interface the processing unit as a soundcard

[1] https://www.asio4all.org/

with MATLAB 2019b, bypassing the standard Windows audio path and enabling access to the multiple audio inputs and outputs. The Audio Device Reader and Audio Device Writer blocks available from the MATLAB Audio Toolbox read and wrote continuous streams of audio data. All four channels were read by the Audio Device Reader, and the channels corresponding to the two outer-ear microphones were played back on the in-ear speakers via the Audio Device Writer. The sampling rate was set at 44,100 Hz.

Multiple configurations were tested: for one set of measurements, all four channels recorded by the ARP were continuously written to PCM audio files and a simple MATLAB Graphical User Interface (GUI) controlled the transmission of audio to the internal loudspeakers with a stop button. The latency of pass-thru was also assessed without saving any audio from the ARP onto the PC, as well as without the GUI. For each of the aforementioned cases, the block size of the data acquisition was varied from 64 to 2048.

To estimate the latency of each of these configurations, an earpiece was placed on one ear of a head and torso simulator (HATS) model 4128C (Brüel & Kjær, Nærum, Denmark) in an audiometric booth. As depicted in Fig. 2.2, the HATS' microphones were connected to a NiDAQ data acquisition system model PXI 1033 (National Instruments, Austin, TX, USA), with which a separate PC outside of the audiometric booth recorded via MATLAB at a sampling rate of 48,000 Hz. Using the same recording PC, a pulse signal with a 5 ms duration, shown in Fig. 2.3, was played on four surrounding loudspeakers within the booth.

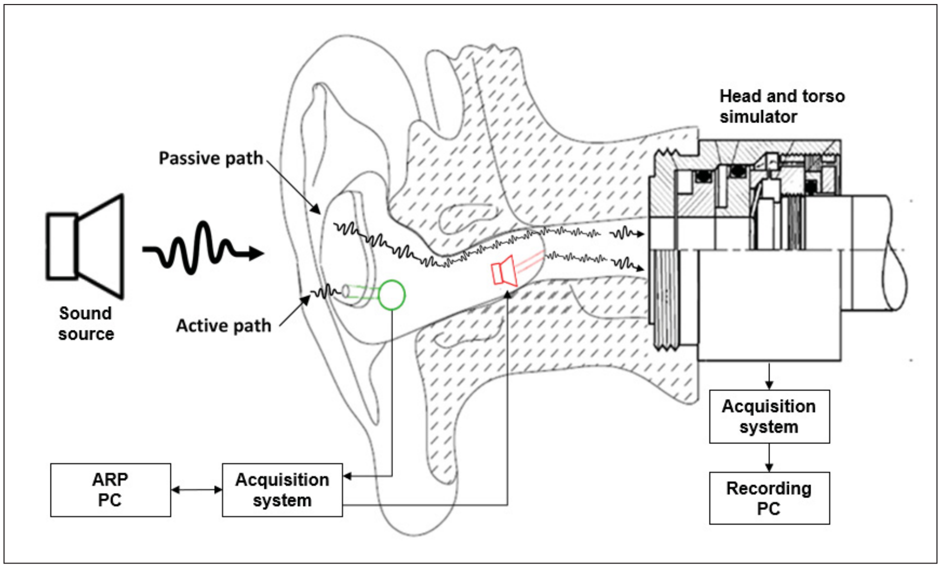


Figure 2.2 Diagram of the setup used to estimate the latency of audio pass-thru with the Auditory Research Platform (ARP) 3.1 and an external computer (PC): Audio is continuously sent from the ARP earpiece to the PC and back to the earpiece. A separate PC records the microphone signal of the head and torso (HATS) simulator. When an external sound is played, it is transmitted through the eartip (passive path), as well as through the electroacoustic system composed of the external microphone, soundcard and internal miniature speaker (active path)

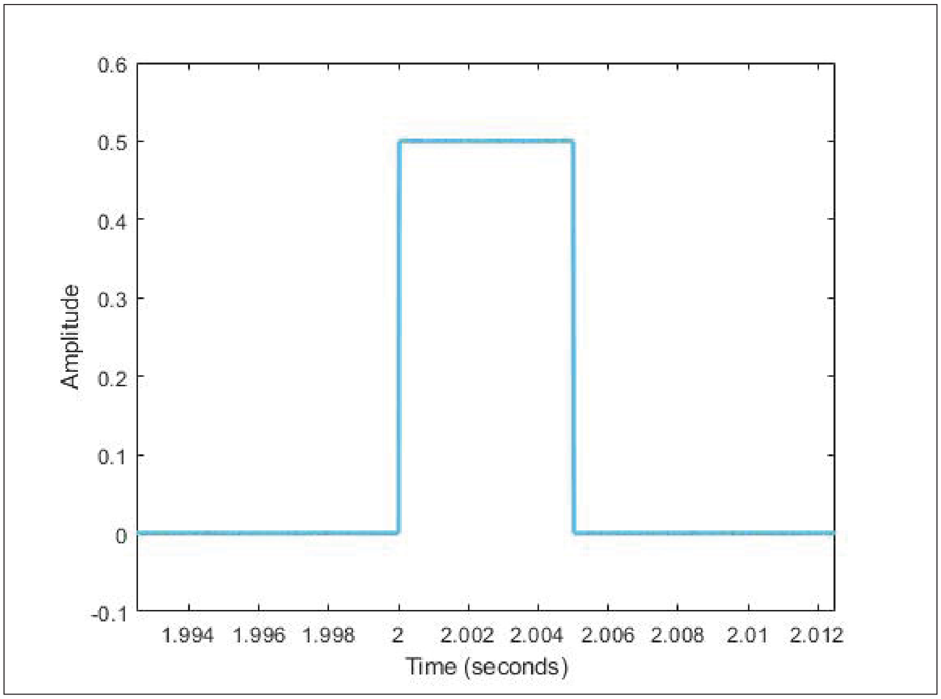


Figure 2.3 Square pulse edge signal of 5 ms played on audiometric booth speakers to measure audio pass-thru latency

The stimulus signal played on the audiometric booth loudspeakers took two paths through the earpiece, depicted in Fig. 2.2: the signal is passively transmitted through the eartip material, and it was also recorded by the earpiece's outer-ear microphone and actively transmitted to the earpiece's internal loudspeaker. The duration of the stimulus signal was set to be below the threshold of minimal noticeable latency so that if the latency of audio pass-thru was larger than this threshold, two distinct signals (corresponding to the passive and active paths) would be recorded by the HATS' microphone. The pass-thru latency was calculated by subtracting the timestamp of the actively transmitted signal from the timestamp of the passively transmitted signal, as visible from the time waveform, such as the one shown in Fig. 2.4.

To determine the interval between the two signals, the audio recorded by the HATS' microphone was plotted in MATLAB, and the timestamps corresponding to the passively and actively transmitted signals were manually selected from the time waveform, as the one shown in Fig. 2.4. Each hardware configuration was tested in a single recording in which the stimulus signal was played repeatedly. Five measurements were taken per recording, corresponding to 90 total measurements for the 18 configurations tested.

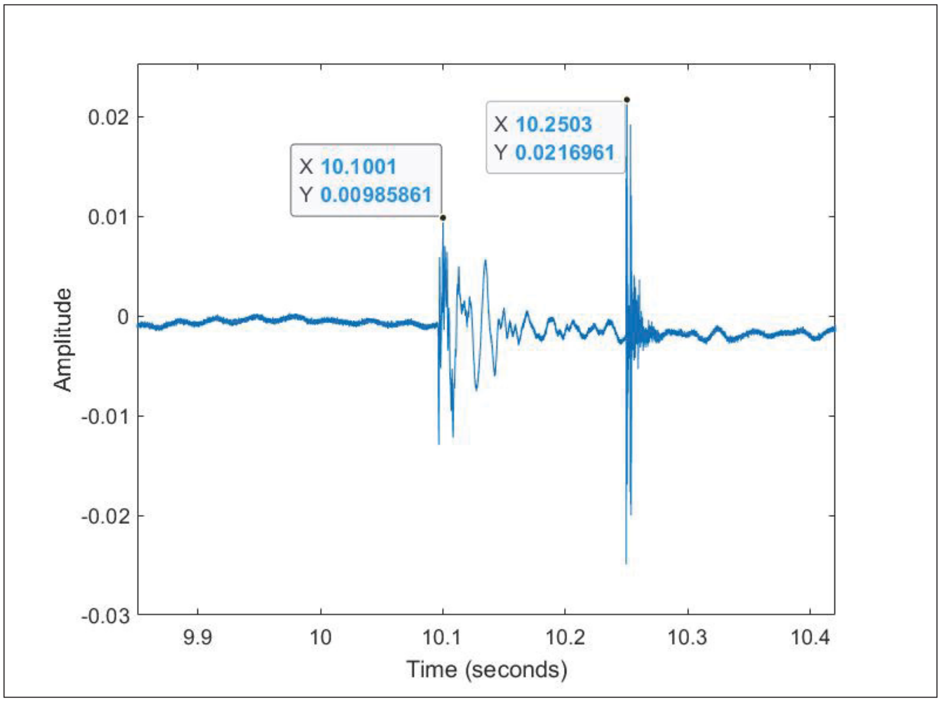


Figure 2.4 Example signal recorded by HATS' microphone used to estimate the latency of audio pass-thru with a block size of 2048, while saving 4 channels of audio data, and running a GUI. The first marker at 10.10 seconds corresponds to the passively transmitted pulse sound, while the second marker at 10.25 seconds corresponds to the pulse sound actively transmitted through the in-ear speaker, indicating a latency of 0.15 s for this pass-thru configuration

2.4.1.2 Latency test results

The mean latency for audio pass-thru via the PC ranged from 22 ms to 150 ms, depending on the configuration. The configuration with the highest mean latency was a block size of 2048 while running a simple GUI with a stop button and writing all four recorded channels to audio files. The configuration with the lowest mean latency was a block size of 256 without running the GUI and without writing any audio files. The mean and standard deviation of the latency measurements taken for each configuration are summarized in Table 2.1. These measured latencies all exceed the maximum value generally considered acceptable for real-time audio processing (Alexander *et al.*, 2018), rendering this hardware configuration unsuitable for the envisioned real-time application. A processing approach with lower latency is needed and a possible solution, the Tympan hardware, is evaluated in the next section.

Table 2.1 Estimates of the latency of audio pass-thru with varying features and block sizes. Values are displayed as mean ± standard deviation (SD) and expressed in milliseconds (ms)

Block size	**With no additional features**	**With only GUI**	**With GUI + saving data**
64	29.42 ± 0.04	47.52 ± 2.14	45.72 ± 0.02
128	29.41 ± 0.04	57.78 ± 0.01	62.05 ± 0.02
256	22.28 ± 0.03	53.59 ± 0.02	62.27 ± 0.03
512	45.9 ± 0.02	53.76 ± 0.01	51.6 ± 0.03
1024	76.91 ± 0.03	76.84 ± 0.02	78.45 ± 0.02
2048	146.81 ± 0.04	147.39 ± 1.75	150.14 ± 0.03

2.4.2 Standalone Tympan hardware

The Tympan is an open-source research platform intended for hearing aid prototyping (Alexander *et al.*, 2018). Shown in Fig. 2.5, the Tympan hardware includes microphone and headphone jacks and a wearable processing unit, based on the Teensy™microcontroller development board and Arduino™software platform. It is designed for real-time audio processing, with a latency considered acceptable for a hearing aid (5.7 ms in one test (Alexander *et al.*, 2018)). The Tympan's software library contains various extensible examples that can be run on the processing unit including audio playback/recording with a Secure Digital (SD) memory card, sound level measurement, equalization, noise reduction, and wide dynamic range compression (WDRC). While the Tympan was originally intended for hearing aid research, it may be suited for decreased

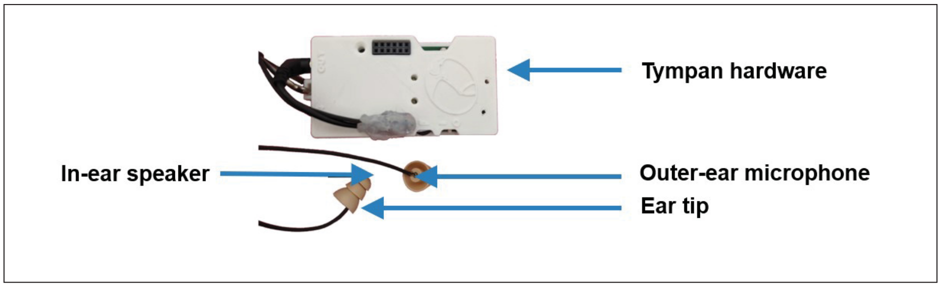


Figure 2.5 The Tympan Hearing Aid Platform Rev D pictured with earpieces containing an outer-ear microphone and in-ear speaker

sound tolerance research as well. For example, while WDRC is generally used in hearing aids to amplify softer sounds and attenuate louder sounds, the parameters could be changed such that it functions as a dynamic range limiter, attenuating louder sounds while preserving the level of softer sounds. Furthermore, since it is based on the Teensy/Arduino development system, it can also be extended with algorithms beyond those available in the Tympan library. However, high-resource algorithms, such as many machine-learning based classifiers, may need to be adapted to run on microcontrollers (Tsutsui da Silva, Souza & Batista, 2022; Wang *et al.*, 2022). Therefore, it was of interest how much of the Tympan's computational resources are used by the core processor unit (CPU) to run the WDRC algorithm. The CPU usage indicates how much additional processing power would be available for other features.

2.4.2.1 CPU test setup

The CPU usage was tested on the "RevD" version of the Tympan hardware, containing a Teensy 3.6 Processor (NXP 180MHz ARM 32 Processor) with codec model TLC320AIC3206 (Texas Instruments, Dallas, TX, USA). A WDRC algorithm available from the Tympan library[2] was used, including adaptive feedback cancellation on eight frequency bands. The processing was monaural, i.e. a single channel of audio was played into both in-ear speakers of the earpieces connected to the Tympan. The sampling rate and block size, known to affect audio quality and

[2] https://github.com/Tympan/Tympan_Library/blob/v1.6.0/examples/05-FullSystems/WDRC_8BandIIR_wAFC/WDRC_8BandIIR_wAFC.ino

latency (Balling, Mosgaard & Helmink, 2022; Wagenknecht, 2018; Dillon, 2012), were varied respectively from 16,000 Hz to 32,000 Hz, and from 16 to 128 samples. The CPU usage for each configuration was then printed in the Serial Monitor in the Arduino IDE.

2.4.2.2 CPU test results

The CPU usage for each configuration tested is shown in Table 2.2. CPU usage ranged from 49% to 98%. At the highest sampling frequency tested, 32,000 Hz, it was not possible to run block sizes of 64 or less, so the CPU usage is not shown for these configurations. Running all processing on the Tympan would limit possible applications that require additional features on top of conventional hearing aid algorithms. Therefore, an approach that would keep the audio processing low latency but offer more computation power is preferable. To address this limitation, the Tympan hardware connected to a PC may be a viable solution and is evaluated in the next section.

Table 2.2 CPU Usage of wide dynamic range compression on Tympan Rev D with varying parameters

Sampling frequency (Hz)	**Block size**	**CPU Usage (%)**
16,000	16	60.7
	32	54.2
	64	50.9
	128	49.1
24,000	16	90.9
	32	81.4
	64	76.5
	128	73.9
32,000	128	98.5

2.4.3 Tympan hardware connected to a PC

The third approach was to interface the Tympan with a PC, such that the Tympan could be used for real-time audio processing and the PC could be used for resource-intensive applications for which latency is not as crucial. As with the standalone Tympan hardware, the Tympan acquired audio for real-time audio processing, but it was additionally exposed to an external PC as a USB soundcard, simultaneously sending audio to the PC. The PC could also send serial commands back to the Tympan, in order to change the parameters of the audio processing blocks without adding extra overhead.

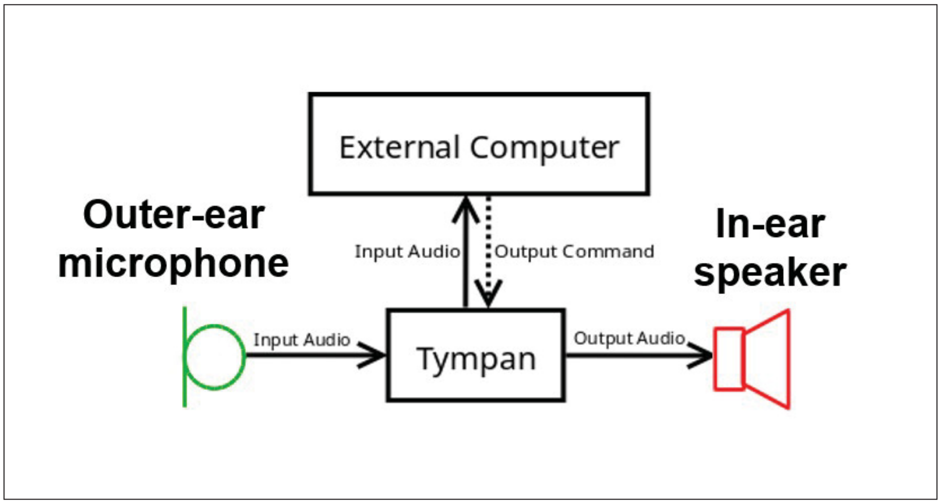


Figure 2.6 Diagram of the proposed pipeline interfacing the Tympan Hearing Aid Platform and an external computer. Input audio is sent to the Tympan, processed, and simultaneously sent to the external computer over serial communication

2.4.3.1 Proof of concept of processing pipeline

As a proof of concept to demonstrate the functionality of the processing pipeline connecting the Tympan to a PC, a clap detection application was developed. As illustrated in Fig. 2.7, audio acquired from the microphone could be passed through directly to the earpiece's loudspeakers or optionally mixed with white noise first. A clap detection algorithm running on the PC was used to control the Tympan's audio processing blocks such that the detection of clap sounds in the audio sent to the PC could turn the white noise playback on and off.

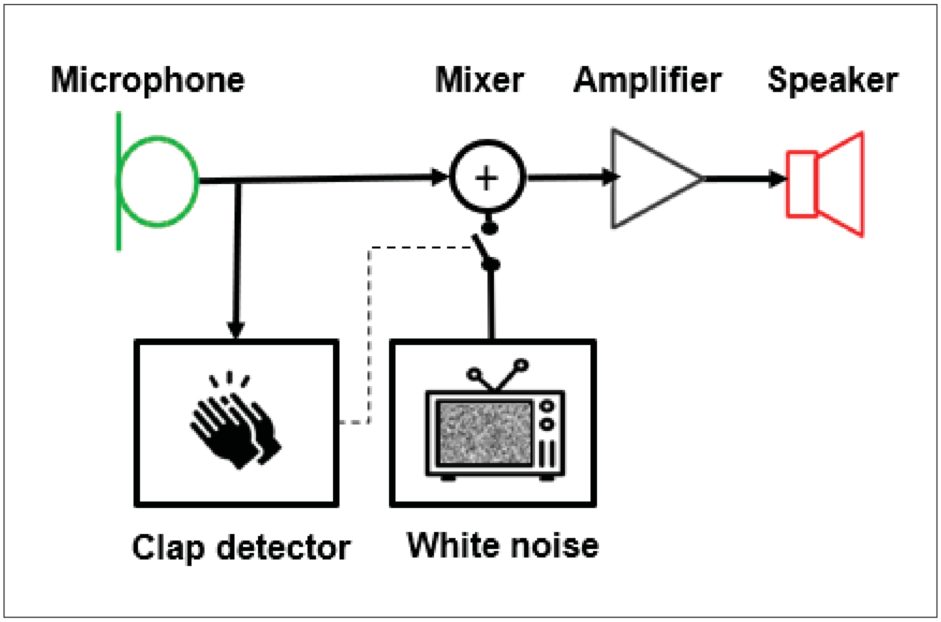


Figure 2.7 Diagram of the example application. Audio is sent from the microphone connected to the Tympan to the PC, where a clap detection algorithm is running. If two claps are detected, the white noise player on the Tympan is toggled and white noise is played on the in-ear speaker

In this example application, the Tympan acquired audio input at 44,117 Hz[3] using the I2S protocol, which was passed through to the in-ear speakers of the earpieces connected to the Tympan. This audio was also sent to the external PC via serial communication, with the Tympan exposed to the PC as a USB soundcard, and was captured by a Python script using PyAudio to set up the audio stream. The audio stream was processed on a frame-by-frame basis using the pi-clap library [4], which detects claps using a basic thresholding algorithm. When two claps were detected, a command (in this case, the letter “m”) was sent back to the Tympan via serial communication, with a baud rate of 115200. This serial command was used by the Tympan to toggle the white noise mixer, controlling whether white noise is sent to the in-ear speakers.

[3] The sampling rate was set to 44,117 Hz for compatibility with the Teensy Library. The latest version of the Tympan hardware, Rev E, is less limited than the Rev D in audio processing capabilities for higher sampling rates.

[4] https://github.com/nikhiljohn10/pi-clap/releases/tag/1.4.2

The code for this example has been made publicly available in a GitHub repository. The general processing pipeline interfacing the Tympan with the PC is intended to be algorithm-agnostic, that is, the white noise mixer may be replaced by any combination of audio processing blocks on the Tympan and the clap detection may be replaced by any algorithm taking signals from the Tympan as input and optionally commands as output (Benesch, Raj, Bouserhal & Voix, 2022b). However, it should be noted that the algorithm running on the PC ran at a latency substantially longer than the acceptable latency for real-time audio processing: as shown in Fig. 2.8, the delay between the claps and the playback of white noise was approximately 800 ms. Therefore, the applications running on the PC should be limited to those where low latency is not required.

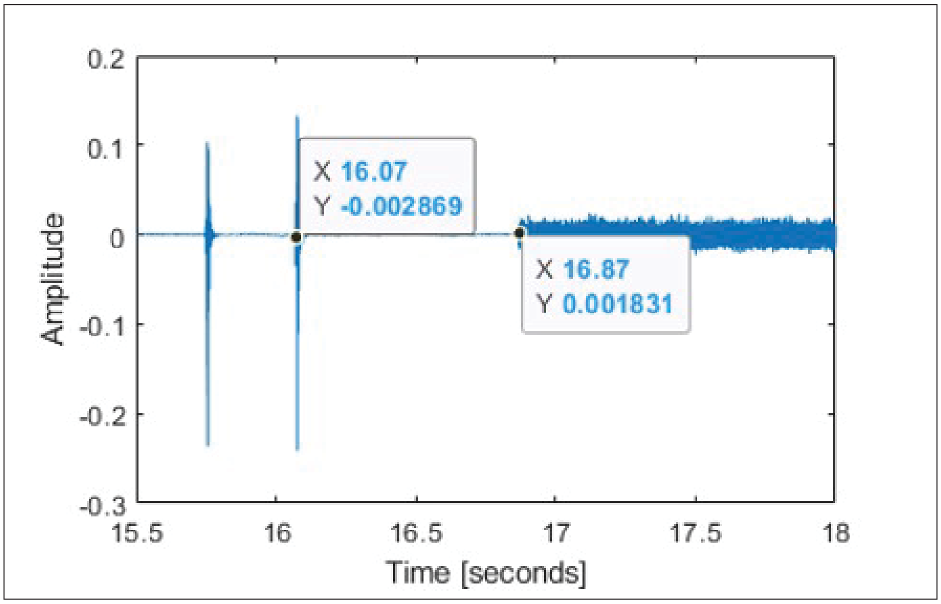


Figure 2.8 Signal recorded by the head and torso simulator demonstrating the white noise toggle with clap detection. The first marker at 16.07 seconds corresponds to the second clap sound, while the second marker at 16.87 seconds corresponds to the onset of the white noise, indicating a delay of 0.80 s to toggle the white noise player

2.5 Discussion

The aim of this project was to find a suitable hardware and software platform for decreased sound tolerance research. Two existing platforms were considered: the Auditory Research Platform 3.1, where all signal processing runs on an external PC, and the Tympan Open Hearing Aid Platform Rev D, where all processing runs on a microcontroller. A key limitation of running all processing solely on the PC was latency: the shortest possible delay of audio pass-thru, before any audio processing was applied and without any additional features included, was 22 ms, already well beyond the 10 ms limit usually imposed on hearing aids to prevent perceptual disturbances (Alexander, 2016). On the other hand, the processing power of the Tympan Rev D was limited even for a conventional hearing aid algorithm without additional features. To circumvent this limitation, the Tympan was interfaced with an external PC such that the Tympan could be used for real-time audio processing and the PC could be used for more resource-intensive processing. Additionally, the PC could make prototyping certain algorithms more accessible by allowing the use of higher-level languages such as Python, MATLAB, and Julia, and their respective libraries, the advantages of which have already been highlighted in the context of hearing aid development (Villescas, de Vries, Stuijk & Corporaal, 2020). However, for certain research applications such as in-situ data collection, a more portable solution such as the “standalone” Tympan would likely be preferable (Lentz, Yun & Smiley, 2021). On that note, it is encouraging that the latest version of the Tympan, the Rev E, has greatly improved processing compared to the Rev D: the underlying Teensy 4.1 has a CoreMark value (metric of CPU performance) of 2314, compared to the Rev D Teensy 3.6’s CoreMark value of 441. As microcontrollers are gaining capabilities over time, the processing power will likely become less of a barrier.

Despite the current limitations of existing research platforms, open-source hardware and software offers a valuable starting point for decreased sound tolerance researchers. In addition, the evaluation approach described in this work could be reused for the benchmarking of future hardware platforms, and the software pipeline that was developed enables the real-time use of a PC, thereby increasing the possibilities for rapid prototyping and research applications.

2.6 Acknowledgments

The authors would like to thank Corentin Delain for the recordings on ARP 3.1, as well as all the team from Tympan for their technical support and generous donation of a Rev. E unit.

The financial support received from the Natural Sciences and Engineering Research Council of Canada (NSERC) and EERS Global Technologies Inc, through the NSERC-EERS Industrial Research Chair in In-Ear Technologies (CRITIAS) is gratefully acknowledged. The first author acknowledges the financial support received from the German Academic Exchange Service (DAAD), and the first two authors acknowledge the financial support received from MITACS Globalink programs.

CHAPTER 3

EVALUATING THE EFFECTS OF AUDIOVISUAL DELAYS ON SPEECH UNDERSTANDING WITH HEARABLES: A PILOT STUDY

Danielle Benesch[1] , Juliane Schwab[2] , Jérémie Voix[1] , Rachel Bouserhal[1]

[1] NSERC-EERS Industrial Research Chair in In-Ear Technologies, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

[2] Department of Linguistics, University of Tübingen,
Keplerstraße 2, Tübingen, Baden-Württemberg, Germany 72074

Article submitted to "Speech Communication" in October 2022.

3.1 Abstract

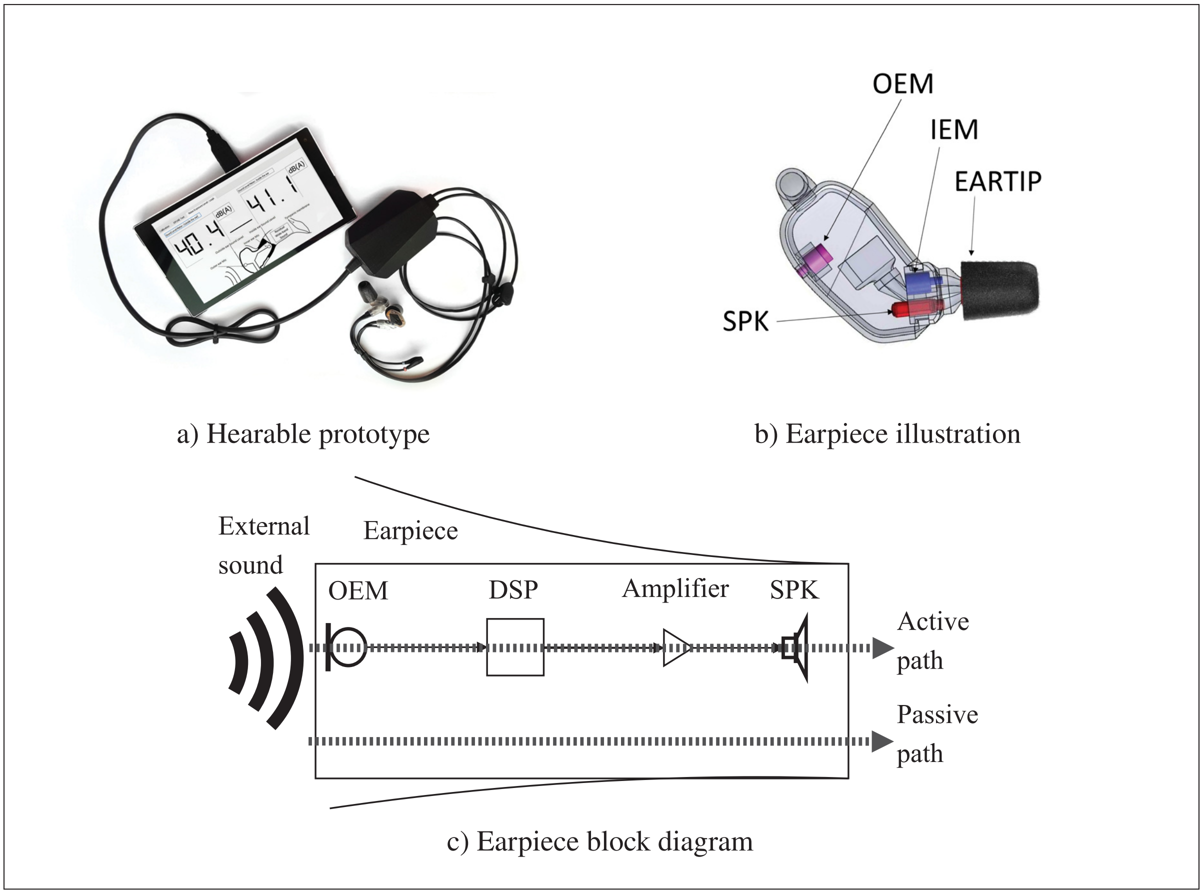
A key consideration in the development of real-time speech processing algorithms is latency between when speech is first captured and when the processed speech is transmitted. Much of the previous work informing the acceptability of processing latency has focused on the thresholds at which delays are detectable or annoying. This study aimed to explore the effects of delays on responding to a task necessitating speech understanding. Sentence-length audiovisual speech stimuli were modified such that the audio stream was delayed relative to the visual stream as well as to an attenuated copy of the audio stream by 10 and 400 ms, respectively. These stimuli were presented in conjunction with a pictorial referent-choice task to 50 normal-hearing participants. Participants responded slower to the task when the audio stream was delayed by 400 ms relative to the video stream. With an additional attenuated copy of the audio stream synchronized with the video stream, reaction time decreased, but this effect was limited to when the target word appeared near the onset of the sentence, highlighting the complexity of delay perception in the context of continuous speech. Implications for determining permissible processing delays for digital electronic hearing devices are discussed.

3.2 Introduction

Hearables, wearable electronic and electroacoustic devices worn in or around the ear, have the potential to improve communication in a wide variety of contexts (Voix, 2014), from workers in noisy industrial environments (Bou Serhal, Falk & Voix, 2013) to people with (central) auditory processing disorders (Moshgelani, Parsa, Allan, Veeranna & Allen, 2020) or decreased sound tolerance conditions such as misophonia and hyperacusis (Pfeiffer *et al.*, 2019a; Lezzoum *et al.*, 2013). These applications require transmission of speech in "real time", that is, with sufficiently low latency. However, the time constraints that define real-time processing are determined by the target application and can involve trade-offs between latency and the performance of audio processing algorithms such as speech enhancement or denoising.

The time it takes to capture audio signals, process them, and transmit them with a hearable can be noticeable and disturbing to the wearer (Dillon, 2012; Lezzoum, Gagnon & Voix, 2016; Goehring *et al.*, 2018; Stone & Moore, 2003). When a speech signal actively transmitted by a hearable is substantially delayed, it may also negatively impact speech understanding. In this work, two types of delays are investigated. The first stems from the asynchrony between the audio signal and the associated visual information, such as lip movements. The second type of delay occurs when external sound is not fully attenuated by the hearable. Even with an earpiece that completely occludes the ear canal (see e.g. Fig. 3.1a and 3.1b), the attenuation provided by the earpiece is limited, and therefore, if the external sound level is high enough, some residual audio will still be audible. When audio is actively transmitted following a processing delay (as shown in Fig. 3.1c), there will be an asynchrony between the residual audio signal inside the ear canal, after the passive attenuation of the earpiece (heard first), and the louder actively transmitted signal (heard second).

These two delays have been previously investigated in hearing aid research to determine recommendations for the maximum acceptable latency (Dillon, 2012). However, given the large range of potential users of hearables, there is a need to reassess the acceptability of delays in different contexts and for specific target populations. Previous research has shown that the



a) Hearable prototype

b) Earpiece illustration

c) Earpiece block diagram

Figure 3.1 Overview of a hearable. (a) depicts the Auditory Research Platform version 3.1, a hearable prototype developed at the NSERC-EERS Industrial Research Chair in In-Ear Technologies (CRITIAS), composed of 2 wired earpieces, one digital signal processor (DSP), and one mini-computer; (b) shows a detailed view of one earpiece, including an outer-ear microphone (OEM), an in-ear microphone (IEM), and an internal loudspeaker (SPK); (c) illustrates the active path of the delayed audio and the passive path of the residual audio. External sound is picked up by the OEMs and can be processed by the mini-computer before being played back on the SPK

perception and tolerance of delays can depend on a range of factors. For audiovisual delays, these factors include the type of audio signal, e.g. speech vs. nonlinguistic stimuli (Noel, De Niear, Stevenson, Alais & Wallace, 2017; Bebko, Weiss, Demark & Gomez, 2006; Dixon & Spitz, 1980), the presence of background noise (Pandey, Kunov & Abel, 1986), native language (Navarra, Alsius, Velasco, Soto-Faraco & Spence, 2010), age (Chan, Pianta & McKendrick, 2014; Gordon-

Salant, Schwartz, Oppler & Yeni-Komshian, 2022) and neurodevelopmental conditions such as autism (Noel *et al.*, 2017; Bebko *et al.*, 2006; de Boer-Schellekens, Eussen & Vroomen, 2013).

Likewise, the perceptual effects of an asynchrony between actively transmitted and residual audio can vary. Given a shorter delay, the actively transmitted and residual audio may be perceived together as a single, unnatural-sounding signal due to the superposition of these two signals resulting in spectral ripples, referred to as comb-filtering effects (Denk, Schepker, Doclo & Kollmeier, 2020; Goehring *et al.*, 2018). Longer delays can result in the perception of two distinct signals, with the delayed signal being perceived as an "echo" of the passively transmitted signal (Lezzoum *et al.*, 2016). The delay thresholds at which these perceptual effects are experienced can depend on a number of factors, including hearing impairment, the type of audio signal, the presence of background noise, and the attenuation provided by the earpiece (Lezzoum *et al.*, 2016; Goehring *et al.*, 2018).

Furthermore, there are multiple measures that could be considered when assessing the acceptability of a delay. One common measure is the threshold at which a delay between two signals is considered to be perceptible, generally derived from explicit judgments of either the signals' simultaneity (e.g. Noel *et al.*, 2017) or temporal order (e.g. de Boer-Schellekens *et al.*, 2013), that is, which of the signals is perceived first. Another measure is the threshold at which a delayed signal is rated as disturbing (Goehring *et al.*, 2018; Stone & Moore, 2003). The subjective perception of delays can also be measured more implicitly with behavioral measures such as whether participants gaze more towards a screen playing asychronous or synchronous stimuli (Bebko *et al.*, 2006). However, perceptibility and preference are distinct from the effect a delay has on communication, which may also be relevant when considering the real-time requirements of audio processing algorithms. Measures that can be used to directly assess the impact of delays on communication include the rate of unintentional interruptions by conversation partners (Egger, Schatz & Scherer, 2010), as well as recognition of syllables (Stone & Moore, 2003; Grant & Seitz, 1998) or words in a sentence (Arai & Greenberg, 1998; Pandey *et al.*, 1986; Gordon-Salant *et al.*, 2022; Grant & Seitz, 1998). These measures can lead to distinct conclusions; for example, Gordon-Salant *et al.* (2022) found that older listeners were less likely than younger listeners to detect audiovisual delays in a simultaneity judgment task, while in a speech recognition task, both populations' performance still decreased similarly with delays.

In some circumstances, the extent to which delays affect speech understanding could be relevant independent of whether they are perceptible: On the one hand, a perceptible delay may be deemed acceptable as long as latency does not impede the wearer's ability to appropriately respond to their surroundings. For example, an industrial worker who would otherwise be exposed to dangerous sound levels could find a perceptible delay acceptable if they can still correctly respond to their colleague's instructions. Similarly, a person with decreased sound tolerance may prefer to hear denoised speech with a perceptible delay instead of the sounds they are intolerant towards. On the other hand, if even imperceptible delays can negatively impact the efficiency of speech processing, this may be deemed unacceptable for some populations. For example, if a person with a (central) auditory processing disorder wears a hearable to improve speech understanding, it is crucial that the underlying speech enhancement algorithm does not introduce a delay that degrades speech understanding to the point that it outweighs the benefits of the speech enhancement.

Audiovisual speech understanding involves multiple (mutually informative) stages, from bottom-up sensory processing to top-down influences from the linguistic system (Grant & Bernstein, 2019). As such, the operationalization of audiovisual integration in experimental tests depends, in part, on which aspects of the language stimulus are deemed most relevant: In a comparison of various tests used to assess audiovisual integration, Grant & Seitz (1998) found that while syllable-based measures were correlated with each other, they were not correlated with more complex sentence-based measures. The present investigation aimed to assess higher-level processes involved in the comprehension of continuous speech, and therefore, a task requiring recognition of words within sentence-length stimuli was included. Furthermore, as the quality of the audio signal can interact with the benefit that visual cues provide in understanding speech (Grant & Bernstein, 2019), in this study, the residual audio that would be passively transmitted through a hearable was simulated in addition to an audiovisual delay.

The overall focus of this pilot study was to examine how measures of asynchrony detection and speech understanding were affected by the delays expected when wearing a hearable. The influence of task-specific factors, such as the syntactic complexity of the speech stimuli, was also explored. By investigating the processing of delayed speech in a sample drawn from the general population, the results may inform future efforts to determine the permissible delay of a hearable for a wide variety of applications and user groups.

3.3 Methods

Sentence-length audiovisual speech signals were modified such that two types of delays were produced and controlled. The first type of delay was an audiovisual asynchrony: the auditory speech signal was delayed relative to the visual speech signal (i.e. the visual stream of the video showing the speaker's face and lip movements during speech production). The second type of delay corresponds to the auditory speech signal lagging a residual audio signal, which was created with an attenuated copy of the original auditory signal and intended to simulate the attenuation provided by a real earpiece. The impact of these two types of delays was assessed both in terms of perceptibility and speech understanding.

Speech understanding was assessed in a pictorial referent choice task, in which participants were given an adjective and instructed to select an image representing the noun modified by this adjective in the audiovisual speech stimulus. This task was intended to index speech processing at a level of complexity higher than the perceptual processing involved in simultaneity judgment tasks or syllable-based recognition tasks. The task required recognition of keywords within the sentence stimulus, similarly to tasks previously used to measure intelligibility (e.g. Grant & Seitz, 1998). As the effects of sensory-level manipulations on speech understanding can be more pronounced with additional strain on cognitive resources such as with linguistically more complex stimuli (Carroll & Ruigendijk, 2013; Grant & Bernstein, 2019), the effects of delays may interact with task demands. Therefore, the stimuli used in the referent choice task were controlled for their sentence structure as well as the position of the target word within the sentence, both of which have been previously shown to interact with the addition of noise (Carroll & Ruigendijk, 2013). Furthermore, rather than instructing participants to respond verbally or in written form, a pictorial task was used such that the target words needed to be semantically mapped to the image representation (Schmithorst, Holland & Plante, 2007). The forced-choice format also allowed for straightforward measurement of reaction time and accuracy as indicators of processing difficulty (Uslar *et al.*, 2013), as well as an assessment of speech comprehension independent of production, thus being more inclusive towards potential target populations with limited expressive language abilities (Plesa Skwerer, Jordan, Brukilacchio & Tager-Flusberg, 2016).

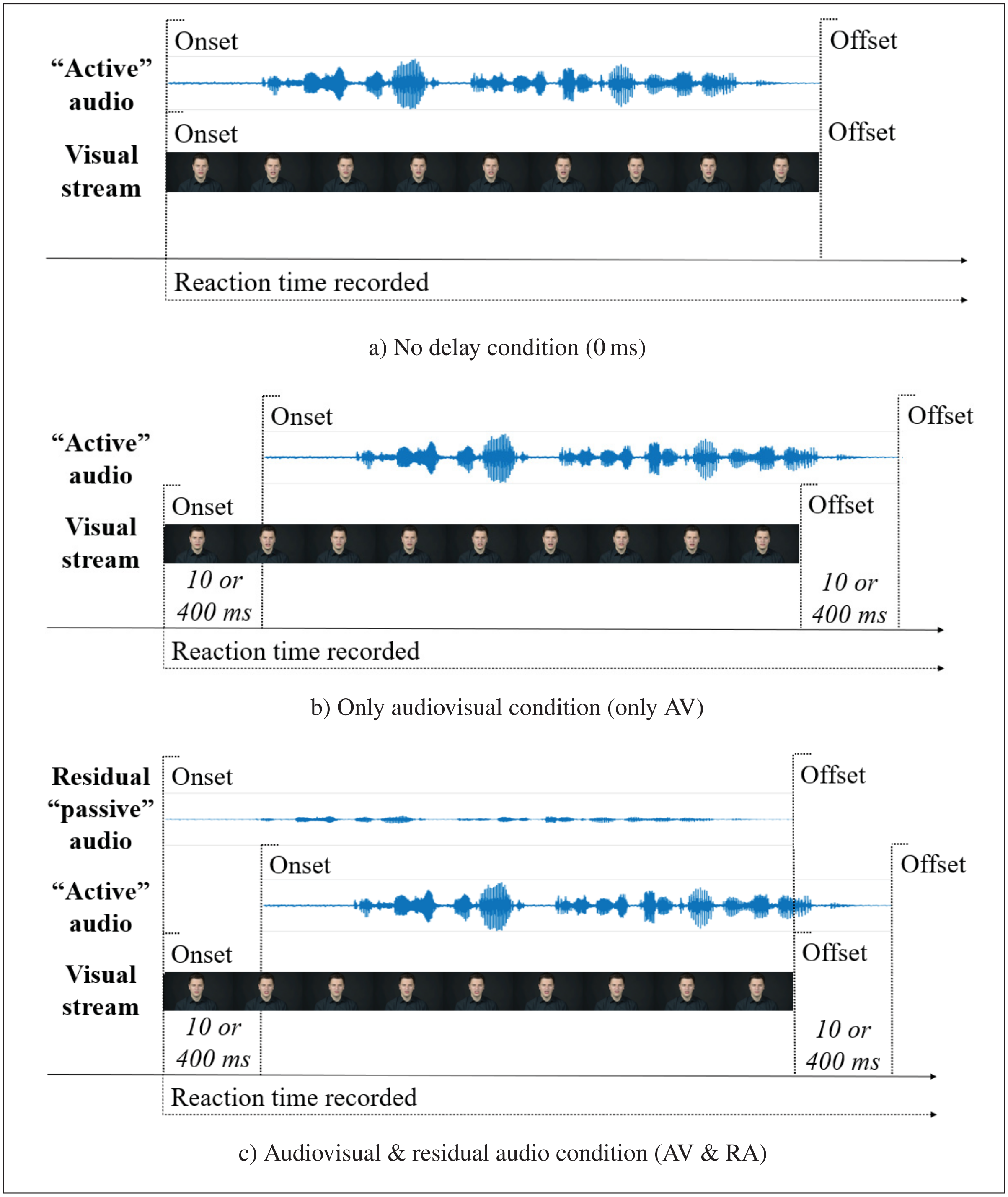
3.3.1 Materials

To investigate the effects of delays of different magnitude, two fixed values were chosen: the smaller delay, 10 ms, is widely considered to not impede perception of others' speech and seems to be an unofficial limit for hearing aid latency (Alexander, 2016). The larger delay, 400 ms, is generally above asynchrony detection thresholds for external speech (Summerfield, 1992; Lezzoum *et al.*, 2016), but is still in the range in which the delay in itself does not strongly impact conversational interactivity (Egger *et al.*, 2010). The stimuli were generated by modifying audiovisual speech recordings (Rosemann & Thiel, 2018) such that the actively transmitted audio was delayed by these two values either with respect to the visual signal alone, creating an audiovisual (AV) delay, or with respect to both the visual signal and the residual audio (RA). This resulted in the following five delay conditions, listed in Table 3.1:

Table 3.1 The names and descriptions of the delay conditions included in this study

Name	**Description**
0 ms	No delay
10 ms only AV	0 ms delay, only audiovisual
10 ms AV & RA	10 ms delay, audiovisual & residual audio
400 ms only AV	400 ms delay, only audiovisual
400 ms AV & RA	400 ms delay, audiovisual & residual audio

The 0 ms condition was the unmodified recording, depicted in Fig. 3.2a. In the 10 ms only AV and 400 ms only AV conditions, the auditory speech signal was modified to lag the visual speech signal, by 10 and 400 milliseconds, respectively (as depicted in Fig. 3.2b). Silence was inserted at the beginning of the audio stream before the start of the speech signal by appending the original auditory signal to the end of an array of zeros in MATLAB. The video stream was extended to match the length of the modified audio stream by showing the last frame repeatedly when the audio stream was still playing at the end. In the 10 ms AV & RA and 400 ms AV & RA conditions, the original auditory speech signal was also modified to lag the visual speech signal by the same delays. In addition, an attenuated copy of the original auditory speech signal, synchronized with the visual speech signal, was added to the audio stream in order to simulate the residual audio that would be passively transmitted through the earpiece material (as depicted in Fig. 3.1c). The attenuated copy was generated by applying the attenuation function shown in Fig. 3.3 to the original signal. Note that using the original auditory speech signal to simulate the delayed actively transmitted signal was a simulation of basic acoustical transparency; however, this method could be generalized to simulate the effects of any audio processing algorithm by delaying the processed signal rather than the original signal.



a) No delay condition (0 ms)

b) Only audiovisual condition (only AV)

c) Audiovisual & residual audio condition (AV & RA)

Figure 3.2 Representation of the delay conditions, depicting the onsets and offsets of the visual and auditory streams. The "active" audio simulates the audio actively transmitted through the earpiece speakers, while the residual "passive" audio simulates the the audio passively transmitted through the earpiece material. After the offset of the visual stream, the last frame of the video remains on the screen

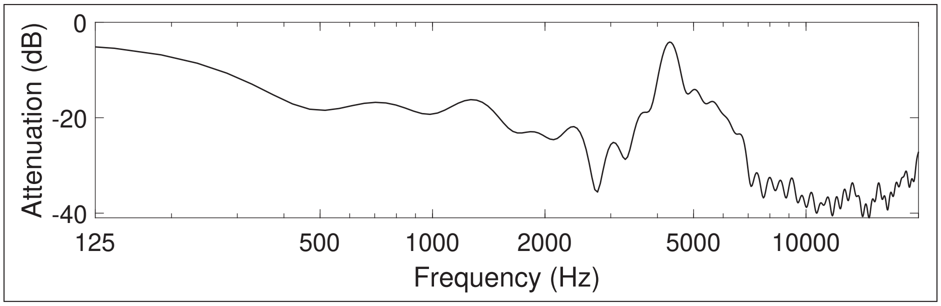


Figure 3.3 Magnitude of the finite impulse response filter computed from the measured transfer function of the earpiece, which was used to simulate the residual audio passively transmitted to the ear for generating delayed stimulus material in the residual audio condition

The attenuation function (shown in Fig. 3.3) was predicted with the field microphone-in-real-ear technique (Voix & Laville, 2009) using data recorded with one earpiece of the Auditory Research Platform version 3.1, a hearable prototype developed at the NSERC-EERS Industrial Research Chair in In-Ear Technologies (CRITIAS) containing microphones outside and inside the ear, shown in Fig. 3.1a and 3.1b. While the wearer was exposed to pink noise played from an external loudspeaker at 85 dB (SPL), the earpiece's outer-ear microphone recorded the external noise, and the in-ear microphone simultaneously recorded the residual audio passively transmitted to the ear (depicted in Fig. 3.1c). The in-ear and outer-ear recordings were then used to estimate the transfer function of the earpiece (Bouserhal, Falk & Voix, 2017). To obtain the attenuation function applied to simulate residual audio, the weighted least squares method was applied to compute the coefficients of a finite impulse response filter from the transfer function estimate. All steps of processing the audio stream were completed offline in MATLAB (MATLAB, 2019), after which the audio and video streams were merged and compressed into h.264 mpeg4 format using FFmpeg (Tomar, 2006).

Each recording contained one sentence from the Oldenburg linguistically and audiologically controlled sentences (OLACS) corpus (Uslar *et al.*, 2013), for example:

(1) *Der* *stille* *Postbote* *grüßt* *den* *dicken*
The-NOMINATIVE silent-NOMINATIVE mailman greets the-ACCUSATIVE big-ACCUSATIVE
Frisör.
hairdresser

'The silent mailman greets the big hairdresser.'

(2) *Den* *alten* *Pfarrer* *grüßt* *der* *kluge*
The-ACCUSATIVE old-ACCUSATIVE pastor greets the-NOMINATIVE clever-NOMINATIVE
Pilot.
pilot

'The clever pilot greets the old pastor.'

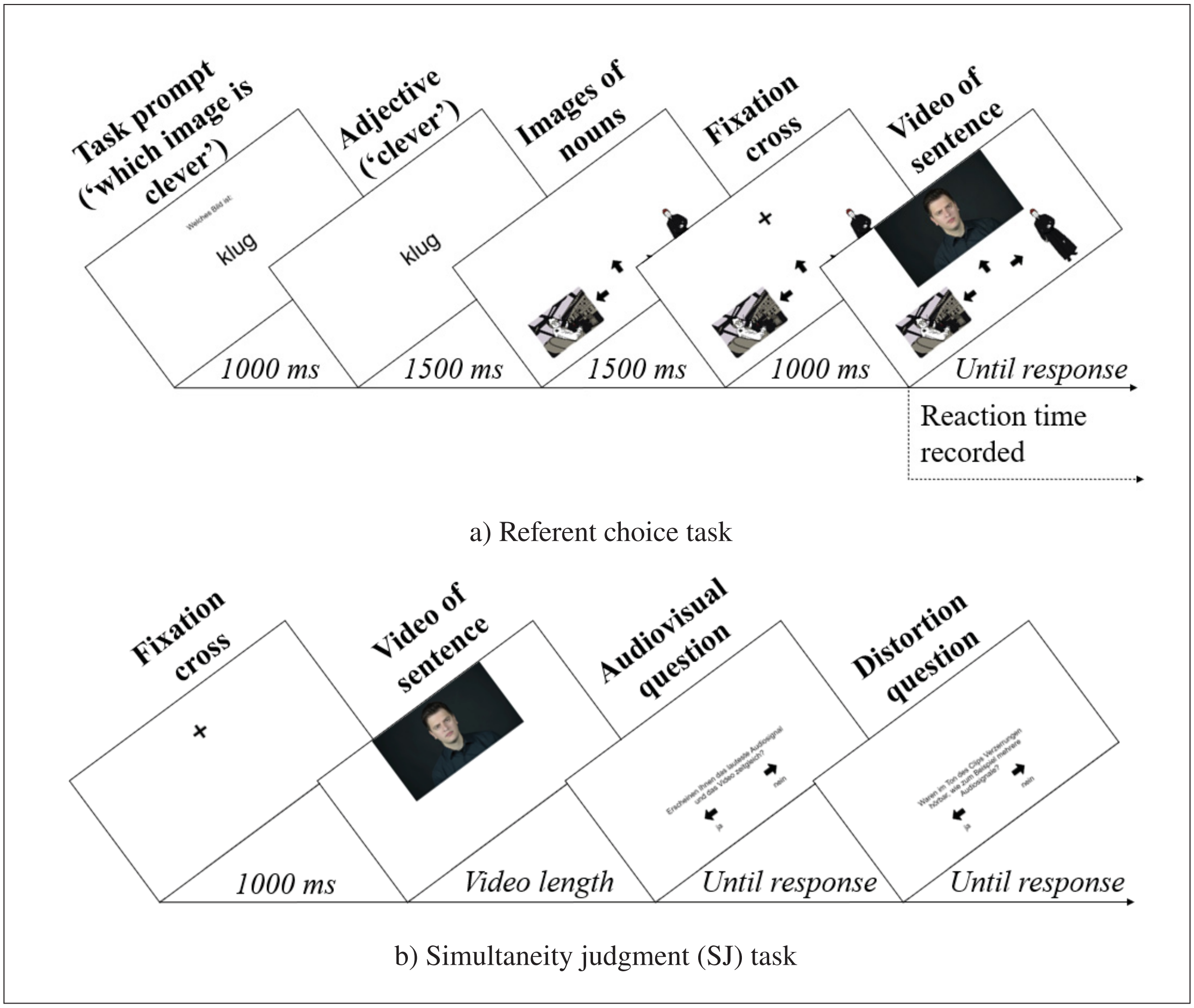
(3) *Der* *Frisör,* *der* *die* *Köchinnen* *erschießt,*
The-NOMINATIVE hairdresser who-NOMINATIVE the-ACCUSATIVE cooks shoots
grinst.
grins

'The hairdresser who shoots the cooks grins.'

Half of the sentences selected as stimuli for the referent choice task followed a subject-verb-object structure, as in example sentence (1). The other half followed an object-verb-subject structure as in example sentence (2). In German, the subject can appear before or after the object without changing the meaning of the sentence due to morphological case marking; in example sentences (1) and (2), the mailman and the pilot are in different positions in the sentence but are both in the nominative case, indicating that each are the subject. Nevertheless, the object-first structure is less common and widely considered to be non-canonical in German (Uslar *et al.*, 2013; Carroll & Ruigendijk, 2013; Bader & Häussler, 2010-03, 2010). All of the subjects and objects in the selected sentences were singular male nouns, for which the determiner unambiguously marks whether the noun phrase is in the nominative case (which in German is generally the subject) or the accusative case (generally the object). Including both subject-verb-object and object-verb-subject sentences in the referent choice task provided a manipulation of syntactic complexity, in which the object-first structure is considered more complex.

All of the sentences selected as stimuli for the simultaneity judgment task followed the same structure as in example sentence (3). Each sentence contained a male singular subject, an intransitive verb (a verb that does not require an object, e.g. 'grins'), and an embedded subject relative clause (e.g. 'who shoots the cooks'). Due to morphological case marking on the relative pronoun (e.g. 'who-NOMINATIVE'), there was no ambiguity of semantic roles within the relative clause.

For the referent choice task, the nouns in the sentences were represented by images (e.g. the pastor and pilot in Fig. 3.4a. There were 57 unique nouns in the sentences used for the referent choice task (some nouns appeared in multiple sentences). Of the 57 images used to represent nouns in the sentences, 43 were taken from the MultiPic corpus (Duñabeitia *et al.*, 2018) and 14 were found via Google Images. A list of the images and sentences used in this study can be found in Appendix I and Appendix II, respectively.



a) Referent choice task

Fixation cross
+
Video of sentence
Audiovisual question
Erscheinen Ihnen das lauteste Audiosignal und das Video zeitgleich?
ja
nein
Distortion question
Waren im Ton des Clips Verzerrungen hörbar, wie zum Beispiel mehrere Audiosignale?
ja
nein
1000 ms
Video length
Until response
Until response

b) Simultaneity judgment (SJ) task

Figure 3.4 Example screens from the referent choice and simultaneity judgment (SJ) tasks: (a) depicts the referent choice task: Following a prompt containing the target adjective (e.g. 'Which image is clever?'), the screen displays two images representing the two nouns in the sentence (e.g. a pilot and a pastor). The audiovisual speech stimulus, i.e. the sentence read by the speaker shown at the top of the screen, is then presented (e.g. 'The clever pilot greets the old pastor.') while reaction times are being recorded. (b) depicts the SJ task: A fixation cross is shown for 1000 ms, the audiovisual sentence stimulus is presented, and the participant is prompted to judge the audiovisual synchrony of the stimulus and whether there was any auditory distortion

The experiment was created using PsychoPy3 Experiment Builder v2021.1.2 (Peirce *et al.*, 2019) and ported to PsychoJS.

3.3.2 Participants

Fifty adult German native speakers participated in the online experiment (29 female and 21 male, mean age of 25.4, ranging from 19 to 56). All participants self-reported normal or corrected-to-normal vision and normal hearing, and no participants reported a diagnosis of Autism Spectrum Disorder. Participation in the study was reimbursed with partial course credit. The experiment was approved by the ethics committee of Osnabrück University.

3.3.3 Procedure

Participants were first requested to complete a questionnaire asking about their age, gender, handedness, native language, hearing impairment, and visual impairment. As optional questions, participants could volunteer whether they were diagnosed with Autism Spectrum Disorder, as well as whether they had any previous neurological or psychiatric disorders. Additionally, participants confirmed that they had no difficulties sitting quietly for 45 minutes, that they were wearing wired headphones, and that they were in a quiet room with comfortable lighting.

After providing consent and completing the questionnaire, participants were provided with a link to the experiment on *pavlovia.org*. All of the experimental stimuli were downloaded before the start of the experiment to mitigate playback issues. Before the experiment began, participants were instructed to adjust their volume such that the loudness of an example stimulus was comparable to a face-to-face conversation.

The main experimental task was the referent choice task, in which participants chose a picture that indicates a noun in a sentence that was modified by the target adjective. At the beginning of each trial, the task prompt containing the target adjective (e.g. “which image is clever”) was shown in the center of the screen for 1000 ms. Only the target adjective (e.g. “clever”) remained on the screen for another 1500 ms. Then, two images representing the two nouns in the sentence that was presented later in the trial appeared alone at the bottom of the screen for 1500 ms. These images remained at the bottom of the screen, and a fixation cross appeared at the top of the screen. After 1000 ms, the sentence was presented in a video that played once at the top of the screen. Participants indicated their answer based on a key press in the direction of the image representing the noun as shown in Fig. 3.4a. There was no time limit for the response, but participants were instructed to answer as quickly as possible, using the middle and index finger of their dominant hand. Accuracy and reaction time were recorded for each response. Reaction time was defined as relative to the onset of the visual stream of the sentence stimulus, as shown in Fig. 3.2 and 3.4a. In case participants had not responded by the offset of the sentence, the video would remain on a still image of its last frame until a response was recorded.

Following an introduction to the referent choice task, participants completed five training trials. During the experiment, participants completed 160 referent choice trials, in which the five delay conditions were each presented 28 times. Of the 160 trials, 20 were filler trials, in which the adjective shown was not corresponding to either of the two target images. In this case the participant was instructed to press the upper arrow key. There were 10 sentences selected for the 20 filler trials and 70 sentences selected for the other 140 referent choice trials. Each sentence was presented twice with two different target adjectives. In the filler trials, the adjective in the prompt was randomly chosen from another sentence in the corpus and was not present anywhere in the current sentence. In the other trials, the target adjective for each sentence was, in one trial, the adjective in the first half of the sentence and, in another trial, the adjective in the second half of the sentence. The order of the trials was pseudorandomized such that trials in immediate sequence never had identical target words, delay conditions, images, and sentences. The trials were then split into 10 blocks containing 16 trials each. The five training trials prior to the experiment used five different sentences that were later presented during the filler trials, each in a different delay condition. No feedback on correct answers was given during the training trials or the experiment. After each of the 10 main trial blocks, participants could choose to take a break for an unrestricted time.

In addition to the referent choice task, participants were asked to judge the simultaneity of audiovisual stimuli in three blocks, before, after, and in the middle of the 10 main trial blocks, respectively. In this modified simultaneity judgment task, participants answered two questions during each trial: first (1) whether they perceived the loudest audio signal to be synchronous with the visual stimulus or not, and then (2) whether they perceived any auditory distortion, such as multiple overlapping audio signals. The wording of this question was chosen to remain agnostic as to which type of auditory distortions might be perceived with the addition of the residual audio, since there are multiple possible perceptual effects (e.g. comb filtering). Each simultaneity judgment block consisted of 10 randomly ordered trials, and each of the five conditions was repeated twice in each block with a different sentence. There were 10 sentences selected for the 30 simultaneity judgment trials, and these 10 sentences were repeated in each block but always appeared in different delay conditions. As in the referent choice task, the sentence stimuli in the simultaneity judgment task were preceded by a fixation cross and played only once per trial (see Fig. 3.4b).

3.3.4 Analysis

Prior to analysis, four participants were excluded because they consistently judged the 0 ms condition as asynchronous and distorted on the modified simultaneity judgment task. In addition, participants' accuracy on the target and filler trials in the referent choice task was examined to determine attentiveness throughout the experiment, but no participants were removed on that basis (all participants had an accuracy > 80 %). All further analyses were carried out using the lme4 package (version 1.1-23) (Bates *et al.*, 2018) in R (version 4.0) (R Core Team, 2019). Reaction times on the referent choice task were cleared of outliers by removing all reaction times that were more than 2.5 standard deviations away from the individual participant's mean, and were subsequently analyzed using linear mixed effects models. The binary responses provided in the referent choice task and the simultaneity judgment task were analyzed using binary logistic regression models. Two models were fit to each of these data sets: The first model, which was utilized for a baseline comparison, treated the five conditions as treatment-coded predictor variable using the 0 ms condition as intercept level. The second model fit to the data was utilized to directly compare the critical condition with a (10 ms or 400 ms) only AV delay to the respective AV & RA condition. To that end, two sum-coded comparisons were entered into the model, which compared the two 400 ms conditions and the two 10 ms conditions to each other, respectively.

Since the target word in the referent choice task could appear in either the first or second half of the sentence, an interaction with the target position was added to the models as a treatment coded predictor. Finally, as task performance may change with increased exposure to the task, an interaction with the trial or block number was included: the trial number as a numeric predictor for the referent choice task or the block number as a treatment-coded predictor for the modified SJ task, respectively. All models were initially fit using their maximal random effects structure (Barr, Levy, Scheepers & Tily, 2013). Since these models did not converge, random effects were removed in a stepwise procedure, with the most complex converging models being ones that used random by-subject and by-item intercepts only. The data and code are available at a repository on Open Science Framework (Benesch, Schwab, Voix & Bouserhal, 2022c).

3.4 Results

3.4.1 Reaction times

Reaction times (RTs), as depicted in Fig. 3.5, were significantly higher in the two 400 ms conditions than at baseline level, whereas there was no evidence for a difference between baseline and the two 10 ms conditions (Table 3.2). As expected, two main effects of target position and trial number were found, with a higher RT if the target appeared in the second half of the sentence and a reduced RT with increased exposure to the task. Interestingly, there was also a significant interaction between target position and trial number, such that RTs were reduced more strongly over the course of the experiment for the target in the first half of the sentence compared to the target in the second half of the sentence. None of the three-way interactions were significant, and details on the model output can be found in Appendix III. The secondary model further elucidates these findings through two interaction effects involving the two 400 ms conditions (see Table 3.3): The significant interaction with target position indicates that participants were faster to respond in the AV & RA condition, but only if the target appeared in the first half of the sentence (see also Fig. 3.5. Moreover, the interaction with the trial number showed a larger reduction in RTs over the course of the experiment for the 400 ms AV & RA condition compared to the 400 ms only AV condition.

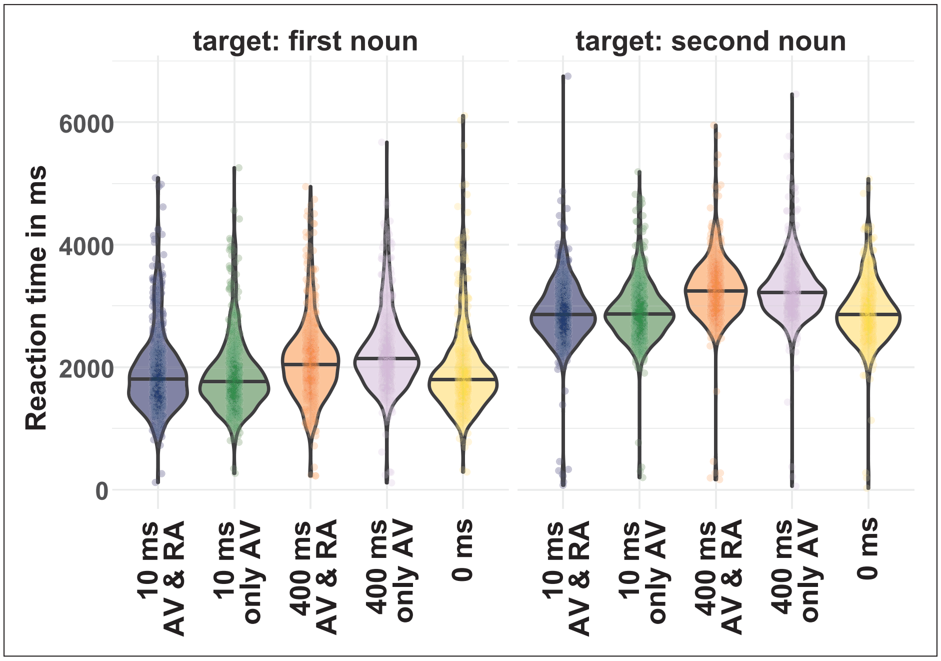


Figure 3.5 Reaction times in ms on the referent choice task for the five delay conditions, showing a slower response in the two 400 ms conditions than in the 0 or 10 ms conditions. The reaction times to the task are plotted separately according to where the target word was located in the sentence, demonstrating a relatively faster response for the 400 ms delay when residual audio was present, but only when the target was in the first part of the sentence. Violin plots show the distribution of the data. The thick black line indicates the mean

Table 3.2 Model coefficients for the linear mixed effects model fit to reaction times and the binary logistic regression model fit to accuracy in the referent choice task (comparison against baseline). Details on the full model, including three-way interactions, are included in Appendix III. Estimates for reaction time are expressed in milliseconds, and estimates for accuracy are expressed in log odds ratios. * = p ≤ 0.05, ** = p ≤ 0.01, *** = p ≤ 0.001

	Reaction times			**Accuracy**		
Measure	Estimate	SE	*z*-value	Estimate	SE	*z*-value
(Intercept)	1907.20	50.97	37.42***	4.07	0.31	13.27***
10 ms only AV	-27.54	26.54	-1.04	0.21	0.38	0.54
400 ms only AV	370.55	26.69	13.89***	-0.59	0.33	-1.77
10 ms AV & RA	28.02	26.51	1.06	-0.31	0.34	-0.90
400 ms AV & RA	219.91	26.67	8.25***	-0.43	0.34	-1.28
position (second)	993.33	26.53	37.45***	-0.78	0.32	-2.40*
trial number	-176.87	20.03	-8.83***	0.53	0.27	1.96*
position:trial	81.64	27.40	2.98**	-0.24	0.33	-0.75
10 ms only AV:position	54.31	37.20	1.46	-0.24	0.45	-0.52
400 ms only AV:position	13.46	37.42	0.36	0.42	0.41	1.01
10 ms AV & RA:position	-24.86	37.22	-0.67	0.06	0.42	0.14
400 ms AV & RA:position	158.43	37.42	4.23***	0.09	0.41	0.22

An additional analysis was conducted for syntactic complexity, that is, differences between subject-verb-object and object-verb-subject order sentences. The models indicated that participants had a significantly lower reaction time in the subject-verb-object condition (β = 80.00, SE = 32.09, t = 2.49, see Appendix III), but crucially there were no interactions with the five delay conditions, the position of the target, or the trial number.

In summary, compared to baseline, participants responded more slowly on the referent choice task if the auditory information was delayed by 400 ms. The contrast between the 400 ms only AV condition and the 400 ms AV & RA condition showed that participants were *faster* to respond to the cue if the residual audio was present than if it was not. This effect was restricted, however, to the conditions where the target word appeared in the first half of the sentence (see the interaction effect in Table 3.3).

Table 3.3 Model coefficients for the linear mixed effects model fit to reaction times and the binary logistic regression model fit to accuracy in the referent choice task (contrast only AV vs. AV & RA). Details on the full model, including all two-way and three-way interactions, are included in Appendix III. Estimates for reaction time are expressed in milliseconds, and estimates for accuracy are expressed in log odds ratios. * = p ≤ 0.05, ** = p ≤ 0.01, *** = p ≤ 0.001

	Reaction times			**Accuracy**		
Measure	Estimate	SE	*z*-value	Estimate	SE	*z*-value
(Intercept)	2052.95	47.03	43.66***	3.79	0.17	21.87***
10 ms only AV vs. AV & RA	-54.33	27.68	-1.96*	0.42	0.31	1.36
400 ms only AV vs. AV & RA	149.25	28.01	5.33***	-0.20	0.31	-0.67
position (second)	1043.50	13.90	75.08***	-0.68	0.12	-5.58***
trial number	-171.80	9.92	-17.32***	0.31	0.10	3.15**
position:trial	93.12	13.98	6.66***	-0.19	0.12	-1.57
10 ms only AV vs. AV & RA:position	75.48	39.10	1.93	-0.22	0.39	-0.56
400 ms only AV vs. AV & RA:position	–144.96	39.54	-3.67***	0.37	0.38	0.96
10 ms only AV vs. AV & RA:trial	6.05	27.51	0.22	-0.19	0.30	-0.66
400 ms only AV vs. AV & RA:trial	58.82	28.53	2.06*	0.25	0.31	0.79

3.4.2 Response accuracy

3.4.2.1 Referent choice task

Accuracy for the task was high overall. The conditions with the lowest accuracy were the two 400 ms conditions with 94% correct responses on both, whereas the 10 ms only AV condition received the highest accuracy with 96% correct responses, with no significant differences for the delay conditions compared to the 0 ms baseline (Table 3.2). A significant effect of target position was found (*z* = -2.40, see Table 3.2), with lower accuracy if the target appeared in the second half of the sentence (as shown in Fig. 3.6). A marginal effect of trial number was found with increasing accuracy over the course of the experiment (*z* = -1.96, see Table 3.2). There was no significant effect of syntactic complexity on accuracy (see Appendix III).

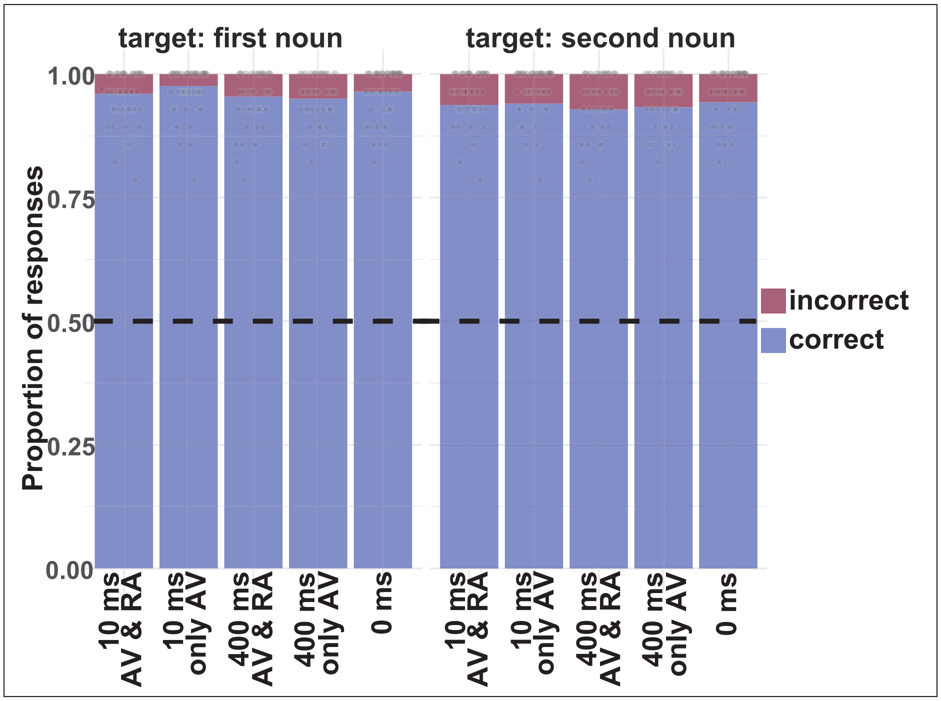


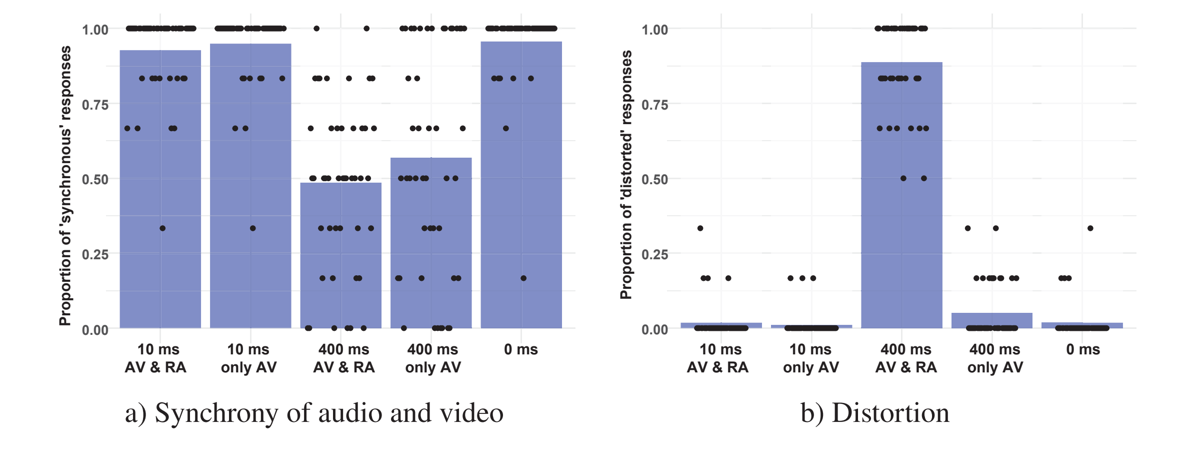
Figure 3.6 Accuracy as the proportion of correct responses on the referent choice task for the five delay conditions, showing an overall high accuracy regardless of whether the stimuli were delayed. Dots indicate the mean accuracy of individual participants

3.4.2.2 Simultaneity judgments

The modified simultaneity judgment task contained two questions, one regarding the synchrony of the (loudest) auditory signal and the video, and one regarding any audible auditory distortions. Responses on the first question revealed a significant difference between the baseline 0 ms condition and the two 400 ms conditions (Table 3.4), as participants were less likely to respond that the signals were synchronous in the latter two conditions. Moreover, the secondary model revealed a difference between the two 400 ms conditions (β = -0.49, SE = 0.23, z = -2.18), such that there were fewer "synchronous" responses in the 400 ms AV & RA condition (48.6% for the AV & RA condition vs. 56.9% for the only AV condition). There were no differences found between the 0 ms baseline condition and the two 10 ms conditions, suggesting that participants may not have consciously perceived this small audiovisual delay.

3.4.2.3 Distortion judgments

Reflecting the findings on the audiovisual synchrony question, responses on the auditory distortion question revealed significant differences between the 0 ms baseline condition and the two 400 ms conditions (Table 3.4). Furthermore, the secondary model showed a clear difference between the two 400 ms conditions (β = 9.26, SE = 0.46, z = 20.13), demonstrating that participants recognized the presence of two overlapping auditory signals. The two 10 ms conditions did not differ from each other nor from the baseline condition. Responses on the audiovisual synchrony and distortion judgments were moderately correlated when pooling the data from all five conditions (r(1378) = -0.34, $p < 0.001$), which was expected as the only condition in which participants reliably recognized the residual audio was also a condition with an audiovisual delay. More importantly, given substantial variance on the audiovisual synchrony judgments for the two 400 ms conditions, the responses on these two conditions were also submitted to separate correlation tests. Here, however, no significant correlations were apparent (400 ms only AV: r(274) = -0.07, p = 0.28; 400 ms AV & RA: r(274) = -0.02, p = 0.72).



a) Synchrony of audio and video

b) Distortion

Figure 3.7 Responses for the two parts of the modified simultaneity judgment task: (a) responses to the question concerning whether audio and video were synchronous, showing that both 400 ms conditions were rated less frequently as synchronous compared to the 0 ms condition; (b) responses to the question concerning whether there was audio distortion such as multiple signals, showing that the 400 ms residual audio condition was rated more frequently as distorted. Dots indicate individual participants' mean response across repeated measures

Table 3.4 Model coefficients for the binary logistic regression models fit to the responses on the simultaneity judgment task (comparison against baseline). Full model coefficients (including the block number) are included in Appendix III. Estimates for accuracy are expressed in log odds ratios. * = p ≤ 0.05, ** = p ≤ 0.01, *** = p ≤ 0.001

	Audiovisual synchrony			**Distortion**		
Measure	Estimate	SE	*z*-value	Estimate	SE	*z*-value
(Intercept)	-3.71	0.39	-9.43***	3.84	0.51	7.60***
10 ms only AV	0.17	0.42	0.41	0.52	0.74	0.71
400 ms only AV	3.43	0.37	9.32***	-1.08	0.53	-2.03*
10 ms AV & RA	0.65	0.40	1.61	-0.00	0.64	0.00
400 ms AV & RA	3.84	0.36	10.68***	-6.45	0.55	-11.75***

3.5 Discussion

This experiment aimed to explore the effects of a hearable's audio processing latency on language understanding. To this end, participants completed a referent choice task while the perceptual delays expected when wearing a hearable were simulated: the auditory speech signal was delayed relative to the corresponding visual speech signal, as well as to the residual audio not fully attenuated by the earpiece. The results of this experiment demonstrated a slower response to the referent choice task when the auditory speech signal lagged the visual speech signal by 400 ms. This could indicate that the efficiency of speech processing suffered under the audiovisual delay. However, one complicating factor is that a longer reaction time could also be caused by a reliance on the delayed auditory stream, that is, that participants waited for the delayed auditory information and therefore took more time to complete the task.

The analysis of reaction times also revealed that participants benefited from the presence of the residual audio when the original auditory signal was delayed by 400 ms, but that benefit was diminished when the target word appeared in the second half of the sentence. It is possible that cognitive resources required to process the delayed speech increase with the complexity and length of the stimulus, highlighting the need to investigate continuous speech rather than only single words or syllables. This explanation would be in line with previous research that found that given only visual speech signals, speech perception was less accurate for sentence-length stimuli compared to single words (Kyle, Campbell, Mohammed, Coleman & MacSweeney, 2013). The effect of the sentences' syntactic complexity further supported that reaction time to the referent choice task at least partially reflected efficiency of higher-order language processing. Participants responded more quickly to sentences in the canonical subject-first order, consistent with previous studies that have found a preference for subject-first constructions in German (Bader & Häussler, 2010-03, 2010; Carroll & Ruigendijk, 2013) and other languages (see e.g. Gordon, Hendrick & Levine, 2002). The lack of an interaction between syntactical complexity and either target position or delay condition may be due to the simplicity of the task, as the way that syntactic complexity interacts with other factors could depend on how cognitive resources are allocated. The question used for the referent choice task (e.g. "Which image was clever?" for the sentence "The old pastor greets the clever pilot.") allowed participants to solve it simply by attaching the adjective ("clever") to the noun it modifies ("pilot"). Alternatively, one could increase the complexity of the required linguistic processing operations by probing the argument structure ("Which image was greeted?"), for which the answer depends on the morphological case marking (Carroll & Ruigendijk, 2013). This would tap directly into the different complexity of subject-verb-object and object-verb-subject sentences, for which an effect on reaction times was already found, even with a task that did not specifically target this difference in complexity.

For both syntactic complexity and the delay conditions, there were no significant differences in referent choice task accuracy. This may have been a ceiling effect due to the previously noted simplicity of the task: participants achieved high accuracy in this task in all conditions. While access to temporally congruent visual speech signals has been shown to enable more accurate speech perception (Summerfield, 1992), the speech tested has generally been conceptually challenging, e.g. a passage from Kant's Critique of Pure Reason (Arnold & Hill, 2001-05, 2001; Reisberg, Mclean & Goldfield, 1987) or perceptually challenging, e.g. degraded with babble noise (Pandey *et al.*, 1986; Sumby & Pollack, 1954). A relatively simple task was chosen to allow for the protocol to be adaptable to different populations; however, future work could investigate the impact of task difficulty by having a range of tasks. For example, as previously mentioned, the task could potentially be made more difficult by including questions that exploit the non-canonical word order in German, or, for a more language-independent modification, by including novel pairings between (unknown) words and referents in addition to known words (see e.g. "modi" in Weatherhead, Arredondo, Nácar Garcia & Werker, 2021).

Overall, the main trends for the two simultaneity judgment questions were as expected: in line with prior research, when the auditory speech signal lagged the visual speech signal by 400 ms, the stimulus was rated significantly more frequently as audiovisually asynchronous, although it should be noted that the proportion of synchronous responses (shown in Fig. 3.7a) was high for a 400 ms audio lag relative to existing literature (cf. Gordon-Salant *et al.*, 2022; Navarra *et al.*, 2010). When the unmodified auditory speech signal lagged an attenuated copy by 400 ms, the stimulus was rated as distorted significantly more often. However, the presence of residual audio also played a role in the audiovisual synchrony question; participants were less likely to judge the 400 ms delay condition as synchronous when there was residual audio, that is, when there was an attenuated copy of the audio signal synchronized with the visual signal. Similarly, for the question on distortion intended to evaluate the effect of residual audio, the audiovisual delay on its own in the 400 ms condition without residual audio had marginally higher distortion responses than the 0 ms reference. One possible explanation for these differences could be that after exposure to the stimuli, some participants may have associated the 400 ms audiovisual asynchrony with the 400 ms residual audio condition, using one asynchrony to detect the other. To further investigate this, the relationship between the two SJ task questions for the 400 ms delay conditions was examined. Within each condition, there was no significant correlation between the audiovisual synchrony responses and the distortion responses, suggesting that most participants were not using the fact that there was always an audiovisual asynchrony when there was residual audio as a strategy when answering the two questions. Note also that in contrast to the high variance of the audiovisual synchrony responses within each 400 ms condition (see Fig. 3.7a), the distortion judgments were more consistent, with a high proportion of responses correctly indicating that the 400 ms condition with residual audio was distorted and the 400 ms condition without residual audio was not distorted (see Fig. 3.7b).

For the 10 ms delay conditions, there were no significant differences in reaction time and accuracy in both tasks. While it is always difficult to interpret null results, here it should be noted that due to the format of the study, there may have been unintentional variation in the true audiovisual delay of the stimuli, as well as the measured reaction time. In an analysis of delay variability in online studies on *pavlovia.org* (Bridges, Pitiot, MacAskill & Peirce, 2020), PsychoPy was found to have inter-trial standard deviations of less than 1 ms for reaction time and audiovisual asynchrony; however, the constant lag for each operating system ranged from 8.43 ms to 22.02 ms for reaction time and 5.43 ms to 17.70 ms for audiovisual asychrony, with the audio leading in all cases. Therefore, for all participants, the audiovisual delay relative to the reference 0 ms condition likely remained roughly constant, but between participants, the absolute audiovisual delay could have varied. Although these variations are not as problematic for larger delays, evaluating the effects of small imperceptible delays in an online study is especially challenging because participants cannot directly confirm that the stimuli are presented as intended.

3.6 Conclusions

As hearables show promise to improve communication in a wide variety of contexts, the minimum acceptable latency value may need to be reevaluated for these new applications. This research laid the groundwork for a protocol to investigate the audio delays expected with hearables and their impact on higher-order speech understanding, which could be applied to subjectively test a wide variety of algorithms. Certain parameters should be adapted to the envisioned application, such as the complexity of the task, the addition of background noise, and the latency values tested, as well as how results translate into concrete latency requirements.

3.7 Acknowledgements

The authors would like to thank Stephanie Rosemann and Christiane M. Thiel for sharing the audiovisual speech recordings used in this study, as well as Aron Petau for implementing the study and collecting data as part of his undergraduate thesis. The financial support received from the Natural Sciences and Engineering Research Council of Canada (NSERC) and EERS Global Technologies Inc, through the NSERC-EERS Industrial Research Chair in In-Ear Technologies (CRITIAS) is gratefully acknowledged. The first author acknowledges the financial support received from the German Academic Exchange Service (DAAD), as well as from the MITACS Globalink Graduate Fellowship.

CHAPTER 4

AUTOMATIC STRESS RECOGNITION WITH IN-EAR HEARTBEAT SOUNDS

Danielle Benesch[1] , Bérangère Villatte[2] , Alain Vinet[3] , Sylvie Hébert[2] , Jérémie Voix[1] , Rachel Bouserhal[4]

[1] Department of Mechanical Engineering, École de Technologie Supérieure, 1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

[2] School of Speech-Language Pathology and Audiology, Université de Montréal, 7077 Av du Parc, Montréal, Québec, Canada, H3N 1X7

[3] Centre de Recherche en Physiologie Cardiovasculaire, Hopital du Sacré Coeur, 5400 Boul Gouin Ouest, Montréal, Québec, Canada H4J 1C5

[4] Department of Electrical Engineering, École de Technologie Supérieure, 1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Article submitted to “IEEE Transactions on Affective Computing” in December 2022.

4.1 Abstract

Although stress plays a key role in tinnitus and decreased sound tolerance, conventional hearing devices used to manage these conditions are not currently capable of monitoring the wearer’s stress level. To assess the feasibility of stress monitoring with an in-ear device, in-ear heartbeat sounds and clinical-grade electrocardiography (ECG) signals were simultaneously recorded while 30 healthy young adults underwent a stress protocol. Heart rate variability features were extracted from both signals to train classification algorithms to predict stress vs. rest. Models trained and tested using in-ear heartbeat sounds appeared to perform better than the models trained and tested using the ECG signals. However, further analyses comparing heart rate variability features extracted from ECG and the in-ear heartbeat sounds suggest that the improvement in stress prediction performance was driven by the increased presence of artifacts during the stress tasks. To address this difference in error between rest and stress conditions, a data augmentation method was proposed to balance the error. The final proposed system demonstrates the viability of robust stress recognition with only in-ear heartbeat sounds.

4.2 Introduction

Psychological stress is closely associated with tinnitus, the subjective perception of sound in the absence of an external source (Mazurek, Boecking & Brueggemann, 2019), and decreased sound tolerance, the intolerance to sounds that do not bother the average listener (Jastreboff & Jastreboff, 2014). However, the causal relationship between these conditions and stress is not fully understood; stress may be both a direct consequence of tinnitus and decreased sound tolerance and an external factor in their development and severity (Mazurek, Szczepek & Hebert, 2015; Mazurek *et al.*, 2019; Hasson *et al.*, 2013; Jastreboff, 2004; Schlee *et al.*, 2023; Hébert, Canlon & Hasson, 2012). The time-varying and situation-dependent nature of tinnitus and decreased sound tolerance may restrict the conclusions that can be drawn from laboratory studies, which are often limited in duration and external validity (Wilhelm & Grossman, 2010).

Recent advancements in biosignal-based stress recognition with wearables have offered the potential for automatically monitoring stress over longer periods of time in real-life settings (Giannakakis *et al.*, 2019a). Hearables, wearable devices worn at the level of the ear, could be especially well-suited for research on tinnitus and decreased sound tolerance as hearables are already commonly used to manage these conditions (Jastreboff, 2004; Jastreboff & Jastreboff, 2014; Cavanna & Seri, 2015; Neave-DiToro *et al.*, 2021; Sammeth *et al.*, 2000; Neave-DiToro *et al.*, 2021). Given real-time information on the wearer's psychophysiological state and external environment, hearables may also be able to deliver improved therapies over time (Benesch, Raj, Bouserhal & Voix, 2021c; Searchfield *et al.*, 2021).

Heartbeat sounds, amplified inside the earcanal when wearing an occluding device, are among the biosignals that can be captured by a hearable and detected automatically (Martin & Voix, 2018). Heartbeat signals are widely used in stress research (Giannakakis *et al.*, 2019a), and previously, automatic stress and emotion recognition has been achieved with only heartbeat signals (Giannakakis, Marias & Tsiknakis, 2019b; Melillo, Bracale & Pecchia, 2011; Szakonyi, Vassányi, Schumacher & Kósa, 2021; Kaur *et al.*, 2014; Cho *et al.*, 2019; Ferdinando, Ye, Seppänen & Alasaarela, 2014; Xiefeng, Wang, Dai, Zhao & Liu, 2019; Prajod & André, 2022; Arquilla, Webb & Anderson, 2022). However, features derived from different types of heartbeat signals are not identical (Yuda *et al.*, 2020) and can be sensitive to artifacts common in wearable sensor data (Nikolic-Popovic & Goubran, 2013). While stress monitoring with electrocardiography (ECG) signals has been well-established (Giannakakis *et al.*, 2019a), it is currently unclear whether similar results could be achieved with automatically processed in-ear heartbeat sounds. Therefore, the aim of this work is to assess the feasibility of stress monitoring using heartbeat sounds recorded with an in-ear device, by evaluating a stress recognition system to predict acute stress induced by experimental tasks in a typical population.

To this end, a dataset was created with clinical-grade ECG and in-ear audio signals collected from 30 healthy young adults when they were at rest and performing two stress tasks: mental arithmetic (Kirschbaum, Pirke & Hellhammer, 1993) and the Cold Pressor Test (Lovallo, 1975). This dataset was used to train systems for stress recognition. These systems were first benchmarked using heart rate variability (HRV) features extracted from the ECG signals. Then, the same features were extracted from audio signals captured within the occluded earcanal, and the performance of the resulting classifiers were compared to those trained on the ECG. Finally, as the error between features extracted from ECG and in-ear audio was higher during the stress conditions, a data augmentation method was designed to control for this error, and its utility was empirically verified.

4.3 Background

Psychological stress can affect the way the autonomic nervous system regulates bodily functions such as heart activity, respiration, and perspiration (Giannakakis *et al.*, 2019a). It has been suggested that tinnitus and decreased sound tolerance involve activation of the autonomic nervous system (Jastreboff, 2004), and measurements of autonomic nervous system-dependent biosignals have been used to study these conditions (Kumar *et al.*, 2017; Grossini *et al.*, 2022; Ferrer-Torres & Giménez-Llort, 2021; Shepherd, 2016; Reinhart, Griffin & Micheyl, 2021; Pfeiffer *et al.*, 2019b; Edelstein *et al.*, 2013; Choi *et al.*, 2013). In the general context of ambulatory stress monitoring, automatic stress recognition systems have been developed using biosensors, with some multimodal systems relying on multiple biosignal inputs (Jongyoon Choi, Ahmed & Gutierrez-Osuna, 2012; Mishra *et al.*, 2020; Akmandor & Jha, 2017), while others have been developed using only one type of sensor (Giannakakis *et al.*, 2019b; Melillo *et al.*, 2011; Szakonyi *et al.*, 2021; Kaur *et al.*, 2014; Cho *et al.*, 2019; Prajod & André, 2022; Arquilla *et al.*, 2022).

Heartbeat signals have shown promise for measuring stress, as both the parasympathetic "rest and digest" and sympathetic "fight or flight" responses of the autonomic nervous system influence heart activity (Kim *et al.*, 2018). Beat-to-beat variability in the heart rate can interact with different bodily functions such as respiration and blood pressure (Shaffer & Ginsberg, 2017). These systems can influence the heart rate simultaneously in different ways, and changes in the heart rate at different time scales can reflect separate neurophysiological mechanisms (Pham, Lau, Chen & Makowski, 2021a). Given the exact number of milliseconds between consecutive heartbeats, referred to as interbeat intervals, multiple HRV measures can be extracted. These measures can be broadly categorized into time-domain, frequency-domain, and nonlinear features depending on the signal processing methodology used to extract them from the interbeat interval signal (Shaffer & Ginsberg, 2017). ECG is considered to be the gold standard heartbeat signal for HRV analysis, as ECG recordings are less likely to be distorted by artifacts and because the interbeat intervals can be precisely computed from the difference between the sharp, spike-like R-peaks in the ECG signal (Laborde *et al.*, 2017). Accordingly, much of the research on automatic stress recognition to date has focused on HRV features extracted from ECG (Giannakakis *et al.*, 2019a; Szakonyi *et al.*, 2021; Giannakakis *et al.*, 2019b; Melillo *et al.*, 2011; Kaur *et al.*, 2014; Cho *et al.*, 2019; Prajod & André, 2022; Arquilla *et al.*, 2022). Nevertheless, in recent years, numerous novel data acquisition methods have been proposed for measuring HRV, such as contactless microwave radar (Shi *et al.*, 2021), wristband photoplethysmogram (PPG) (Ollander, Godin, Campagne & Charbonnier, 2016), and audio signals recorded from the shoulder (Xiefeng *et al.*, 2019).

4.4 Methods

4.4.1 Dataset

4.4.1.1 Participants

In order to first establish the feasibility of stress monitoring with an in-ear device in a typical population, 30 healthy and normal hearing young adults were included in this study (15 men, age 27.3±4.7 years and 15 women, age 27.1±3.1 years). Recruitment was conducted through advertisements within the Université de Montréal, social networks, and word-of-mouth. To be included in the study, participants had to be aged between 18 and 45 years old and consider themselves to be physically and psychologically healthy. Exclusion criteria were having hearing disorders including tinnitus or hyperacusis, cardiovascular and cerebrovascular disorders, respiratory diseases, Raynaud's syndrome, diabetes, hyperglycemia, or taking any medication that could interfere with stress reactivity. Each participant signed a written informed consent before starting the experiment and the project was approved by the Comité d'éthique pour la recherche clinique (CERC), the Institutional Review Board of Université de Montréal (QC, Canada).

4.4.1.2 Materials

Hearing thresholds. Audiometry was performed in a soundproof audiometric booth using a calibrated AC40 clinical audiometer (Interacoustics, Middelfart, Denmark) with Telephonics TDH-39 headphones. Thresholds were measured with pure tones in octave bands between 250 Hz and 8,000 Hz using the Hughson-Westlake auditory threshold tracking procedure. Participants' eligibility was determined based on having a pure tone average (PTA) no greater than 15 dB HL (i.e., within the normal hearing limits) at frequencies of 0.5, 1, 2 and 4 kHz.

Auditory Research Platform In-ear heartbeat sounds were recorded with the Auditory Research Platform (ARP), an in-ear wearable technology developed within the ÉTS-EERS Industrial

Research Chair in In-Ear Technologies (CRITIAS). Each earpiece, illustrated in Fig. 4.1, contains an outer-ear microphone (OEM), an in-ear microphone (IEM), and an internal miniature loudspeaker (SPK). Four channels of audio were recorded from the two OEMs and IEMs at a sampling rate of 44,100 Hz with MATLAB 2020a (Mathworks, Natick, MA, USA). The right earpiece was in "transparency mode", meaning that external sound picked-up by the OEM was played back on the SPK inside the wearer's ear at unity gain. Only the IEM signal acquired from the left earpiece was used for the experiments described in this paper.

Electrocardiogram (ECG) A clinical Burdick Altair-disc Holter (Spacelabs, Deerfield, USA) with five Ag/AgCl self-adhesive electrodes (3M Healthcare, Canada) positioned in 5-lead configuration was used for continuous heart rate recording. Raw ECG signal was sampled at a rate of 500 Hz.

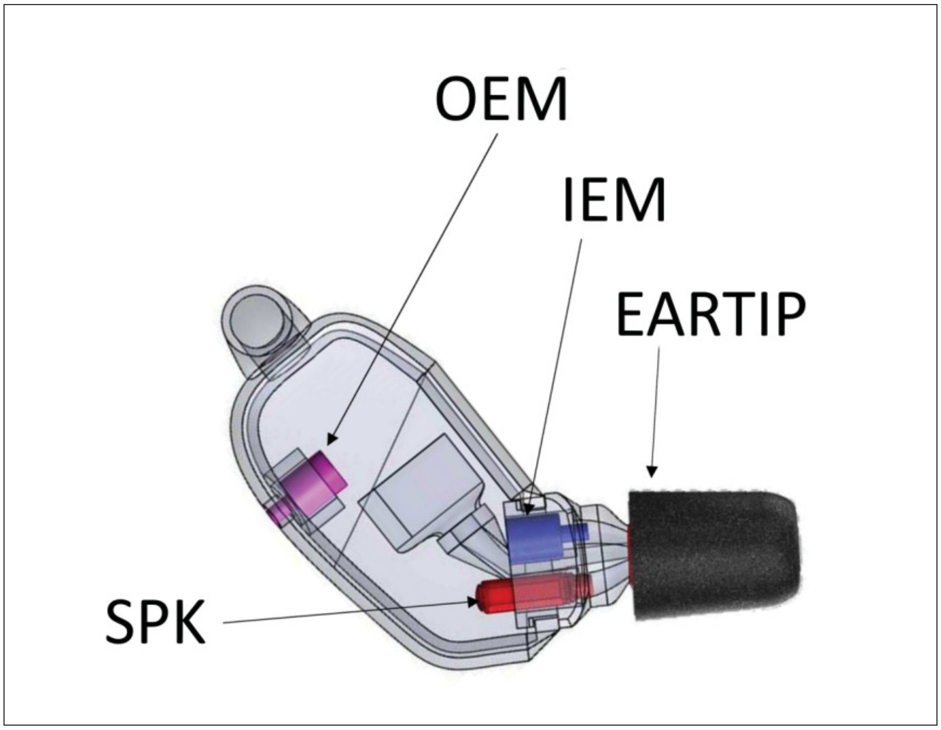


Figure 4.1 Diagram of an Auditory Research Platform earpiece, containing an outer-ear microphone (OEM), an in-ear microphone (IEM), an internal miniature loudspeaker (SPK), and a high-attenuation foam eartip

4.4.1.3 Procedure

Once inclusion criteria were verified, participants were asked to enter a double-wall audiometric booth where the experiment took place, and general instructions about the experiment were given. Participants were then equipped with the Holter and ARP, which were respectively recording cardiac and audio signals simultaneously and continuously until the end of the experiment. The proper insertion of the ARP earpieces was checked with a fit-test (Voix & Laville, 2002). Then, participants began the stress measurement protocol, which consisted of three stress tasks in a pre-defined order (1. mental task, 2. noise exposure, and 3. cold pressor test) and 5-minute rest periods (sitting in silence) in between each stress task. Specific instructions about the tasks were given by the experimenter to participants directly before each task started.

Mental arithmetic task Mental arithmetic tasks have been widely used to experimentally induce acute stress (Cho *et al.*, 2019; Brindle *et al.*, 2016; Visnovcova *et al.*, 2014; Monaco, Cattaneo, Ortu, Constantinescu & Pietropaoli, 2017; Mishra *et al.*, 2020, 2018; Kirschbaum *et al.*, 1993). A stress task based on the Trier Social Stress Test (TSST) (Kirschbaum *et al.*, 1993) was adapted to be performed silently, since the participant's voice, amplified by the occluded earcanal (Bouserhal *et al.*, 2017), would have otherwise contaminated the in-ear heartbeat sounds recorded by the ARP. Moreover, speech is also known to modify the breathing pattern, and consequently, the heart rate (Brugnera *et al.*, 2018). The task was developed on MATLAB and displayed on a PC screen using PsychToolBox library functions. Participants were instructed to count down from 1022 with a decrease of 13 (hence reported values of 1009, 996, etc.) until the five-minute task was over. The socioevaluative component usually present in the TSST was replaced in this task by a visual timer limiting response time to 7.5 seconds. Additionally, participants received positive or negative feedback depending on the answer's accuracy. When a wrong or no answer was given, the participant had to restart the equations from the beginning (i.e., subtracting 13 from 1022). Regardless of how frequently the task was restarted, the total duration was limited to five minutes.

Noise task A broadband noise stimulus based on Ref. (Waye *et al.*, 2002) was generated with the Audacity audio editing software (www.audacityteam.org). This noise was chosen as it had previously been used on a population with noise sensitivity and tinnitus (Hébert & Lupien, 2009; Waye *et al.*, 2002). The stimulus magnitude was of rising intensity (ramped from 40 to 90 dBA) for 2 minutes and kept constant at 90 dBA for the remaining 3 minutes of the 5 minutes stimulus. The rising ramp was intended to be unpredictable for the participant, which is believed to induce stress (Bach *et al.*, 2008; Pagé *et al.*, 2021). The noise was sent through the AC40 audiometer, and played on two free field speakers inside the audiometric booth, oriented at 45° on either side of the participant's head. This task was included in the protocol to evaluate its ability to induce stress in a separate analysis, since noise is thought to be implicated in triggering or modulating tinnitus (Baigi, Oden, Almlid-Larsen, Barrenäs & Holgers, 2011; Colagrosso, Fournier, Fitzpatrick & Hébert, 2019). However, as this task has not been well-validated as a stressor, it was not intended to serve as ground truth stress data. Consequently, recordings from this noise task were not used as labeled data for training the stress recognition system, as further described in 4.4.2.1.

Cold Pressor Test During the cold pressor test, a standard pain task commonly used in pain and stress research (Mishra *et al.*, 2020, 2018; Mourot, Bouhaddi & Regnard, 2009; Hasson *et al.*, 2013), participants immersed one hand in cold water for 3 minutes. The water was contained in a cooler and the temperature was kept at 6.5°C thanks to a thermal controller of 0.1°C-sensitivity. A recirculating pump was installed inside the cooler to ensure water-circulation and thus preventing local warming of water temperature around the immersed hand.

4.4.2 Preprocessing

4.4.2.1 Task segmentation

The time points corresponding to each step of the experimental protocol were labeled by the experimenter using the Audacity software. In order to have a uniform segment duration across tasks and participants, for each task and rest period, a single three-minute segment was chosen. While the recommended segment duration for short-term HRV analysis is five minutes (Malik *et al.*, 1996; Shaffer & Ginsberg, 2017; Giannakakis *et al.*, 2019a; Laborde *et al.*, 2017), some HRV features have been estimated with shorter durations (Shaffer & Ginsberg, 2017; Laborde *et al.*, 2017). Furthermore, shorter segment durations are commonly used for stress prediction, in particular when real-time analysis is envisioned (Giannakakis *et al.*, 2019b; Momeni, Arza, Rodrigues, Sandi & Atienza, 2021; Jongyoon Choi *et al.*, 2012; Hovsepian *et al.*, 2015; Akmandor & Jha, 2017; Suzuki, Laohakangvalvit, Matsubara & Sugaya, 2021; Mishra *et al.*, 2020; Sun *et al.*, 2012; Kaur *et al.*, 2014).

The three-minute segment used for the rest period was chosen to maximize the time between steps in the protocol (i.e. when the last task ended and when preparation for the next task began). Though the participants were at rest for longer than three minutes, discarding the data from the start and end of the rest period accounted for recovery from the previous task and anticipation of the next task, as “residual” stress induced by experimental tasks may linger during the rest period (Mishra *et al.*, 2018, 2020). The segment used for the tasks started 30 seconds before and ended 150 seconds after the start of each task, in order to include data recorded while participants were anticipating the task (which can be a stressor on its own (Pulopulos, Vanderhasselt & De Raedt, 2018)), as well as during the task itself.

For each participant, four three-minute segments were used to train and evaluate a binary stress classification model: two segments recorded during the mental arithmetic and cold pressor tasks were labeled as “stress”, and two segments recorded during the rest periods prior to these two tasks were labeled as “rest”. The segment recorded during the noise task was not used to train the stress classification model, as it has not been well-validated as a stress-induction task and may have been perceived differently by participants while they were wearing the earpiece. However, both the segments recorded prior to and during the noise task were used for synthesizing features with error, which is further explained in section 4.4.3.2. Additionally, the rest segment recorded prior to the noise task was used for baseline normalization, further explained in section 4.4.3.3.

4.4.2.2 Synchronization

The ECG and IEM data were synchronized as follows: at each step of the experimental protocol, an event marker button on the Holter monitor was pressed, triggering a sound that was recorded by the outer-ear microphones of the Auditory Research Platform and saving the corresponding timestamp on the Holter. The event marker sounds were labeled in Audacity by the experimenter. The Holter and Auditory Research Platform data were then aligned such that the mean difference between all Audacity labels and Holter event times was minimized. After synchronization, the maximum absolute difference between individual Audacity labels and Holter event times across all participants was 3.59 seconds. This range of synchronization error was considered acceptable for the current analysis which used three-minute long segments.

4.4.2.3 Heartbeat annotations

The ECG R-Peaks were visually inspected and corrected for artifact and ectopy using Burdick Vision Premier Holter analysis software (Cardiac Science-Quinton-Burdick, Bothell, WA, USA) and MATLAB. Only sinus beats were included in the computation of interbeat intervals. A small number of interbeat intervals were missing due to poor signal quality. However, as the total duration of missing intervals was never longer than 25% of the total segment length, none of the segments were excluded for having too much missing data (Cajal *et al.*, 2022).

The peaks corresponding to the first heartbeat sound in the IEM were automatically annotated following the methods described in Ref. (Martin & Voix, 2018). There was no manual correction of the IEM annotations, as some errors would be expected in a real-world monitoring context and it was of interest to assess how the stress classification would be affected by these errors.

4.4.3 Prediction

4.4.3.1 Feature extraction

HRV features were extracted using the Python toolbox Neurokit2 (Makowski *et al.*, 2021). All time-domain, frequency-domain, and non-linear features available in Neurokit2 were used, excluding features that had invalid values; for example the standard deviation of the average of normal-to-normal intervals over two minutes (SDANN2) requires at least three two-minute windows, which was not possible with the three-minute segment length used in this analysis. A description of all 73 features used can be found in the supplemental material.

There was no empirical selection of features using the current dataset, however, the stress classification performance was also evaluated using two smaller feature sets selected based on previous work. The first feature set consisted of only the median heart period (MedianNN), as the MedianNN was expected to be relatively robust to errors in the automatic heartbeat annotation and it has previously been used for stress recognition (Prajod & André, 2022; Arquilla *et al.*, 2022; Mishra *et al.*, 2020, 2018). The second feature set consisted of the MedianNN and the root-mean-square of successive differences (RMSSD). RMSSD is one of the most commonly used HRV features and is well-suited for shorter segment durations (Pham *et al.*, 2021a).

4.4.3.2 Error-balanced data synthesis

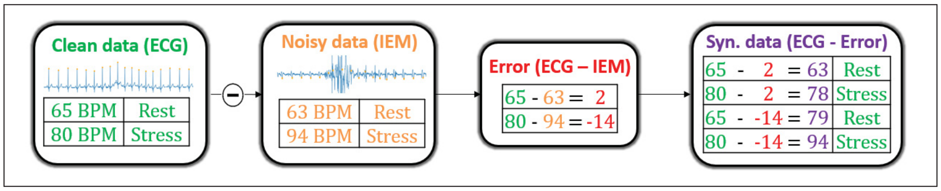


Figure 4.2 Diagram of synthetic data generated with error for one feature and two samples. Note that for illustrative purposes, heart rate expressed in beats per minute (BPM) is given as an example feature, though this was not actually a feature used in this study (the mean and median heart period were in milliseconds)

A variety of sounds beyond the heartbeat can be amplified in an occluded earcanal (Chabot *et al.*, 2021). Some of these sounds, such as noise artifacts caused by the wearer moving or swallowing, may induce errors in the automatic extraction of heartbeats from in-ear audio (Martin & Voix, 2018). Heartbeat signal artifacts can distort heartbeat variability features, affecting the predictions of classifiers trained on these features (Nikolic-Popovic & Goubran, 2013).

Previous research using another type of wearable sensor to monitor stress (a wristband photoplethysmogram) has found that performing experimental tasks, regardless of whether they are intended to induce stress, can affect signal quality and reduce the number of accurately detected heartbeats compared to rest recordings (Ollander *et al.*, 2016). In this work, the only "non-stress" class corresponded to the rest recordings, during which the participants were asked to sit in silence and do nothing rather than to perform an experimental task. Therefore, there is a risk that a classification system could learn task-related artifacts as a way to discriminate between rest and stress conditions, rather than physiologically meaningful differences between these conditions (for an empirical justification of this claim see Section 4.5.2).

To address this risk, a synthetic dataset containing equal errors across classes was generated, such that the learned classifier could be robust to the errors caused by artifacts in the IEM data without using these errors as a way to discriminate between classes. The method used for the synthetic data generation is presented in Algorithm 4.1, and a diagram is shown in Fig. 4.2.

Data synthesis was applied to the current dataset as follows: First, the error was computed by subtracting each HRV feature extracted from IEM from each feature extracted from simultaneously recorded ECG. This process resulted in one vector of feature errors per sample. Synthetic data was separately generated for each participant, and therefore, the errors from the test set were not used in generating the training data (the classifiers were trained and tested on different groups of participants, as further explained in section 4.4.3.5). For the purposes of obtaining the error, the stress condition was considered to be irrelevant, and therefore, all six segments recorded during the three rest periods and the three experimental tasks (mental, noise and cold) were used to

Algorithm 4.1 NumPy-style pseudocode for error-balanced data synthesis

```
def get_error(X_ecg, X_iem):
    error = X_ecg - X_iem
    return error

def synthesize_with_error(X_ecg, error, y):
    # N: number of samples used to compute error
    # D: dimensionality of the feature vector
    N, D = error.shape
    # subtract all error vectors
    # from every ECG sample
    X_syn = (X_ecg[:,None,:] - error[None,:,:])
    X_syn = X_syn.reshape(-1, D)
    # generate synthetic labels
    # by repeating labels corresponding to ECG
    y_syn = y.repeat(N)
    return X_syn, y_syn
```

compute the feature errors. These six feature error vectors were then individually subtracted from the four ECG feature vectors to be used for stress prediction (corresponding to the two rest periods and the mental and cold tasks), resulting in 24 synthetic samples per participant.

4.4.3.3 Feature normalization

Normalization of heart rate variability features can be achieved at different steps in the prediction pipeline: the interbeat intervals can be normalized prior to HRV feature extraction (Mishra *et al.*, 2020; Hovsepian *et al.*, 2015), or the HRV features themselves can be normalized (Parent *et al.*, 2019; Giannakakis *et al.*, 2019b; Nardelli, Valenza, Greco, Lanata & Scilingo, 2015; Sun *et al.*, 2012; Ollander *et al.*, 2016; Nardelli *et al.*, 2015). As certain features are invariant to scaling of the interbeat intervals, such as the ratio of low frequency to high frequency power (LFHF), normalization was always implemented at the HRV feature level such that all features would be transformed by each normalization step. Furthermore, normalization can be applied within participants (Parent *et al.*, 2019; Giannakakis *et al.*, 2019b; Nardelli *et al.*, 2015; Sun *et al.*, 2012; Ollander *et al.*, 2016; Nardelli *et al.*, 2015), as well as using the statistics from participants in the training set (Nardelli *et al.*, 2015). In this work, two types of normalization were applied to the HRV features:

Baselining. Normalizing physiological signals according to each individual is a common feature transformation prior to stress prediction; however, it is unclear how baseline normalization would be implemented in real-time systems in which an individual's physiology would likely vary over longer periods of time (Mishra *et al.*, 2020). Therefore, one aim of this work was to evaluate the feasibility of stress prediction without normalizing using an individual's baseline. When baselining was applied, each feature value was divided by the mean of the value of all baseline feature values for that given participant (Parent *et al.*, 2019). Since the rest periods were used for the stress classification itself (as the non-stress class) and there were no separate baseline recordings, the baseline feature values were extracted from the rest periods prior to the two tasks besides the task at hand, such that these rest periods could essentially be treated as separate baseline recordings. For example, features from both the rest period prior to the cold task and the cold task itself were normalized with the features from the rest periods prior to the mental and noise tasks. The two other rest periods were used instead of the three rest periods or the entire dataset in order to simulate baselining with features from separate baseline recordings.

Z-scoring. As a separate normalization step, regardless of whether baseline normalization was applied, each feature was z-scored with StandardScaler in the Scikit-learn pipeline (Pedregosa *et al.*, 2011). The mean and standard deviation were computed based on the training data, that is, the samples corresponding to two rest and two stress tasks for all participants selected as the training set for the current iteration of the cross-validation (further explained in Section 4.4.3.5). All samples in both the training and test sets were z-scored using the mean and standard deviation derived from the training set. The z-scoring operation was the final feature transformation before classification, and it was applied in all experiments.

4.4.3.4 Classification algorithms

Prior work has proposed automatic stress recognition systems based on a wide variety of classification methods ranging in complexity and interpretability, including linear models, ensemble learning, and deep learning (Giannakakis *et al.*, 2019a). In this work, two classification algorithms were tested: logistic regression, implemented in Scikit-learn (Pedregosa *et al.*, 2011), and extreme gradient boosting (XGBoost), implemented in the XGBoost Python package (Chen & Guestrin, 2016). Logistic regression is a simple linear model that has been previously used to classify heartbeat data (Sadeghi, McDonald & Sasangohar, 2021; Jongyoon Choi *et al.*, 2012; Parent *et al.*, 2019; Masino *et al.*, 2019; Gupta *et al.*, 2021; Kaur *et al.*, 2014; Arquilla *et al.*, 2022), with the advantage of being intrinsically interpretable (Molnar, 2020). On the other hand, XGBoost is a more complex ensemble model, with previously demonstrated performance classifying various types of data including heartbeat data (Gupta *et al.*, 2021; Sadeghi *et al.*, 2021; Liew, Loo & Wermter, 2021; Momeni *et al.*, 2021; Guo *et al.*, 2020). The default hyperparameters were used for both of these algorithms, with the exception of setting the maximum number of iterations to 1000 for logistic regression. However, the results of supplementary analysis that was later conducted with optimized hyperparameters can be found in Appendix V. The code for the classification pipeline is available upon request from an online repository at https://critias.etsmtl.ca/the-technology/critias-db/db2022-in-ear-stress.

4.4.3.5 Evaluation

Different combinations of heartbeat signals, HRV features, feature transformations, and classifiers (depicted in Fig. 4.3) were tested in a series of experiments. In all experiments, each stress recognition system was evaluated with a repeated five-fold participant-wise cross-validation scheme, in order to test how the system would generalize to unseen individuals. Participants were randomly divided into five groups, and the samples from participants in four of the five groups were used to train the classifier, rotating which group was used to test the classifier. This procedure was repeated with 10 random seeds, resulting in 50 evaluations of each system tested (with the same 10 random seeds used for each system). Performance was assessed with the mean and standard deviation of the accuracy over these 50 evaluations.

The performance of the stress recognition systems was evaluated in several analyses guided by the following questions:

1. How does normalizing to each individual's baseline affect stress recognition performance?

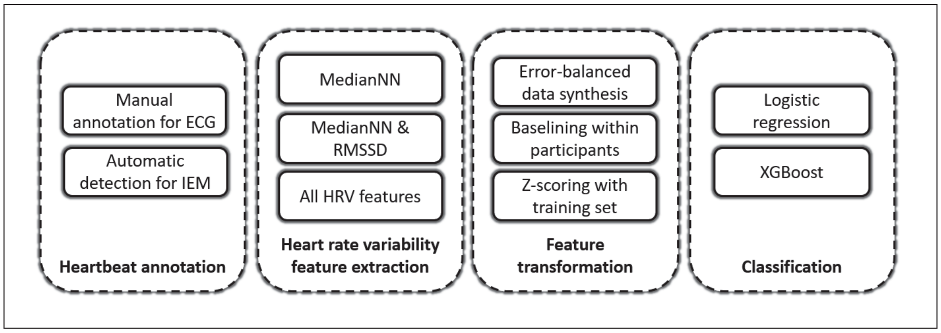

Manual annotation for ECG
Automatic detection for IEM
Heartbeat annotation
MedianNN
MedianNN & RMSSD
All HRV features
Heart rate variability feature extraction
Error-balanced data synthesis
Baselining within participants
Z-scoring with training set
Feature transformation
Logistic regression
XGBoost
Classification

Figure 4.3 Diagram depicting possible components of the prediction pipeline

2. How does the performance of stress recognition systems trained on in-ear audio compare to those trained on clinical-grade ECG?
3. How does performance change when using a synthetic dataset in which the feature error is balanced across rest and stress conditions?

4.5 Results

4.5.1 Baseline normalization

To assess the upper bound of stress classification performance with a heartbeat signal with baselining and without, classifiers trained and tested on the ECG were compared when the features were baselined vs. when their original values were used. Additionally, the two classifiers, as well as the three feature sets, were compared. The accuracy of classifiers trained on the ECG and tested on the ECG with baselining is presented in Table 4.1. With baselining, the highest mean accuracy achieved was 78.9% using all HRV features; however, reducing the feature set had a relatively small impact on performance: the highest mean accuracy using MedianNN & RMSSD was 76.7% and the highest mean accuracy using only MedianNN was 76.0%. In contrast, without baselining, using the two smaller feature sets led to a larger drop in performance: compared to 76.5% with all HRV features, the highest accuracy using MedianNN & RMSSD was 59.9%, and the highest accuracy using only MedianNN was 63.7%. The accuracy of all classifiers trained on the ECG and tested on the ECG without baselining is listed in Table 4.2.

Table 4.1 Accuracy of Stress Classifiers Trained on the ECG and Tested on the ECG with Baselining

Features	Classifier	Acc. (%)
MedianNN	Log. Reg.	76.0 ± 7.3
MedianNN	XGBoost	67.5 ± 8.4
MedianNN & RMSSD	Log. Reg.	76.7 ± 8.5
MedianNN & RMSSD	XGBoost	72.0 ± 9.4
All HRV features	Log. Reg.	**78.9 ± 8.1**
All HRV features	XGBoost	76.3 ± 7.8

Table 4.2 Accuracy of Stress Classifiers Trained on the ECG and Tested on the ECG without Baselining

Features	Classifier	Acc. (%)
MedianNN	Log. Reg.	57.2 ± 5.8
MedianNN	XGBoost	63.7 ± 7.1
MedianNN & RMSSD	Log. Reg.	59.5 ± 6.1
MedianNN & RMSSD	XGBoost	59.9 ± 7.9
All HRV features	Log. Reg.	**76.5 ± 8.4**
All HRV features	XGBoost	73.0 ± 8.5

4.5.2 Comparison of ECG and in-ear audio

After the feasibility of stress classification without baselining was successfully established with the ECG, it was of interest to evaluate whether a satisfactory performance could be achieved even when using a less reliable heartbeat signal, extracted from the IEM audio. To this end, the same combinations of feature sets and classifiers were trained and tested with HRV features extracted from the IEM data. The performance of these models is presented in Table 4.3. Surprisingly, the models trained and tested on the IEM had higher accuracy than the models trained and tested on ECG for the feature sets containing MedianNN & RMSSD and all HRV features. When training and testing on the IEM, the model with the highest mean accuracy without baselining was Logistic Regression using all HRV features: 81.1% (Table 4.3), compared to 76.5% (Table 4.2) when training and testing with ECG.

Table 4.3 Accuracy of Stress Classifiers Trained on the IEM and Tested on the IEM without Baselining

Features	Classifier	Acc. (%)
MedianNN	Log. Reg.	62.2 ± 6.1
MedianNN	XGBoost	58.7 ± 10.3
MedianNN & RMSSD	Log. Reg.	75.9 ± 6.5
MedianNN & RMSSD	XGBoost	79.1 ± 7.1
All HRV features	Log. Reg.	**81.1 ± 8.0**
All HRV features	XGBoost	76.1 ± 6.7

A substantial difference in accuracy between the classifiers trained and tested on the IEM vs. ECG was observed for the MedianNN & RMSSD feature set. For this feature set, the best-performing model trained and tested on the IEM had a mean accuracy of 79.1% (Table 4.3), while the best-performing model trained and tested on ECG had a mean accuracy of 59.9% (Table 4.2). With MedianNN as the only feature, the performance was more similar between the classifiers trained and tested on the IEM (62.2%; Table 4.3) vs. ECG (63.7%; Table 4.2).

To assess whether the models trained on HRV features extracted from the IEM could generalize to conventional HRV features, all models trained on the IEM were tested on the ECG. The classification performance is presented in Table 4.4. With the feature set consisting of only MedianNN, there was a relatively small reduction in mean accuracy when comparing the best-performing model trained on the IEM and tested on ECG (62.7%, Table 4.4) to the best-performing model both trained and tested on ECG (63.7%, Table 4.2). However, with the feature sets containing MedianNN & RMSSD and all HRV features, mean accuracy training on the IEM and testing on the ECG ranged from 50.0% to 50.8%.

The coefficients of the logistic regression models separately trained on the ECG and IEM, each using the MedianNN & RMSSD feature set, are shown in Fig. 4.4. As the feature values were z-scored with the mean and standard deviation of all the samples in the training set, the

Table 4.4 Accuracy of Stress Classifiers Trained on the IEM and Tested on the ECG without Baselining

Features	Classifier	Acc. (%)
MedianNN	Log. Reg.	57.8 ± 7.5
MedianNN	XGBoost	**62.7 ± 8.5**
MedianNN & RMSSD	Log. Reg.	50.8 ± 1.7
MedianNN & RMSSD	XGBoost	50.1 ± 0.6
All HRV features	Log. Reg.	50.0 ± 0.0
All HRV features	XGBoost	50.2 ± 0.8

magnitude of the coefficient for each feature was considered to be a measure of importance (Masino *et al.*, 2019). In terms of which of the two features was more important, the models trained on the two signals had opposite trends. In the case of the model trained on the ECG, MedianNN (mean coefficient of -0.8) had higher importance than RMSSD (mean coefficient of 0.3). In contrast, when training on the IEM, MedianNN (mean coefficient of -1.0) had lower importance than RMSSD (mean coefficient of 1.7). As demonstrated in Fig. 4.4, this pattern was consistent across cross-validation iterations.

Aiming to understand the source of the difference in feature importance, the error between HRV features extracted from ECG and from IEM was assessed. Unlike the manually-corrected ECG annotations, the heartbeats in the IEM data were annotated automatically (without manual correction of missed or extra beats) such that the IEM annotations would be representative of a real-world monitoring context. After manual review of segments of the in-ear audio where the HRV feature error was high, it became apparent that the IEM signal contained many artifacts such as movement and speech, despite the fact that participants were instructed to remain still and silent during the experimental tasks. An example of an IEM signal with artifacts and the corresponding heartbeat annotation is shown in Fig. 4.5. As some of these artifacts may have been related to the task (e.g. keyboard sounds during the computer-based mental task), the error was assessed separately for the rest and stress conditions.

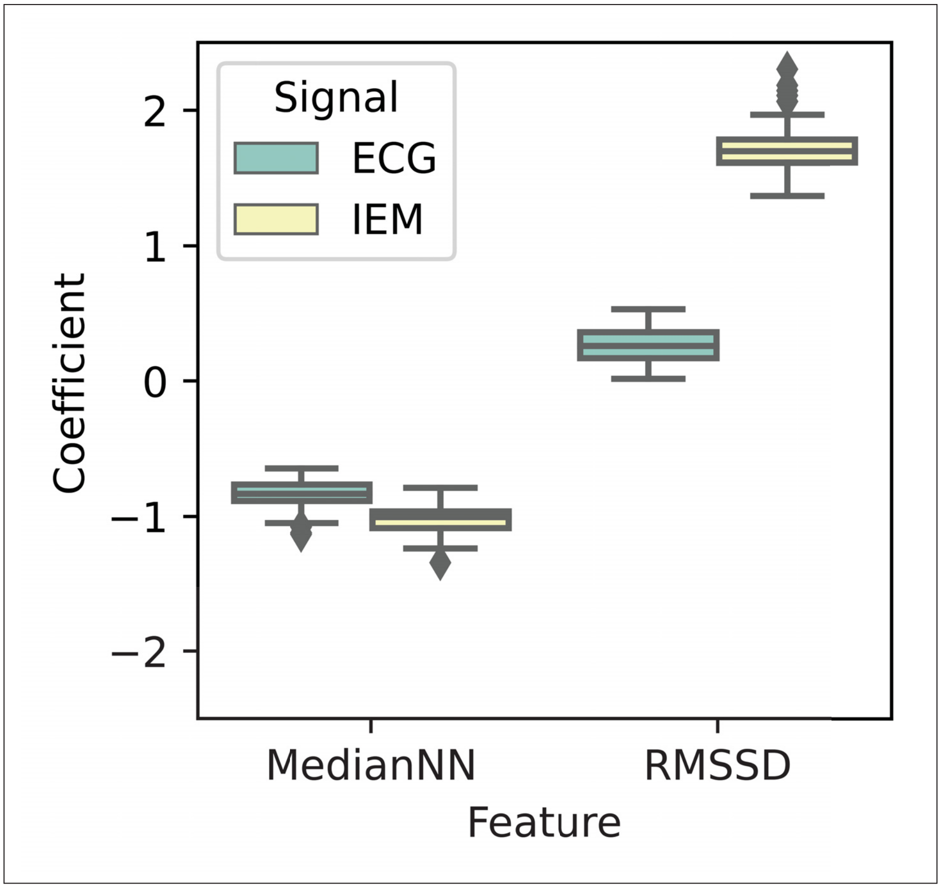


Figure 4.4 Box plot of feature coefficients in logistic regression models trained on the ECG and IEM. The coefficients were derived from models using the MedianNN & RMSSD feature set without baselining, trained over all cross-validation iterations. MedianNN had a higher absolute coefficient value than RMSSD in the model trained on the ECG, while RMSSD had a higher absolute coefficient value than MedianNN in the model trained on the IEM. Note that a negative coefficient indicates increasing probability of predicting stress when the standardized feature values are negative, while a positive coefficient indicates increasing probability of predicting stress when the standardized feature values are positive

The agreement between ECG and IEM for MedianNN and RMSSD was visually assessed with Bland-Altman plots, shown in Fig. 4.6. Visual inspection of the plots suggests that MedianNN extracted from IEM data in some cases overestimated and in other cases underestimated

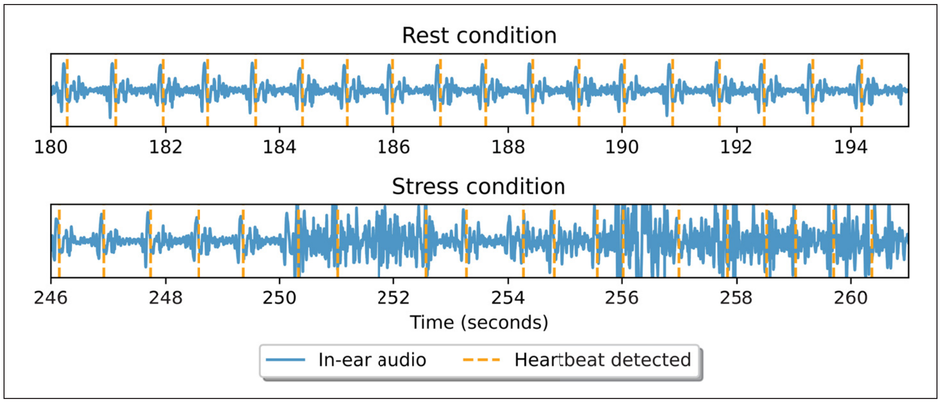
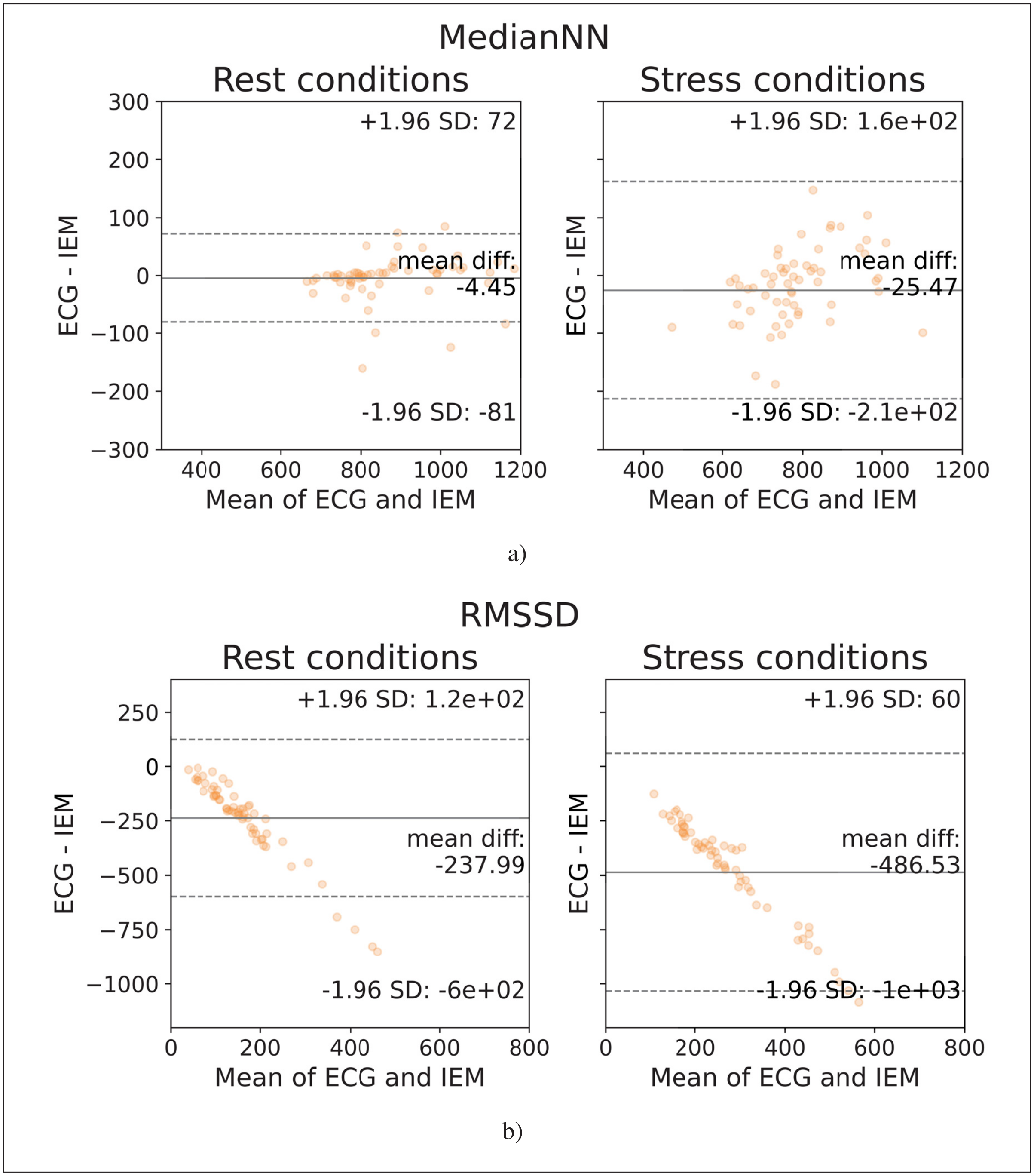


Figure 4.5 Example IEM recording and automatically detected heartbeats during a rest and stress condition. It can be seen that when there were artifacts in the IEM signal, the heartbeats were misdetected

MedianNN extracted from ECG data (though the mean difference was negative). In contrast, the plots demonstrate that RMSSD extracted from IEM data tended to consistently overestimate RMSSD extracted from ECG data, particularly in the stress conditions: the mean difference is -486.53 in stress conditions but is -237.99 in rest conditions. In both rest and stress conditions, it can be seen that as RMSSD extracted from IEM increasingly overestimated RMSSD extracted from ECG (shown on the y-axis), the mean of the RMSSD values extracted from ECG and IEM (shown on the x-axis) tended to increase, suggesting that the error greatly influenced the value of RMSSD for the IEM.



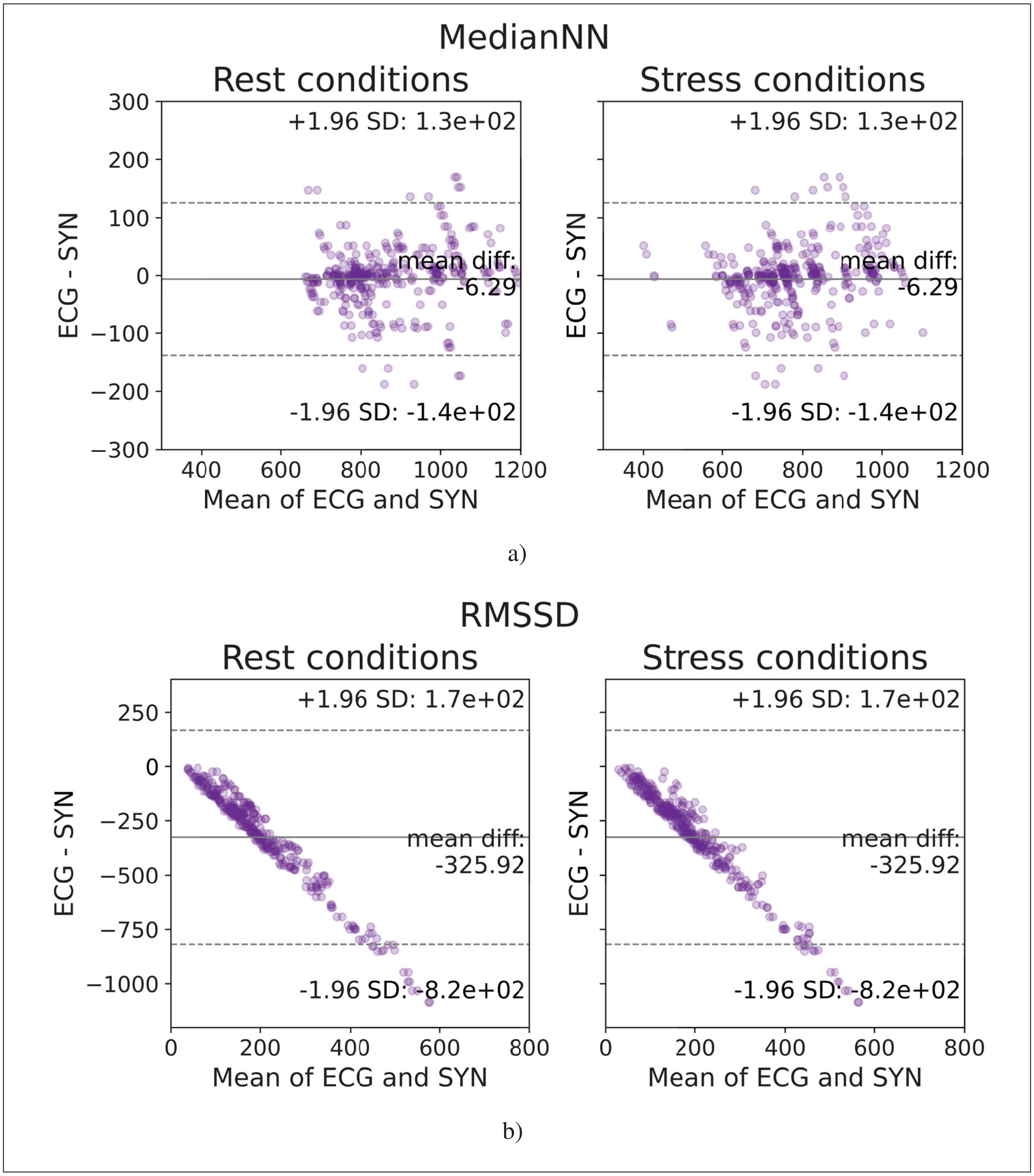
a)

b)

Figure 4.6 Bland-Altman plots assessing agreement between the electrocardiogram (ECG) and in-ear microphone (IEM) on the HRV features (a) MedianNN and (b) RMSSD, for data recorded during rest and stress conditions. Each point represents the difference between a HRV feature extracted from ECG and the same feature extracted from simultaneously recorded IEM data. The solid line in the middle represents the mean difference, while the upper and lower dashed lines represent the upper and lower limits of agreement at 95% confidence

4.5.3 Error-balanced data synthesis

After it was determined that the HRV feature error differed in the rest and stress conditions, synthetic data (SYN) was generated to balance the errors across classes, following the methods described in Section 4.4.3.2. The agreement between ECG and SYN for MedianNN and RMSSD was assessed with Bland-Altman plots, shown in Fig. 4.7. When comparing to the Bland-Altman plots assessing agreement between ECG and IEM (Fig. 4.6), it can be seen that the distribution of the synthetic errors between ECG and SYN in each of the rest and stress conditions is similar to the distribution of the original errors between ECG and IEM: SYN MedianNN sometimes overestimated and sometimes underestimated ECG MedianNN, while SYN RMSSD largely overestimated ECG RMSSD. However, unlike the original errors between ECG and IEM, the mean difference between ECG and SYN is identical across classes, for MediannNN (mean difference of -6.29) and RMSSD (mean difference of -325.92).



a)

b)

Figure 4.7 Bland-Altman plots assessing agreement between the electrocardiogram (ECG) and synthetic data (SYN) on the HRV features (a) MedianNN and (b) RMSSD, for data recorded during rest and stress conditions. Each point represents the difference between a HRV feature extracted from ECG and the same feature synthesized with one error value and the original ECG feature

All models trained on the IEM were tested on the SYN, in order to evaluate whether the models trained on HRV features extracted from the IEM would perform as well when tested on data where the overall distribution of the errors was similar to IEM but where these errors were balanced across classes (and therefore could not be used as a way to discriminate between rest and stress). The classification performance of the models trained on the IEM and tested on the SYN is presented in Table 4.5. With the feature set consisting of only MedianNN, the mean accuracy was similar when comparing the best-performing model trained on the IEM and tested on the SYN (61.1%, in Table 4.5) to the best-performing model both trained and tested on the IEM (62.2%, in Table 4.3). With the feature sets containing MedianNN & RMSSD and all HRV features, there was a larger drop in performance: mean accuracy training on the IEM and testing on the SYN ranged from 54.0% to 64.9% (see Table 4.5), compared to a mean accuracy ranging from 75.9% to 81.1% when both training and testing on the IEM (see Table 4.3).

Table 4.5 Accuracy of Stress Classifiers Trained on the IEM and Tested on the SYN without Baselining

Features	Classifier	Acc. (%)
MedianNN	Log. Reg.	61.1 ± 5.4
MedianNN	XGBoost	59.7 ± 4.7
MedianNN & RMSSD	Log. Reg.	59.1 ± 4.5
MedianNN & RMSSD	XGBoost	54.0 ± 2.3
All HRV features	Log. Reg.	**64.9 ± 5.3**
All HRV features	XGBoost	60.9 ± 5.0

As the performance of models trained on the IEM decreased when the error was balanced across classes compared to testing on the IEM, it was of interest to assess whether augmenting the training data with the SYN would allow for sufficient performance when testing on IEM without the possibility of learning class-dependent differences in errors. To determine the added benefit of augmenting the training data with errors, models were first trained on just the clean (manually-corrected) ECG and tested on the noisy (automatically-annotated) IEM. The classification performance of these models is presented in Table 4.6. The best-performing model, logistic regression trained on all HRV features had a mean accuracy of 65.4%, still well above chance accuracy but a substantial drop from the same model trained and tested on the ECG (76.5%, Table 4.2). The same models were then trained on a dataset consisting of both ECG and SYN, and tested separately on ECG, IEM, and SYN to evaluate how augmenting the dataset would affect generalization to features extracted from clean data without error, real data with class-imbalanced errors, and synthetic data with class-balanced errors. The performance of the models trained using the best-performing feature set, all HRV features, is shown in Table 4.7. XGBoost had the highest mean accuracy for all test signals, ranging from 76.6% to 77.3%.

Table 4.6 Accuracy of Stress Classifiers Trained on the ECG and Tested on the IEM without Baselining

Features	Classifier	Acc. (%)
MedianNN	Log. Reg.	61.3 ± 6.3
MedianNN	XGBoost	60.8 ± 9.5
MedianNN & RMSSD	Log. Reg.	54.2 ± 4.9
MedianNN & RMSSD	XGBoost	56.0 ± 9.2
All HRV features	Log. Reg.	**65.4 ± 9.3**
All HRV features	XGBoost	60.7 ± 7.3

Table 4.7 Accuracy of Stress Classifiers Trained on the ECG & SYN Using All HRV Features without Baselining

Test signal	Classifier	Acc. (%)
ECG	Log. Reg.	74.3 ± 8.1
ECG	XGBoost	**76.7 ± 7.9**
IEM	Log. Reg.	69.9 ± 9.0
IEM	XGBoost	**76.6 ± 7.0**
SYN	Log. Reg.	70.6 ± 5.8
SYN	XGBoost	**77.3 ± 4.7**

4.6 Discussion

4.6.1 Baseline normalization

The aim of this work was to determine the feasibility of stress monitoring with an in-ear wearable device. As normalization for each individual's baseline may be difficult to implement in a real-world monitoring context (Mishra *et al.*, 2020), the effect of baselining on classification performance was assessed. It was found that normalizing to each individual's baseline improved classification results, in line with previous research using HRV (Nardelli *et al.*, 2015). However, using all HRV features allowed for nearly as high performance as without baselining, suggesting that concerns about the ecological validity of the baselining procedure could be avoided by including these additional features.

Nonetheless, fewer features may be preferred for certain applications, as the number of features plays a major role in the interpretability of a model (Poursabzi-Sangdeh, Goldstein, Hofman, Wortman Vaughan & Wallach, 2021), which may be especially important in clinical contexts (Kelly, Karthikesalingam, Suleyman, Corrado & King, 2019) such as treatment or diagnosis of tinnitus and decreased sound tolerance. Moreover, fewer features may be preferred to reduce computational cost (Momeni *et al.*, 2021; Suzuki *et al.*, 2021), as hearables, like other wearables, generally have limited computational power. However, it may be possible that a smaller subset of the 73 HRV features used in this work could achieve similar performance without baselining, as many HRV features are correlated with each other and influenced by the same physiological phenomena (Shaffer & Ginsberg, 2017; Pham, Lau, Chen & Makowski, 2021b). Future efforts could aim to identify this smaller subset of features, ideally with a separate dataset for testing (Mishra *et al.*, 2020), as empirically selecting this subset from all the subsets possible given 73 features would require many comparisons, and therefore, there is an increased risk that improvements in mean cross-validation accuracy would not translate to better performance on completely unseen data (Hosseini *et al.*, 2020).

4.6.2 Comparison of ECG and in-ear audio

A subsequent set of experiments compared stress classification systems trained on features extracted from the clean ECG signal to those trained on the features extracted from the IEM audio signal, to determine whether automatically-detected heartbeats in the IEM could estimate HRV well enough to predict stress. Unexpectedly, the best-performing classifier trained on the IEM performed better than the best-performing classifier trained on the manually-annotated, clinical-grade ECG. For models trained on just two features, MedianNN and RMSSD, training and testing with the IEM instead of the ECG improved the mean accuracy by 16 to 19% depending on the classifier used (compare Tables 4.2 and 4.3). However, analyses of the feature importance and error suggest that the improved performance of the models trained on the IEM was not due to learning physiologically meaningful features: RMSSD, the most important feature for one model trained on the IEM, was highly distorted by misdetected heartbeats and had a higher error during the stress conditions than during the rest conditions. This difference in error between conditions may have been due to artifacts more commonly found during the stress tasks such as the presence of movement, speech signals or keyboard typing noise, as these can increase the number of misdetected heartbeats.

Spurious correlations have been known to be an issue in deploying machine learning algorithms across different domains. Biases in the data used to train and evaluate a model can lead to large drops in performance once deployed (Kelly *et al.*, 2019). For example, a pneumonia detection algorithm that appeared to perform well when tested on data from the same hospital did not generalize well to new hospitals, likely because the algorithm exploited confounding information specific to the hospital such as the type of scanner used in departments with different rates of pneumonia (Zech *et al.*, 2018). In the context of stress classification with heartbeat signals, it is also possible that spurious information in the training data can lead to large drops in classification performance when applied in unseen contexts, particularly when the input to a data-driven system is not limited to physiologically meaningful features based on domain knowledge: in one study, deep learning models directly trained on ECG signals appeared to perform better than models trained on HRV features when tested on held-out data from the same dataset used for training, but the HRV feature-based models outperformed the ECG signal-based models when tested on a different dataset (Prajod & André, 2022). The results of the current study suggest that even HRV features can be affected by spurious information if errors are not accounted for, and therefore, the cross-validation performance obtained by training and testing on HRV features extracted from the IEM likely does not reflect how well the model would perform when deployed. If the increased performance for the stress prediction is driven by the increased presence of artifacts during the stress tasks, the learned classifier would not generalize to real-world applications where artifacts such as movement, speech, and keyboard typing may occur when the wearer is not stressed. More broadly, these findings highlight the risk of artifacts unknowingly being exploited by predictive models and could have implications for any applications in which raw or automatically processed biosignals are used as input to a classifier without controlling for the presence of artifacts.

4.6.3 Error-balanced data synthesis

When the models trained on the IEM were evaluated with synthetic data in which the HRV feature errors were equal across classes, performance dropped substantially, further supporting the claim that the models trained on the IEM had relied on this difference in HRV feature error to discriminate between rest and stress. The influence of spurious patterns on predictions has been previously identified and reduced with synthetic data in other domains (Chang, Adam & Goldenberg, 2021; Agarwal, Shetty & Fritz, 2020; Zhao, Wang, Yatskar, Ordonez & Chang, 2018). In the context of the natural language processing task of coreference resolution, imbalanced co-occurrence of gendered pronouns with certain occupations has been corrected for by generating new sentences with different pronouns (e.g. swapping “he” for “she” in “The physician hired the secretary because he was overwhelmed with clients.”). These modified sentences have revealed gender biases in data-driven coreference systems and shown to be an effective way to prevent biases when used as training data (Zhao *et al.*, 2018). The current study extended the existing research on robustness to spurious patterns by proposing a data synthesis method to counter this issue for systems that learn from biosignals that may be corrupted by artifacts. While the analyses presented in this paper focused on the synthesis of HRV features, the proposed method could be applied to any dataset containing both physiologically meaningful “clean” features and “noisy” features from which errors can be computed (e.g. breathing rate derived from a conventional respiration sensor and breathing rate automatically extracted from in-ear audio signals (Martin & Voix, 2018)).

In addition to its utility to evaluate existing models for their reliance on spurious patterns, the synthesized data also proved useful for training: compared to training on only the clean ECG features, augmenting the dataset with the error substantially improved test performance on the IEM. Data augmentation methods that synthesize additional samples with noise are often employed to increase the number of samples, to improve performance of machine learning methods that tend to rely on a large amount of training data, as well as to make the system more robust to noise (Yu & Sano, 2022). It may be possible to further improve the system’s robustness to noise at other steps in the prediction pipeline. When interpretability is crucial, it may be preferable to identify unreliable segments of the heartbeat signal before HRV feature extraction rather than have a classifier robust to errors in the HRV features. That said, if there are many of these unreliable segments, it may not be possible to have a long enough recording for HRV feature extraction without treating unreliable segments as missing data, and methods to treat missing data such as interpolation could still introduce errors in the HRV features (Bernal, Montgomery & Maes, 2021; Ollander *et al.*, 2016).

One possible limitation in the theoretical assumptions made in this work should be noted: by using all of the errors to generate new data for both rest and stress conditions regardless of which condition the original errors came from, the proposed data synthesis method assumed that the difference between the ECG and the IEM should be completely independent of the stress condition. In reality, it is possible that the difference between the ECG and the IEM could also be influenced by factors related to the stress condition to some extent; for example, even without errors in the heartbeat annotations, there are small physiological differences in the interbeat intervals computed from different heartbeat signals, and these differences may constitute novel biomarkers (Yuda *et al.*, 2020). Determining the extent to which physiological differences influence the error would be an interesting avenue for future research. However, the fact that the test performance was high for both IEM and SYN suggests that simply assuming that the error is completely independent of the stress condition is sufficient for the purposes of the data augmentation, that is, to achieve high performance on IEM data containing artifacts without unintentionally learning the artifacts as a way to discriminate between rest and stress.

Finally, future work could investigate the utility of data synthesis with separate datasets. In the current study, the same dataset was used to estimate the error and obtain the clean features that were augmented with the error. However, it is possible that the errors and the clean features could be derived from separate sources: the distribution of the feature errors could be modeled based on any dataset containing simultaneous clean and noisy features, and this distribution could be sampled from to augment a separate dataset containing only clean features recorded under the conditions of interest (e.g. stress tasks). The ability to augment separate datasets could save substantial resources associated with data collection, as it would open the possibility of developing algorithms for hearables with datasets that were not collected with a hearable in mind, only requiring in-ear audio signals to validate that the developed algorithms are robust to real errors.

4.7 Conclusions

This work demonstrated the feasibility of a novel modality of stress monitoring based on heartbeat sounds recorded from inside the human earcanal, which could lead to improved research methods and therapies by extending the capabilities of existing hearing technology for tinnitus and decreased sound tolerance. Moreover, the findings of this work highlighted the possibility that artifacts in biosignal data can be exploited by a classifier to achieve high performance that likely would not generalize to a real-world monitoring context, with implications for the broader field of automatic biosignal monitoring.

4.8 Acknowledgments

The authors would like to thank Ajin Tom and Philippe Chabot for their feedback on the initial analyses, Omar Mahiddine for his expertise on the Holter and for the ECG signal inspection, and Valentin Pintat, Corentin Delain, and Frédérik Desaulniers for data acquisition setup and the initial measurements. The authors would like to acknowledge the support of the Natural Sciences and Engineering Council of Canada (NSERC), EERS Global Technologies Inc. and the ÉTS-EERS Industrial Research Chair in In-Ear Technologies (CRITIAS) through the Alliance Grant (ALLRP 566678 - 21). This work was also made possible via funding from the FRQ AUDACE program (2020-AUDC-267674). The first author acknowledges financial support received from the German Academic Exchange Service (DAAD) and MITACS Accelerate Cluster (IT26677/SUBV-2021-168).

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Potential applications

The methods and tools presented in this work could be applied to improve current approaches to study and manage decreased sound tolerance:

1. **More accessible biofeedback.** Heart rate variability is one of the most common signals used in biofeedback, which may be useful in managing stress associated with decreased sound tolerance. Existing biofeedback systems present representations of changes in biosignals or more abstract representations of psychological states such as stress with different sensory modalities, including only-auditory systems (Yu, Funk, Hu, Wang & Feijs, 2018). Rather than requiring separate dedicated devices for biofeedback, a hearing device capable of stress monitoring could provide auditory biofeedback through the earpieces.
2. **Adapting noise exposure based on stress.** A hearing device administering sound therapy could adapt the level of the stimuli according to the wearer's stress level. Additionally, to reduce over-avoidance associated with the regular use of hearing devices to cope with decreased sound tolerance, a hearing device could adapt exposure to external sounds such that these external sounds are only attenuated when deemed necessary based on the wearer's stress. Noise exposure could be adapted by dynamically changing the parameters of existing algorithms used in hearing devices for decreased sound tolerance such as dynamic range limiting, or potentially via machine learning models personalized to the wearer that learn to selectively remove the specific sounds causing stress over time given simultaneous data on external sounds and stress.
3. **Recording sounds associated with stress.** A device monitoring both external sounds and the wearer's stress level could be used to create a database of sounds that induce stress. These recordings could have multiple uses: for sound therapies relying on recordings of real-life sounds, stimuli are ideally personalized to the individual (Schröder *et al.*, 2017) and ecologically valid (Frank & McKay, 2019). Moreover, these sounds could be used to build a large, comprehensive database with numerous research applications such as for

stimuli in neuropsychological studies, training data for audio processing algorithms, and acoustic analyses to better understand the nature of the sounds that are intolerable and how this varies across individuals.

4. **Assessing how device use relates to noise exposure and stress.** Studying how hearing devices are used in real-world contexts over a longer period of time can offer unique insights on individual usage patterns (Johansen *et al.*, 2017), and psychophysiological measures could provide valuable data on sound-induced distress, particularly in populations with communication challenges (Fodstad *et al.*, 2021). A hearing device simultaneously recording wearer stress level, external sounds, and logs of how the device is used could inform which device features are most useful in reducing sound-induced distress over time.

Moreover, the findings of this work have the potential to advance hearing technology and research beyond decreased sound tolerance. For example, tinnitus is often managed with hearing devices (Searchfield *et al.*, 2021) and is linked to stress (Hébert *et al.*, 2012), and therefore, integrating stress monitoring into these devices could improve treatments and research in this field as well. More broadly, even in the general population, noise exposure and stress are related and are associated with various negative outcomes such as changes in cardiovascular function and reductions in subjective measures of well-being (Lusk, Gillespie, Hagerty & Ziemba, 2004; Sander, Marques, Birt, Stead & Baumann, 2021). As headphones are becoming ubiquitous, these devices have been used for collecting research data on a large scale on noise exposure (Smith *et al.*, 2020). The addition of stress monitoring could allow for new research methods and devices that more effectively reduce noise-induced stress.

5.2 Limitations, considerations and recommendations

This work presented tools to prototype a hearing device for decreased sound tolerance and established the feasibility of stress monitoring with in-ear heartbeat sounds. However, more work needs to be done to assess the utility of stress monitoring in a hearing device as a research tool and management strategy in practice.

Although the stress recognition system presented in Chapter 4 was developed with people with decreased sound tolerance in mind, the system has only been validated in a non-clinical population. There remains a need to assess the utility of the developed system for monitoring the stress associated with decreased sound tolerance. One avenue for this research could be to apply the data augmentation method introduced in Chapter 4 to existing datasets containing conventional heartbeat signals (e.g. ECG rather than in-ear audio) collected from people with decreased sound tolerance. Furthermore, the stress recognition system was only validated using post-processed data collected in a controlled setting. The processing pipeline presented in Chapter 2 could serve as a base for a real-time implementation, allowing an evaluation of the stress recognition system for ambulatory monitoring.

The aforementioned processing pipeline was developed in order to decouple the latency of stress recognition from the latency of audio processing, as the latency requirements for real-time stress recognition are generally in the order of minutes, while the requirements for audio processing are in the order of milliseconds. However, the exact maximum acceptable latency of audio processing for the current application remains unclear, posing a challenge for defining the requirements of novel audio processing algorithms, particularly those based on machine learning. The findings presented in Chapter 3 demonstrated that the effect of the delay condition on reaction time differed depending on the position of the target word in the sentence, suggesting that latency can affect speech comprehension at higher levels of processing. Therefore, previous work concentrating on nonverbal sounds or single syllables/words may not translate to how increased latency would affect the comprehension of continuous speech. The current study was limited in that main effects of the delay condition on reaction time could not be interpreted as solely reflecting cognitive load, as the participants may have also waited for the delayed auditory stimulus. Future studies aiming to assess the effect of latency on speech comprehension could include additional measures of cognitive load, such as accuracy on a more difficult task or physiological measures of mental effort like pupillometry. Furthermore, while it is currently unclear whether higher latency audio processing algorithms would be usable in practice, there are many applications that would not require novel audio processing algorithms, such as continuous monitoring to collect research data or the adaptation of existing audio processing algorithms already shown to be usable in practice when implemented on hearing devices to manage decreased sound tolerance (e.g. dynamic range limiting).

5.3 Contributions

5.3.1 Industrial

The industrial contributions of this work include the following:

1. During a MITACS Accelerate internship with the industrial partner, EERS Global Technologies, the limitations of stress recognition based on in-ear heartbeat sounds were characterized, taking into account the envisioned real-world monitoring application.
2. A framework was developed during the internship to evaluate stress recognition systems on multiple data sets.
3. Methods were developed to improve the robustness of the stress recognition system based on in-ear audio, for which a Declaration of Invention (VAL-361) has been filed.

5.3.2 Scientific

The methods and tools presented in this thesis provided a base for the development of a hearing device capable of monitoring stress associated with decreased sound tolerance, which could also have scientific implications beyond the envisioned application. Chapter 2 presented a generalized processing pipeline to allow for the prototyping of novel algorithms on an external computer to change the parameters of any audio processing running on a hearing device. Chapter 3 laid the groundwork for future protocols to assess the impact of audio processing delays on language understanding, as well as demonstrated the complexity of continuous speech comprehension. Chapter 4 demonstrated the viability of automatic stress recognition with in-ear heartbeat sounds and presented a new method of data augmentation to prevent the stress recognition system from exploiting differences in feature error across classes, which could be applied to improve robustness of other machine learning models trained on automatically processed biosignals.

In addition to the articles included as chapters of this thesis, numerous scientific contributions were made during the course of the research project including:

1. This work was presented at several conferences (Benesch, Raj & Voix, 2021d; Benesch, Bouserhal & Voix, 2021a; Benesch, Delain, Blain-Moraes & Voix, 2020; Benesch, Bouserhal, Blain-Moraes & Voix, 2021b).
2. To accelerate research in the field of in-ear biosignal monitoring, a dataset containing in-ear audio recorded while participants performed tasks intended to induce different emotional states will be made open access (for more information please see Appendix VI). This dataset contains both pilot data collected in the initial stages of the research project, as well as data collected for a MITACS internship with the industrial partner to validate the automatic stress recognition system proposed in this work.
3. During the development of the biosignal processing routines used in this work, multiple contributions were made to the open-source Python neurophysiological signal processing library Neurokit2 (Makowski *et al.*, 2021): adding features, fixing bugs, improving documentation, and reviewing code[1].
4. To increase the diversity of sounds available for decreased sound tolerance research, an open-access database was curated and published on Zenodo (Benesch, Orloff & Hansen, 2022a), funded by a soQuiet Misophonia Student Research Award and publicly presented online (Orloff, Benesch & Hansen, 2022).

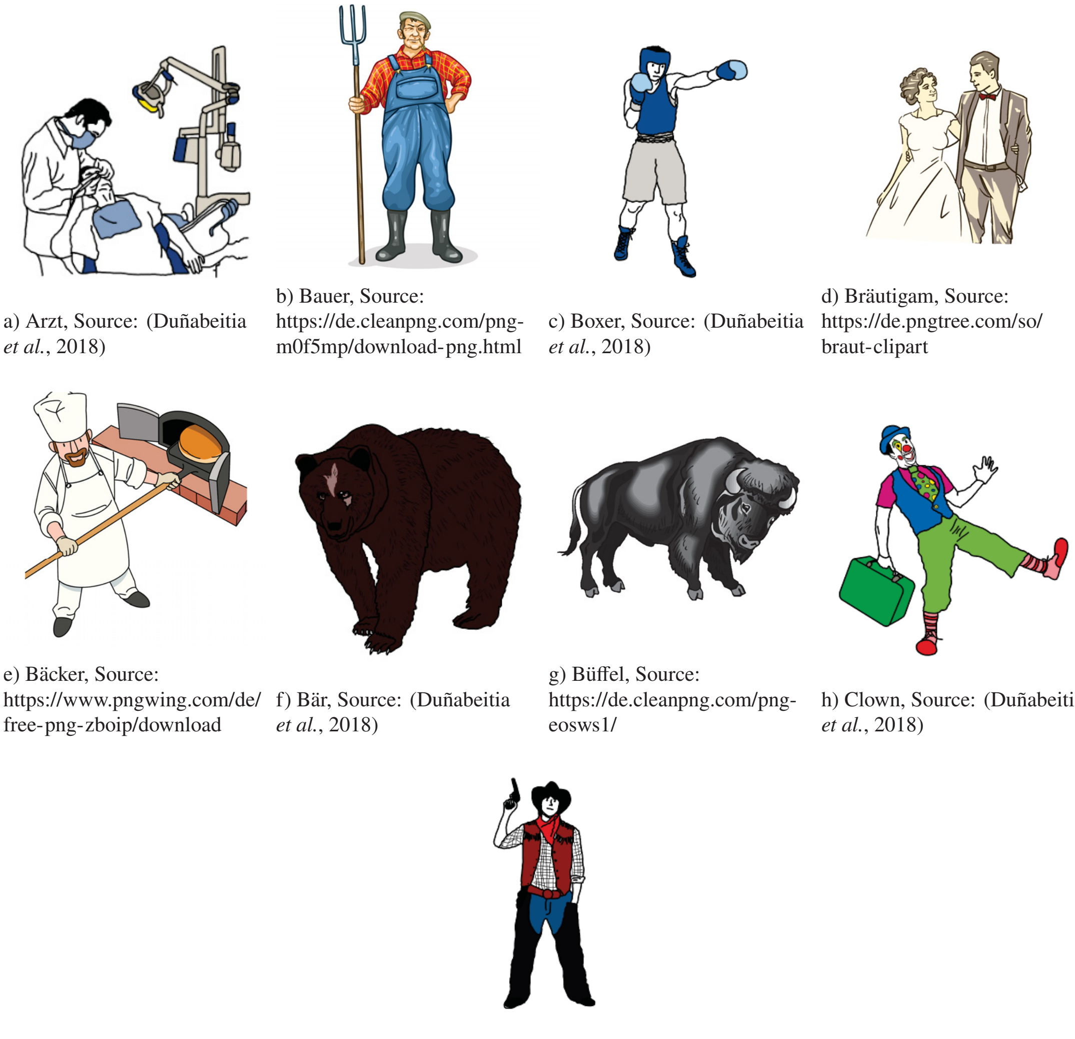
[1] A list of the maintainers and contributors of Neurokit2 as of September 7, 2022 is available at the following link: https://github.com/neuropsychology/NeuroKit/blob/5cd089e9d82e6f7fb98d2028e69840aee08f8ab3/AUTHORS.rst

APPENDIX I

LIST OF IMAGE STIMULI FROM CHAPTER 3

The following list consists of all image stimuli included in this study, the noun they were intended to represent (in German), and their source.

Figure-A I-1 Images used to represent target nouns



a) Arzt, Source: (Duñabeitia *et al.*, 2018)

b) Bauer, Source: https://de.cleanpng.com/png-m0f5mp/download-png.html

c) Boxer, Source: (Duñabeitia *et al.*, 2018)

d) Bräutigam, Source: https://de.pngtree.com/so/braut-clipart

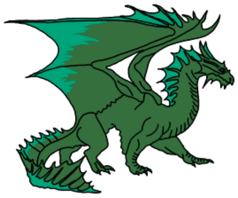
e) Bäcker, Source: https://www.pngwing.com/de/free-png-zboip/download

f) Bär, Source: (Duñabeitia *et al.*, 2018)

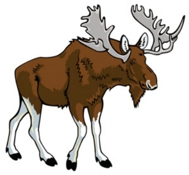
g) Büffel, Source: https://de.cleanpng.com/png-eosws1/

h) Clown, Source: (Duñabeitia *et al.*, 2018)

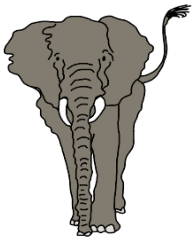
i) Cowboy, Source: (Duñabeitia *et al.*, 2018)



j) Drache, Source: (Duñabeitia *et al.*, 2018)



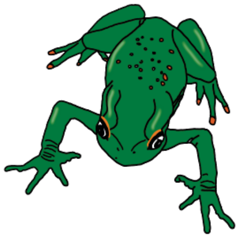
k) Elch, Source: https://www.pngaaa.com/detail/2563273



l) Elefant, Source: (Duñabeitia *et al.*, 2018)

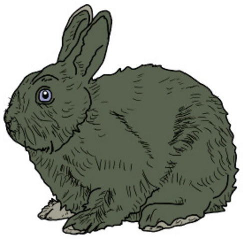
m) Esel, Source: (Duñabeitia *et al.*, 2018)

n) Frisör, Source: (Duñabeitia *et al.*, 2018)

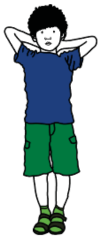


o) Frosch, Source: (Duñabeitia *et al.*, 2018)

p) Gärtner, Source: (Duñabeitia *et al.*, 2018)



q) Hase, Source: (Duñabeitia *et al.*, 2018)



r) Junge, Source: (Duñabeitia *et al.*, 2018)

s) Jäger, Source: (Duñabeitia *et al.*, 2018)

t) Kapitän, Source: (Duñabeitia *et al.*, 2018)

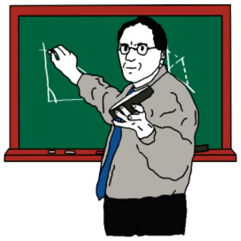
u) Kasper, Source: (Duñabeitia *et al.*, 2018)

v) Koala, Source: (Duñabeitia *et al.*, 2018)

w) Kobold, Source: https://de.cleanpng.com/png-85z9z8/download-png.html

x) Koch, Source: https://pngtree.com/so/chef-hat-clipart

y) König, Source: (Duñabeitia *et al.*, 2018)



z) Lehrer, Source: (Duñabeitia *et al.*, 2018)

aa) Löwe, Source: (Duñabeitia *et al.*, 2018)

ab) Maler, Source: (Duñabeitia *et al.*, 2018)

ac) Mann, Source: (Duñabeitia *et al.*, 2018)

ad) Matrose, Source: https://www.nicepng.com/ourpic/u2w7q8a9q8w7w7u2_navy-drawing-sailor-line-art-soldier-navy-drawing/

ae) Maulwurf, Source: https://www.pngwing.com/de/free-png-tuucp

af) Metzger, Source: (Duñabeitia *et al.*, 2018)

ag) Mönch, Source: https://de.cleanpng.com/png-zsofj1/download-png.html

ah) Nikolaus, Source: https://nikolaus-von-myra.de/de/darstellung/galerie/

ai) Panda, Source: (Duñabeitia *et al.*, 2018)

aj) Papagei, Source: (Duñabeitia *et al.*, 2018)

ak) Papst, Source: (Duñabeitia *et al.*, 2018)

al) Pfarrer, Source: (Duñabeitia *et al.*, 2018)

am) Pilot, Source: (Duñabeitia *et al.*, 2018)

an) Pinguin, Source: (Duñabeitia *et al.*, 2018)

ao) Jäger, Source: (Duñabeitia *et al.*, 2018)

ap) Polizist, Source: (Duñabeitia *et al.*, 2018)

aq) Postbote, Source: (Duñabeitia *et al.*, 2018)

ar) Prinz, Source:https://www.pngegg.com/de/png-iptbx

as) Punker, Source: https://www.vippng.com/preview/hRbTRJR_punk-tribu-urbana-vestimenta-punks/

at) Radfahrer, Source: https://de.pngtree.com/so/die-jungs

au) Riese, Source: https://www.pinterest.de/pin/766386061570400292

av) Ritter, Source: (Duñabeitia *et al.*, 2018)

aw) Roboter, Source: (Duñabeitia *et al.*, 2018)

ax) Räuber, Source: (Duñabeitia *et al.*, 2018)

ay) Soldat, Source: (Duñabeitia *et al.*, 2018)

az) Tiger, Source: (Duñabeitia *et al.*, 2018)

ba) Tourist, Source: (Duñabeitia *et al.*, 2018)

bb) Vater, Source: https://de.cleanpng.com/png-hu6n8o/download-png.html

bc) Wikinger, Source:https://de.cleanpng.com/png-en6r2o/

bd) Zauberer, Source:(Duñabeitia *et al.*, 2018)

be) Zwerg, Source: (Duñabeitia *et al.*, 2018)

APPENDIX II

LIST OF SENTENCE STIMULI FROM CHAPTER 3

The following list consists of all sentences included in this study, which were taken from the Oldenburg linguistically and audiologically controlled sentences (OLACS) corpus (Uslar *et al.*, 2013). An overview of the sentences available from the OLACS corpus, including English translations of those used in previous studies, can be found in the online description of the corpus[1].

Referent choice task

Table-A II-1 Sentences and target adjectives used for the main referent choice task

Item ID	1st adjective	2nd adjective	Target adjective	Sentence
RC 1	blind	brav	blind	Der blinde Jäger erschießt den braven Soldaten.
RC 2	blind	brav	brav	Der blinde Jäger erschießt den braven Soldaten.
RC 3	fies	brav	fies	Der fiese Pirat erschießt den braven Soldaten.
RC 4	fies	brav	brav	Der fiese Pirat erschießt den braven Soldaten.
RC 5	faul	böse	faul	Der faule Bäcker ersticht den bösen Koch.
RC 6	faul	böse	böse	Der faule Bäcker ersticht den bösen Koch.
RC 2	fies	arm	fies	Der fiese Koch ersticht den armen Touristen.
RC 7	fies	arm	arm	Der fiese Koch ersticht den armen Touristen.
RC 8	böse	dreist	böse	Der böse Gärtner erwürgt den dreisten Postboten.
RC 9	böse	dreist	dreist	Der böse Gärtner erwürgt den dreisten Postboten.
RC 10	taub	müde	taub	Der taube Elefant fängt den müden Elch.
RC 11	taub	müde	müde	Der taube Elefant fängt den müden Elch.
RC 12	gut	müde	gut	Der gute Soldat fängt den frechen Cowboy.
RC 13	gut	müde	müde	Der gute Soldat fängt den frechen Cowboy.
RC 14	blind	groß	blind	Der blinde Kasper fesselt den großen Zauberer.
RC 15	blind	groß	groß	Der blinde Kasper fesselt den großen Zauberer.
RC 16	müde	groß	müde	Der müde Drache fesselt den großen Panda.

Continued on next page

[1] http://www.aulin.uni-oldenburg.de/49349.html (Accessed: 2022-03-28)

Item ID	1^{st} adjective	2^{nd} adjective	Target adjective	Sentence
RC 17	müde	groß	groß	Der müde Drache fesselt den großen Panda.
RC 18	klein	süß	klein	Der kleine Pinguin filmt den süßen Koala.
RC 19	klein	süß	süß	Der kleine Pinguin filmt den süßen Koala.
RC 20	still	dick	still	Der stille Postbote grüßt den dicken Frisör.
RC 21	still	dick	dick	Der stille Postbote grüßt den dicken Frisör.
RC 22	müde	laut	müde	Der müde Ritter interviewt den lauten Touristen.
RC 23	müde	laut	laut	Der müde Ritter interviewt den lauten Touristen.
RC 24	dick	klein	dick	Der dicke Bär interviewt den kleinen Pinguin.
RC 25	dick	klein	klein	Der dicke Bär interviewt den kleinen Pinguin.
RC 26	schön	blass	schön	Der schöne Radfahrer jagt den blassen Cowboy.
RC 27	schön	blass	blass	Der schöne Radfahrer jagt den blassen Cowboy.
RC 28	nett	gut	nett	Der nette Papst küsst den guten Soldaten.
RC 29	nett	gut	gut	Der nette Papst küsst den guten Soldaten.
RC 30	klug	alt	klug	Der kluge Pinguin küsst den alten Esel.
RC 31	klug	alt	alt	Der kluge Pinguin küsst den alten Esel.
RC 32	dick	klein	dick	Der dicke Panda malt den kleinen Koala.
RC 33	dick	klein	klein	Der dicke Panda malt den kleinen Koala.
RC 34	stur	alt	stur	Der sture Esel malt den alten Löwen.
RC 35	stur	alt	alt	Der sture Esel malt den alten Löwen.
RC 36	groß	gut	groß	Der große Büffel malt den guten Drachen.
RC 37	groß	gut	gut	Der große Büffel malt den guten Drachen.
RC 38	bunt	wild	bunt	Der bunte Papagei malt den wilden Tiger.
RC 39	bunt	wild	wild	Der bunte Papagei malt den wilden Tiger.
RC 40	dick	stolz	dick	Der dicke Bär massiert den stolzen Tiger.
RC 41	dick	stolz	stolz	Der dicke Bär massiert den stolzen Tiger.
RC 42	nett	still	nett	Der nette Maler massiert den stillen Gärtner.
RC 43	nett	still	still	Der nette Maler massiert den stillen Gärtner.
RC 44	böse	brav	böse	Der böse Räuber schlägt den braven Soldaten.
RC 45	böse	brav	brav	Der böse Räuber schlägt den braven Soldaten.
RC 46	frech	schwach	frech	Der freche Punker schlägt den schwachen Polizis...
RC 47	frech	schwach	schwach	Der freche Punker schlägt den schwachen Polizis...
RC 48	stark	blind	stark	Der starke Koch schubst den blinden Wikinger.
RC 49	stark	blind	blind	Der starke Koch schubst den blinden Wikinger.
RC 50	dick	alt	dick	Der dicke Nikolaus streichelt den alten Mann.
RC 51	dick	alt	alt	Der dicke Nikolaus streichelt den alten Mann.

Continued on next page

Item ID	1^{st} adjective	2^{nd} adjective	Target adjective	Sentence
RC 52	böse	dick	böse	Der böse Wikinger streichelt den dicken Ritter.
RC 53	böse	dick	dick	Der böse Wikinger streichelt den dicken Ritter.
RC 54	arm	nass	arm	Der arme Pinguin tritt den nassen Frosch.
RC 55	arm	nass	nass	Der arme Pinguin tritt den nassen Frosch.
RC 56	wach	müde	wach	Der wache Löwe tritt den müden Tiger.
RC 57	wach	müde	müde	Der wache Löwe tritt den müden Tiger.
RC 58	nett	arm	nett	Der nette Lehrer tröstet den armen Jungen.
RC 59	nett	arm	arm	Der nette Lehrer tröstet den armen Jungen.
RC 60	flink	blass	flink	Der flinke Maler verfolgt den blassen Touristen.
RC 61	flink	blass	blass	Der flinke Maler verfolgt den blassen Touristen.
RC 62	nett	müde	nett	Der nette Maler weckt den müden Gärtner.
RC 63	nett	müde	müde	Der nette Maler weckt den müden Gärtner.
RC 64	groß	still	groß	Der große Bär weckt den stillen Roboter.
RC 65	groß	still	still	Der große Bär weckt den stillen Roboter.
RC 66	brav	blind	brav	Der brave Kasper weckt den blinden Maler.
RC 67	brav	blind	blind	Der brave Kasper weckt den blinden Maler.
RC 68	faul	klug	faul	Den faulen Drachen berührt der kluge Roboter.
RC 69	faul	klug	klug	Den faulen Drachen berührt der kluge Roboter.
RC 70	grau	grün	grau	Den grauen Elefanten berührt der grüne Frosch.
RC 71	grau	grün	grün	Den grauen Elefanten berührt der grüne Frosch.
RC 72	gut	alt	gut	Den guten Lehrer beschattet der alte Metzger.
RC 73	gut	alt	alt	Den guten Lehrer beschattet der alte Metzger.
RC 74	blind	brav	blind	Den blinden Jäger erschießt der brave Soldat.
RC 75	blind	brav	brav	Den blinden Jäger erschießt der brave Soldat.
RC 76	böse	brav	böse	Den bösen Jäger erschießt der brave Polizist.
RC 77	böse	brav	brav	Den bösen Jäger erschießt der brave Polizist.
RC 78	fies	brav	fies	Den fiesen Piraten erschießt der brave Soldat.
RC 79	fies	brav	brav	Den fiesen Piraten erschießt der brave Soldat.
RC 80	faul	böse	faul	Den faulen Bäcker ersticht der böse Koch.
RC 81	faul	böse	böse	Den faulen Bäcker ersticht der böse Koch.
RC 82	arm	fies	arm	Den armen Touristen ersticht der fiese Koch.
RC 83	arm	fies	fies	Den armen Touristen ersticht der fiese Koch.
RC 84	dreist	böse	dreist	Den dreisten Postboten erwürgt der böse Gärtner.
RC 85	dreist	böse	böse	Den dreisten Postboten erwürgt der böse Gärtner.
RC 86	schwarz	stur	schwarz	Den schwarzen Zauberer erwürgt der sture Koch.

Continued on next page

Item ID	1^{st} adjective	2^{nd} adjective	Target adjective	Sentence
RC 87	schwarz	stur	stur	Den schwarzen Zauberer erwürgt der sture Koch.
RC 88	gut	frech	gut	Den guten Soldaten fängt der freche Cowboy.
RC 89	gut	frech	frech	Den guten Soldaten fängt der freche Cowboy.
RC 90	blind	groß	blind	Den blinden Kasper fesselt der große Zauberer.
RC 91	blind	groß	groß	Den blinden Kasper fesselt der große Zauberer.
RC 92	böse	jung	böse	Den bösen Piraten fesselt der junge Prinz.
RC 93	böse	jung	jung	Den bösen Piraten fesselt der junge Prinz.
RC 94	müde	groß	müde	Den müden Drachen fesselt der große Panda.
RC 95	müde	groß	groß	Den müden Drachen fesselt der große Panda.
RC 96	süß	klein	süß	Den süßen Koala filmt der kleine Pinguin.
RC 97	süß	klein	klein	Den süßen Koala filmt der kleine Pinguin.
RC 98	alt	klug	alt	Den alten Pfarrer grüßt der kluge Pilot.
RC 99	alt	klug	klug	Den alten Pfarrer grüßt der kluge Pilot.
RC 100	streng	böse	streng	Den strengen Zauberer jagt der böse Räuber.
RC 101	streng	böse	böse	Den strengen Zauberer jagt der böse Räuber.
RC 102	dick	klein	dick	Den dicken Koala jagt der kleine Maulwurf.
RC 103	dick	klein	klein	Den dicken Koala jagt der kleine Maulwurf.
RC 104	nett	gut	nett	Den netten Papst küsst der gute Soldat.
RC 105	nett	gut	gut	Den netten Papst küsst der gute Soldat.
RC 106	krank	scheu	krank	Den kranken Hasen küsst der scheue Maulwurf.
RC 107	krank	scheu	scheu	Den kranken Hasen küsst der scheue Maulwurf.
RC 108	klein	dick	klein	Den kleinen Koala malt der dicke Panda.
RC 109	klein	dick	dick	Den kleinen Koala malt der dicke Panda.
RC 110	alt	stur	alt	Den alten Löwen malt der sture Esel.
RC 111	alt	stur	stur	Den alten Löwen malt der sture Esel.
RC 112	wild	bunt	wild	Den wilden Tiger malt der bunte Papagei.
RC 113	wild	bunt	bunt	Den wilden Tiger malt der bunte Papagei.
RC 114	brav	böse	brav	Den braven Soldaten schlägt der böse Räuber.
RC 115	brav	böse	böse	Den braven Soldaten schlägt der böse Räuber.
RC 116	dick	alt	dick	Den dicken Nikolaus streichelt der alte Mann.
RC 117	dick	alt	alt	Den dicken Nikolaus streichelt der alte Mann.
RC 118	stolz	frech	stolz	Den stolzen Clown tadelt der freche Kasper.
RC 119	stolz	frech	frech	Den stolzen Clown tadelt der freche Kasper.
RC 120	nass	arm	nass	Den nassen Frosch tritt der arme Pinguin.
RC 121	nass	arm	arm	Den nassen Frosch tritt der arme Pinguin.

Continued on next page

Item ID	1^{st} adjective	2^{nd} adjective	Target adjective	Sentence
RC 122	alt	jung	alt	Den alten König tröstet der junge Prinz.
RC 123	alt	jung	jung	Den alten König tröstet der junge Prinz.
RC 124	arm	nett	arm	Den armen Jungen tröstet der nette Lehrer.
RC 125	arm	nett	nett	Den armen Jungen tröstet der nette Lehrer.
RC 126	dünn	treu	dünn	Den dünnen Arzt umarmt der treue Pilot.
RC 127	dünn	treu	treu	Den dünnen Arzt umarmt der treue Pilot.
RC 128	dick	klein	dick	Den dicken Nikolaus umarmt der kleine Junge.
RC 129	dick	klein	klein	Den dicken Nikolaus umarmt der kleine Junge.
RC 130	schwer	dick	schwer	Den schweren Boxer verfolgt der dicke Postbote.
RC 131	schwer	dick	dick	Den schweren Boxer verfolgt der dicke Postbote.
RC 132	schnell	lahm	schnell	Den schnellen Elefanten verfolgt der lahme Elch.
RC 133	schnell	lahm	lahm	Den schnellen Elefanten verfolgt der lahme Elch.
RC 134	still	groß	still	Den stillen Roboter weckt der große Bär.
RC 135	still	groß	groß	Den stillen Roboter weckt der große Bär.
RC 136	brav	blind	brav	Den braven Kasper weckt der blinde Maler.
RC 137	brav	blind	blind	Den braven Kasper weckt der blinde Maler.
RC 138	arm	groß	arm	Den armen Matrosen weckt der große Kapitän.
RC 139	arm	groß	groß	Den armen Matrosen weckt der große Kapitän.

Sentences and target adjectives used as fillers for the referent choice task

Item ID	1^{st} adjective	2^{nd} adjective	Target adjective	Sentence
RC 140	schlau	faul	gut	Der schlaue Kasper beschattet den faulen Vater.
RC 142	schlau	faul	dick	Der schlaue Kasper beschattet den faulen Vater.
RC 143	flink	träge	müde	Der flinke Zwerg fesselt den trägen Riesen.
RC 144	flink	träge	dick	Der flinke Zwerg fesselt den trägen Riesen.
RC 145	rüde	frech	wach	Der rüde Cowboy jagt den frechen Kobold.
RC 146	rüde	frech	faul	Der rüde Cowboy jagt den frechen Kobold.
RC 147	süß	lieb	nett	Der süße Junge küsst den lieben Vater.
RC 148	süß	lieb	rüde	Der süße Junge küsst den lieben Vater.
RC 149	böse	frech	faul	Der böse Zauberer tadelt den frechen Kobold.
RC 150	böse	frech	grau	Der böse Zauberer tadelt den frechen Kobold.
RC 151	alt	klug	still	Den alten Pfarrer grüßt der kluge Pilot.
RC 152	alt	klug	blind	Den alten Pfarrer grüßt der kluge Pilot.

Continued on next page

Item ID	1st adjective	2nd adjective	Target adjective	Sentence
RC 153	frech	rüde	böse	Den frechen Kobold jagt der rüde Cowboy.
RC 154	frech	rüde	alt	Den frechen Kobold jagt der rüde Cowboy.
RC 155	stark	lahm	klein	Den starken Touristen schubst der lahme Bauer.
RC 156	stark	lahm	schlau	Den starken Touristen schubst der lahme Bauer.
RC 157	frech	böse	dick	Den frechen Kobold tadelt der böse Zauberer.
RC 158	frech	böse	streng	Den frechen Kobold tadelt der böse Zauberer.
RC 159	dick	hübsch	klug	Den dicken Mönch tröstet der hübsche Bräutigam.
RC 160	dick	hübsch	gut	Den dicken Mönch tröstet der hübsche Bräutigam.

Sentences and target adjectives used as practice for the referent choice task at the beginning of the study

Item ID	1st adjective	2nd adjective	Target adjective	Sentence
P 1	flink	träge	flink	Der flinke Zwerg fesselt den trägen Riesen.
P 2	rüde	frech	frech	Der rüde Cowboy jagt den frechen Kobold.
P 3	alt	klug	alt	Den alten Pfarrer grüßt der kluge Pilot.
P 4	dick	hübsch	hübsch	Den dicken Mönch tröstet der hübsche Bräutigam.
P 5	böse	frech	dick	Der böse Zauberer tadelt den frechen Kobold.

Simultaneity judgment task

List of sentences used for the simultaneity judgment task

Item ID	Sentence
SJ 1	Der Punker, der die Maler beschattet, niest.
SJ 2	Der Maler, der die Vampire beschattet, gähnt.
SJ 3	Der Lehrer, der die Models bestiehlt, zittert.
SJ 4	Der Mönch, der die Astronauten erschießt, lacht.
SJ 5	Der Frisör, der die Bäcker erschießt, niest.
SJ 6	Der Frisör, der die Köchinnen erschießt, grinst.
SJ 7	Der Koch, der die Touristinnen erschießt, niest.
SJ 8	Der Bräutigam, der die Riesen ersticht, lacht.
SJ 9	Der Maler, der die Witwen ersticht, zittert.

Continued on next page

Item ID	Sentence
SJ 10	Der Richter, der die Radfahrer erwürgt, weint.

APPENDIX III

SUPPLEMENTARY STATISTICAL ANALYSES FROM CHAPTER 3

Full model coefficients tables for the reported analyses

Below, we list the results tables for the statistical analyses of the reaction times and accuracy on the referent choice task, as well as accuracy on the simultaneity judgment task. With respect to simple effects and two-way interactions, these tables are identical to the ones reported in the main paper. In addition, they report the (non-significant) three-way interactions between factors.

Table-A III-1 Model coefficients for the linear mixed effects model fit to reaction times and the binary logistic regression model fit to accuracy in the referent choice task (comparison against baseline). Estimates for reaction time are expressed in milliseconds, and estimates for accuracy are expressed in log odds ratios. * = p ≤ 0.05, ** = p ≤ 0.01, *** = p ≤ 0.001

	Reaction times			**Accuracy**		
Measure	Estimate	SE	*z*-value	Estimate	SE	*z*-value
(Intercept)	1907.20	50.97	37.42***	4.07	0.31	13.27***
10 ms only AV	-27.54	26.54	-1.04	0.21	0.38	0.54
400 ms only AV	370.55	26.69	13.89***	-0.59	0.33	-1.77
10 ms AV & RA	28.02	26.51	1.06	-0.31	0.34	-0.90
400 ms AV & RA	219.91	26.67	8.25***	-0.43	0.34	-1.28
position (second)	993.33	26.53	37.45***	-0.78	0.32	-2.40*
trial number	-176.87	20.03	-8.83***	0.53	0.27	1.96*
position:trial	81.64	27.40	2.98**	-0.24	0.33	-0.75
10 ms only AV:position	54.31	37.20	1.46	-0.24	0.45	-0.52
400 ms only AV:position	13.46	37.42	0.36	0.42	0.41	1.01
10 ms AV & RA:position	-24.86	37.22	-0.67	0.06	0.42	0.14
400 ms AV & RA:position	158.43	37.42	4.23***	0.09	0.41	0.22
10 ms only AV:trial	-6.17	27.24	-0.23	-0.53	0.37	-1.42
400 ms only AV:trial	40.35	27.77	1.45	-0.02	0.33	-0.07
10 ms AV & RA:trial	-7.90	27.07	-0.29	-0.30	0.34	-0.87
400 ms AV & RA:trial	-14.16	27.48	-0.52	-0.21	0.34	-0.62
10 ms only AV:position:trial	12.91	37.63	0.34	0.15	0.45	0.33
400 ms only AV:position:trial	-1.62	38.06	-0.04	0.01	0.41	0.02
10 ms AV & RA:position:trial	11.26	37.91	0.30	-0.08	0.42	-0.18
400 ms AV & RA:position:trial	27.75	38.52	0.72	0.15	0.42	0.37

Table-A III-2 Model coefficients for the linear mixed effects model fit to reaction times and the binary logistic regression model fit to accuracy in the referent choice task (contrast only AV vs. AV & RA). Estimates for reaction time are expressed in milliseconds, and estimates for accuracy are expressed in log odds ratios. * = p ≤ 0.05, ** = p ≤ 0.01, *** = p ≤ 0.001

	Reaction times			**Accuracy**		
Measure	Estimate	SE	*z*-value	Estimate	SE	*z*-value
(Intercept)	2052.95	47.03	43.66***	3.79	0.17	21.87***
10 ms only AV vs. AV & RA	-54.33	27.68	-1.96*	0.42	0.31	1.36
400 ms only AV vs. AV & RA	149.25	28.01	5.33***	-0.20	0.31	-0.67
position (second)	1043.50	13.90	75.08***	-0.68	0.12	-5.58***
trial number	-171.80	9.92	-17.32***	0.31	0.10	3.15**
position:trial	93.12	13.98	6.66***	-0.19	0.12	-1.57
10 ms only AV vs. AV & RA:position	75.48	39.10	1.93	-0.22	0.39	-0.56
400 ms only AV vs. AV & RA:position	-144.96	39.54	-3.67***	0.37	0.38	0.96
10 ms only AV vs. AV & RA:trial	6.05	27.51	0.22	-0.19	0.30	-0.66
400 ms only AV vs. AV & RA:trial	58.82	28.53	2.06*	0.25	0.31	0.79
position:trial:10 ms only AV vs. AV & RA	6.03	39.02	0.15	0.15	0.38	0.40
position:trial:400 ms only AV vs. AV & RA	-38.83	40.11	-0.97	-0.19	0.39	-0.50

Table-A III-3 Model coefficients for the binary logistic regression models fit to the responses on the simultaneity judgment task (comparison against baseline). Estimates for accuracy are expressed in log odds ratios. * = p ≤ 0.05, ** = p ≤ 0.01, *** = p ≤ 0.001.

	Audiovisual synchrony			**Distortion**		
Measure	Estimate	SE	*z*-value	Estimate	SE	*z*-value
(Intercept)	-3.71	0.39	-9.43***	3.84	0.51	7.60***
10 ms only AV	0.17	0.42	0.41	0.52	0.74	0.71
400 ms only AV	3.43	0.37	9.32***	-1.08	0.53	-2.03*
10 ms AV & RA	0.65	0.40	1.61	-0.00	0.64	0.00
400 ms AV & RA	3.84	0.36	10.68***	-6.45	0.55	-11.75***
block2	-0.10	0.21	-0.49	0.52	0.34	1.51
block3	-0.04	0.21	-0.21	0.58	0.34	1.68

Analysis of the effect of syntactic complexity on the referent choice task

As the referent choice task used two different German word orders, namely the subject-verb-object and the object-verb-subject orders (for details, see the main paper), an additional analysis was conducted for syntactic complexity, that is, whether reaction times or task accuracy differed depending on the word order of the stimulus sentence. To that end, we added the binary predictor *word order* as sum-coded (-1, 1) factor, together with all possible two-way interactions, to the linear mixed effects models and the binary logistic regression models fit to reaction times and task accuracy, respectively. At the same time, to reduce model complexity, we removed the previously non-significant three-way interaction terms between task condition, target position, and trial number. The models indicated that participants had a significantly lower reaction time in the subject-verb-object conditions (β = 80.00, SE = 32.09, t = 2.49), but crucially there were no interactions between the five delay conditions, the position of the target, or the trial number. With respect to task accuracy, the word order had no effects.

Table-A III-4 Model coefficients for the linear mixed effects model fit to reaction times and the binary logistic regression model fit to accuracy in the referent choice task (comparison against baseline), both including word order as predictor. Estimates for reaction time are expressed in milliseconds, and estimates for accuracy are expressed in log odds ratios. * = p ≤ 0.05, ** = p ≤ 0.01, *** = p ≤ 0.001

	Reaction times			**Accuracy**		
Measure	Estimate	SE	*z*-value	Estimate	SE	*z*-value
(Intercept)	2225.58	54.67	40.70***	4.04	0.28	14.37***
10 ms only AV	-27.02	40.50	-0.67	0.23	0.36	0.65
400 ms only AV	301.70	41.32	7.30***	-0.57	0.30	-1.89
10 ms AV & RA	33.23	40.81	0.81	-0.27	0.32	-0.83
400 ms AV & RA	225.70	41.63	5.42***	-0.39	0.31	-1.26
word order	80.00	32.09	2.49**	0.05	0.18	0.29
position (second)	833.10	32.11	25.95***	-0.75	0.29	-2.59**
trial number	-3.96	0.33	-12.06***	0.49	0.17	2.84**
word order:position	8.17	11.70	0.70	-0.07	0.12	-0.58
word order:trial	-0.03	0.13	-0.20	-0.01	0.06	-0.19
position:trial	1.99	0.26	7.74***	-0.19	0.12	-1.56
10 ms only AV:position	52.56	36.94	1.42	-0.27	0.43	-0.62
400 ms only AV:position	11.09	37.16	0.30	0.40	0.38	1.06
10 ms AV & RA:position	-26.58	36.96	-0.72	0.04	0.39	0.11
400 ms AV & RA:position	157.60	37.15	4.24***	0.05	0.38	0.13
10 ms only AV:trial	0.01	0.41	0.02	-0.43	0.21	-2.06*
400 ms only AV:trial	0.86	0.41	2.09*	-0.003	0.20	-0.02
10 ms AV & RA:trial	-0.05	0.41	-0.13	-0.33	0.20	-1.68
400 ms AV & RA:trial	-0.07	0.42	-0.16	-0.11	0.20	-0.54
10 ms only AV:word order	0.04	18.46	0.002	0.03	0.37	-1.42
400 ms only AV:word order	12.40	18.59	0.67	0.04	0.18	0.22
10 ms AV & RA:word order	10.81	18.47	0.59	-0.25	0.19	-1.31
400 ms AV & RA:word order	33.86	18.60	1.82	-0.08	0.18	-0.42

Table-A III-5 Model coefficients for the linear mixed effects model fit to reaction times and the binary logistic regression model fit to accuracy in the referent choice task (contrast only AV vs. AV & RA), both including word order as predictor. Estimates for reaction time are expressed in milliseconds, and estimates for accuracy are expressed in log odds ratios. * = p ≤ 0.05, ** = p ≤ 0.01, *** = p ≤ 0.001

	Reaction times			**Accuracy**		
Measure	Estimate	SE	*z*-value	Estimate	SE	*z*-value
(Intercept)	2354.28	49.14	47.91***	3.79	0.17	21.87***
10 ms only AV vs. AV & RA	-70.39	44.47	-1.58	0.44	0.30	1.44
400 ms only AV vs. AV & RA	79.45	46.08	1.72	-0.23	0.30	-0.77
word order	84.22	30.95	2.72**	0.004	0.13	0.03
position (second)	879.96	28.87	30.48***	-0.68	0.12	-5.53***
trial number	-3.75	0.22	-17.37***	0.32	0.10	3.23**
word order:position	2.12	13.89	0.15	-0.08	0.12	-0.65
word order:trial	0.12	0.15	0.76	-0.02	0.06	-0.26
position:trial	2.02	0.30	6.64***	-0.19	0.12	-1.58
10 ms only AV vs. AV & RA:position	76.10	39.09	1.95	-0.26	0.38	-0.68
400 ms only AV vs. AV & RA:position	-147.02	39.47	-3.72***	0.39	0.38	1.03
10 ms only AV vs. AV & RA:trial	0.20	0.42	0.47	-0.06	0.18	-0.34
400 ms only AV vs. AV & RA:trial	0.89	0.44	2.03*	0.12	0.19	0.65
10 ms only AV vs. AV & RA:word order	-10.49	19.55	-0.54	0.24	0.18	1.32
400 ms only AV vs. AV & RA:word order	-25.59	19.77	-1.30	0.10	0.18	0.57

Analysis of correlations between synchrony and distortion judgments on the simultaneity judgment task

As reported in the main paper, we analyzed responses on the audiovisual synchrony and distortion judgment questions for their within-trial correlations, both when pooling the data from all five conditions and when submitting the responses on the five experimental conditions to separate correlation tests. Correlations were determined using Pearson two-sided product-moment correlation tests. The full results are reported in Table III-6.

Table-A III-6 Correlation coefficients and confidence intervals for the Pearson product-moment correlation between synchrony and distortion judgments for all five experimental conditions

Measure	r	df	*p*-value	95 % CI
Pooled data	-0.33	1498	≤ 0.001	[-0.38, -0.29]
0 ms	-0.02	298	0.70	[-0.13, 0.09]
10 ms only AV	-0.13	298	0.02	[-0.24, -0.02]
10 ms AV & RA	-0.12	298	0.04	[-0.23, -0.01]
400 ms only AV	-0.05	298	0.42	[-0.16, 0.07]
400 ms AV & RA	-0.06	298	0.27	[-0.18, 0.05]

APPENDIX IV

DESCRIPTION OF FEATURES FROM CHAPTER 4

Description of all heart rate variability features included.

Feature	Category	Definition
MeanNN	Time-domain	The mean of the RR intervals.
SDNN	Time-domain	The standard deviation of the RR intervals.
SDANN1	Time-domain	The standard deviation of average RR intervals extracted from 1-minute segments of time series data (requires minimum segment duration of 3 minutes).
SDANN2	Time-domain	The standard deviation of average RR intervals extracted from 2-minute segments of time series data (requires minimum segment duration of 6 minutes).
SDANN5	Time-domain	The standard deviation of average RR intervals extracted from 5-minute segments of time series data (requires minimum segment duration of 15 minutes).
SDNNI1	Time-domain	The mean of the standard deviations of RR intervals extracted from 1-minute segments of time series data (requires minimum segment duration of 3 minutes).

Continued on next page

Feature	Category	Definition
SDNNI2	Time-domain	The mean of the standard deviations of RR intervals extracted from 2-minute segments of time series data (requires minimum segment duration of 6 minutes).
SDNNI1	Time-domain	The mean of the standard deviations of RR intervals extracted from 5-minute segments of time series data (requires minimum segment duration of 15 minutes).
RMSSD	Time-domain	The square root of the mean of the squared successive differences between adjacent RR intervals.
SDSD	Time-domain	The standard deviation of the successive differences between RR intervals.
CVNN	Time-domain	The standard deviation of the RR intervals (SDNN) divided by the mean of the RR intervals (MeanNN).
CVSD	Time-domain	The root mean square of successive differences (RMSSD) divided by the mean of the RR intervals (MeanNN).
MedianNN	Time-domain	The median of the RR intervals.
MadNN	Time-domain	The median absolute deviation of the RR intervals.

Continued on next page

Feature	Category	Definition
HCVNN	Time-domain	The median absolute deviation of the RR intervals (MadNN) divided by the median of the RR intervals (MedianNN).
IQRNN	Time-domain	The interquartile range (IQR) of the RR intervals.
Prc20NN	Time-domain	The 20th percentile of the RR intervals (Han *et al.*, 2017; Hovsepian *et al.*, 2015).
Prc80NN	Time-domain	The 80th percentile of the RR intervals (Han *et al.*, 2017; Hovsepian *et al.*, 2015).
pNN50	Time-domain	The proportion of RR intervals greater than 50ms, out of the total number of RR intervals.
pNN20	Time-domain	The proportion of RR intervals greater than 20ms, out of the total number of RR intervals.
MinNN	Time-domain	The minimum of the RR intervals (Parent *et al.*, 2019; Subramaniam & Dass, 2022).
MaxNN	Time-domain	The maximum of the RR intervals (Parent *et al.*, 2019; Subramaniam & Dass, 2022).
TINN	Time-domain	A geometrical parameter of the HRV; more specifically, the baseline width of the RR intervals distribution obtained by triangular interpolation, where the error of least squares determines the triangle. It is an approximation of the RR interval distribution.

Continued on next page

Feature	Category	Definition
HTI	Time-domain	The HRV triangular index, measuring the total number of RR intervals divided by the height of the RR intervals histogram.
ULF	Frequency-domain	The spectral power of ultra low frequencies (.0 to .0033 Hz).
VLF	Frequency-domain	The spectral power of very low frequencies (.0033 to .04 Hz).
LF	Frequency-domain	The spectral power of low frequencies (.04 to .15 Hz).
HF	Frequency-domain	The spectral power of high frequencies (.15 to .4 Hz).
VHF	Frequency-domain	The spectral power of very high frequencies (.4 to .5 Hz).
LFn	Frequency-domain	The normalized low frequency, obtained by dividing the low frequency power by the total power.
HFn	Frequency-domain	The normalized high frequency, obtained by dividing the low frequency power by the total power.
LnHF	Frequency-domain	The log transformed HF.
SD1	Non-linear	Standard deviation perpendicular to the line of identity. It is an index of short-term RR interval fluctuations, i.e., beat-to-beat variability.

Continued on next page

Feature	Category	Definition
SD2	Non-linear	Standard deviation along the identity line. Index of long-term HRV changes.
SD1SD2	Non-linear	Ratio of SD1 to SD2. Describes the ratio of short term to long term variations in HRV.
S	Non-linear	Area of ellipse described by SD1 and SD2. It is proportional to SD1SD2.
CSI	Non-linear	The Cardiac Sympathetic Index (Toichi, Sugiura, Murai & Sengoku, 1997) is a measure of cardiac sympathetic function independent of vagal activity, calculated by dividing the longitudinal variability of the Poincaré plot (4*SD2) by its transverse variability (4*SD1).
CVI	Non-linear	The Cardiac Vagal Index (Toichi *et al.*, 1997) is an index of cardiac parasympathetic function (vagal activity unaffected by sympathetic activity), and is equal equal to the logarithm of the product of longitudinal (4*SD2) and transverse variability (4*SD1).
CSI_Modified	Non-linear	The modified CSI (Jeppesen, Beniczky, Johansen, Sidenius & Fuglsang-Frederiksen, 2014) obtained by dividing the square of the longitudinal variability by its transverse variability.

Continued on next page

Feature	Category	Definition
GI	Non-linear	Guzik's Index, defined as the distance of points above line of identity (LI) to LI divided by the distance of all points in Poincaré plot to LI except those that are located on LI.
SI	Non-linear	Slope Index, defined as the phase angle of points above LI divided by the phase angle of all points in Poincaré plot except those that are located on LI.
AI	Non-linear	Area Index, defined as the cumulative area of the sectors corresponding to the points that are located above LI divided by the cumulative area of sectors corresponding to all points in the Poincaré plot except those that are located on LI.
PI	Non-linear	Porta's Index, defined as the number of points below LI divided by the total number of points in Poincaré plot except those that are located on LI.
SD1d	Non-linear	Short-term variance of contributions of decelerations (Piskorski & Guzik, 2011).
SD1a	Non-linear	Short-term variance of contributions of accelerations (Piskorski & Guzik, 2011).
C1d	Non-linear	The contributions of heart rate decelerations to short-term HRV (Piskorski & Guzik, 2011).
C1a	Non-linear	The contributions of heart rate accelerations to short-term HRV (Piskorski & Guzik, 2011).

Continued on next page

Feature	Category	Definition
SD2d	Non-linear	Long-term variance of contributions of decelerations (Piskorski & Guzik, 2011).
SD2a	Non-linear	Long-term variance of contributions of accelerations (Piskorski & Guzik, 2011).
C2d	Non-linear	The contributions of heart rate decelerations to long-term HRV (Piskorski & Guzik, 2011).
C2a	Non-linear	The contributions of heart rate accelerations to long-term HRV (Piskorski & Guzik, 2011).
SDNNd	Non-linear	Total variance of contributions of decelerations (Piskorski & Guzik, 2011).
SDNNa	Non-linear	Total variance of contributions of accelerations (Piskorski & Guzik, 2011).
Cd	Non-linear	The total contributions of heart rate decelerations to HRV.
Ca	Non-linear	The total contributions of heart rate accelerations to HRV.
PIP	Non-linear	Percentage of inflection points of the RR intervals series.
IALS	Non-linear	Inverse of the average length of the acceleration/deceleration segments.
PSS	Non-linear	Percentage of short segments.
PAS	Non-linear	Percentage of NN intervals in alternation segments.

Continued on next page

Feature	Category	Definition
ApEn	Non-linear	Approximate entropy (Sabeti, Katebi & Boostani, 2009; Shi, Zhang, Yuan, Wang & Li, 2017).
SampEn	Non-linear	Sample entropy.
FuzzyEn	Non-linear	Fuzzy entropy (Ishikawa & Mieno, 1979; Zadeh, Klir & Yuan, 1996).
MSE	Non-linear	Multiscale entropy (Costa, Goldberger & Peng, 2002).
CMSE	Non-linear	Composite Multiscale Entropy (Wu, Wu, Lin, Wang & Lee, 2013).
RCMSE	Non-linear	Refined Composite Multiscale Entropy (Wu *et al.*, 2013).
CD	Non-linear	Correlation Dimension (Bolea *et al.*, 2014; Boon, Henry, Suttle & Dain, 2008), a lower bound estimate of the fractal dimension of a signal.
HFD	Non-linear	Higuchi's Fractal Dimension (Higuchi, 1988), an approximate value for the box-counting dimension for time series.
KFD	Non-linear	Katz's Fractal Dimension (Katz, 1988), based on euclidean distances between successive points in the signal which are summed and averaged, and the maximum distance between the starting and any other point in the sample.

Continued on next page

Feature	Category	Definition
LZC	Non-linear	Lempel-Ziv Complexity (Lempel & Ziv, 1976), quantifying the regularity of the signal by scanning symbolic sequences for new patterns, increasing the complexity count every time a new sequence is detected. Regular signals have a lower number of distinct patterns and thus have low values whereas irregular signals are characterized by a high value. While often being interpreted as a complexity measure, it was originally proposed to reflect randomness.
DFA_alpha1	Non-linear	Monofractal detrended fluctuation analysis corresponding to short-term correlations.
MFDFA_alpha1_-Width	Non-linear	The width feature of multifractical detrended fluctuation analysis (corresponding to short-term correlations). It is the width of the singularity spectrum and quantifies the degree of the multifractality.
MFDFA_alpha1_-Peak	Non-linear	The peak feature of multifractical detrended fluctuation analysis (corresponding to short-term correlations). It is the value of the singularity exponent H corresponding to the peak of singularity dimension D. It is a measure of the self-affinity of the signal, and a high value is an indicator of high degree of correlation between the data points.

Continued on next page

Feature	Category	Definition
MFDFA_alpha1_-Mean	Non-linear	The mean feature of multifractical detrended fluctuation analysis (corresponding to short-term correlations). It is the mean of the maximum and minimum values of singularity exponent H, which quantifies the average fluctuations of the signal.
MFDFA_alpha1_-Max	Non-linear	The max feature of multifractical detrended fluctuation analysis (corresponding to short-term correlations). It is the value of singularity spectrum D corresponding to the maximum value of singularity exponent H, which indicates the maximum fluctuation of the signal.
MFDFA_alpha1_-Delta	Non-linear	The delta feature of multifractical detrended fluctuation analysis (corresponding to short-term correlations). It is the vertical distance between the singularity spectrum D where the singularity exponents are at their minimum and maximum, corresponding to the range of fluctuations of the signal.
MFDFA_alpha1_-Asymmetry	Non-linear	The asymmetry feature of multifractical detrended fluctuation analysis (corresponding to short-term correlations). The Asymmetric Ratio corresponds to the centrality of the peak of the spectrum (Orozco-Duque, Novak, Kremen & Bustamante, 2015).

APPENDIX V

SUPPLEMENTARY EXPERIMENTS FROM CHAPTER 4

Hyperparameter optimization

Below, supplementary results from the stress recognition experiments are presented. These experiments followed the same methodology as those reported in the main paper, with the exception that grid search was employed to optimize several hyperparameters for each classification algorithm. The range of hyperparameter values used in the grid search was selected based on previous work (Opoku Asare *et al.*, 2022) and is provided in Table V-1:

Table-A V-1 Range of Hyperparameter Values Used in the Grid Search

Classifier	Hyperparameter	Range
Log. Reg.	C	[0.01, 0.1, 1, 10, 100]
Log. Reg.	solver	[“newton-cg”, “lbfgs”, “liblinear”, “saga”]
XGBoost	learning_rate	[0.05, 0.10, 0.15, 0.20]
XGBoost	max_depth	[3, 4, 5, 6]
XGBoost	gamma	[0.0, 0.1, 0.2, 0.3, 0.4]
XGBoost	min_child_weight	[1, 3, 5, 7]

In these experiments, the hyperparameter values for each iteration of the cross-validation process were separately selected through a nested cross-validation procedure. Specifically, the training data for each iteration of the outer five-fold cross-validation (24/30 participants) was divided into four inner folds. Three of these folds (18/24 participants) were used for training the classification algorithm with different hyperparameter values, while the remaining fold (6/24 participants) was used to evaluate the classification accuracy with these hyperparameters. The hyperparameter values that resulted in the highest mean accuracy across the four inner cross-validation iterations were then used to train the classification algorithm on the outer training data for that iteration. This process was repeated for all 50 iterations of the outer cross-validation, such that the hyperparameter values were optimally chosen for each iteration.

The following tables present the accuracy obtained with optimized hyperparameters (Acc. opt.), as well as the accuracy achieved with the fixed hyperparameters (Acc. fix.) reported in the main paper:

Table-A V-2 Results of Hyperparameter Optimization for Stress Classifiers Trained on the ECG and Tested on the ECG with Baselining

Features	Classifier	Acc. opt. (%)	Acc. fix. (%)
MedianNN	Log. Reg.	75.5 ± 7.4	76.0 ± 7.3
MedianNN	XGBoost	70.5 ± 8.6	67.5 ± 8.4
MedianNN & RMSSD	Log. Reg.	76.3 ± 8.1	76.7 ± 8.5
MedianNN & RMSSD	XGBoost	72.4 ± 8.3	72.0 ± 9.4
All HRV features	Log. Reg.	78.2 ± 8.2	78.9 ± 8.1
All HRV features	XGBoost	76.3 ± 8.8	76.3 ± 7.8

Table-A V-3 Results of Hyperparameter Optimization for Stress Classifiers Trained on the ECG and Tested on the ECG without Baselining

Features	Classifier	Acc. opt. (%)	Acc. fix. (%)
MedianNN	Log. Reg.	57.3 ± 6.1	57.2 ± 5.8
MedianNN	XGBoost	63.6 ± 7.5	63.7 ± 7.1
MedianNN & RMSSD	Log. Reg.	59.2 ± 6.9	59.5 ± 6.1
MedianNN & RMSSD	XGBoost	62.7 ± 7.5	59.9 ± 7.9
All HRV features	Log. Reg.	74.4 ± 7.9	76.5 ± 8.4
All HRV features	XGBoost	74.2 ± 8.9	73.0 ± 8.5

Table-A V-4 Results of Hyperparameter Optimization for Stress Classifiers Trained on the IEM and Tested on the IEM without Baselining

Features	Classifier	Acc. opt. (%)	Acc. fix. (%)
MedianNN	Log. Reg.	62.6 ± 6.2	62.2 ± 6.1
MedianNN	XGBoost	58.8 ± 9.2	58.7 ± 10.3
MedianNN & RMSSD	Log. Reg.	75.8 ± 6.5	75.9 ± 6.5
MedianNN & RMSSD	XGBoost	77.6 ± 6.7	79.1 ± 7.1
All HRV features	Log. Reg.	79.2 ± 6.7	81.1 ± 8.0
All HRV features	XGBoost	76.5 ± 6.2	76.1 ± 6.7

Table-A V-5 Results of Hyperparameter Optimization for Stress Classifiers Trained on the IEM and Tested on the ECG without Baselining

Features	Classifier	Acc. opt. (%)	Acc. fix. (%)
MedianNN	Log. Reg.	57.8 ± 7.5	57.8 ± 7.5
MedianNN	XGBoost	60.3 ± 8.4	62.7 ± 8.5
MedianNN & RMSSD	Log. Reg.	50.7 ± 1.6	50.8 ± 1.7
MedianNN & RMSSD	XGBoost	50.1 ± 0.6	50.1 ± 0.6
All HRV features	Log. Reg.	50.0 ± 0.0	50.0 ± 0.0
All HRV features	XGBoost	50.2 ± 0.8	50.2 ± 0.8

Table-A V-6 Results of Hyperparameter Optimization for Stress Classifiers Trained on the IEM and Tested on the SYN without Baselining

Features	Classifier	Acc. opt. (%)	Acc. fix. (%)
MedianNN	Log. Reg.	61.1 ± 5.4	61.1 ± 5.4
MedianNN	XGBoost	60.6 ± 4.7	59.7 ± 4.7
MedianNN & RMSSD	Log. Reg.	59.2 ± 4.7	59.1 ± 4.5
MedianNN & RMSSD	XGBoost	53.3 ± 3.1	54.0 ± 2.3
All HRV features	Log. Reg.	62.0 ± 6.3	64.9 ± 5.3
All HRV features	XGBoost	59.1 ± 4.2	60.9 ± 5.0

Table-A V-7 Results of Hyperparameter Optimization for Stress Classifiers Trained on the ECG and Tested on the IEM without Baselining

Features	Classifier	Acc. opt. (%)	Acc. fix. (%)
MedianNN	Log. Reg.	61.2 ± 6.0	61.3 ± 6.3
MedianNN	XGBoost	59.5 ± 9.5	60.8 ± 9.5
MedianNN & RMSSD	Log. Reg.	52.6 ± 5.3	54.2 ± 4.9
MedianNN & RMSSD	XGBoost	58.6 ± 9.2	56.0 ± 9.2
All HRV features	Log. Reg.	66.3 ± 7.3	65.4 ± 9.3
All HRV features	XGBoost	59.7 ± 7.5	60.7 ± 7.3

Table-A V-8 Results of Hyperparameter Optimization for Stress Classifiers Trained on the ECG & SYN Using All HRV Features without Baselining

Test signal	Classifier	Acc. opt. (%)	Acc. fix. (%)
ECG	Log. Reg.	74.0 ± 7.6	74.3 ± 8.1
ECG	XGBoost	75.4 ± 7.5	76.7 ± 7.9
IEM	Log. Reg.	75.7 ± 9.7	69.9 ± 9.0
IEM	XGBoost	76.1 ± 6.7	76.6 ± 7.0
SYN	Log. Reg.	72.3 ± 5.4	70.6 ± 5.8
SYN	XGBoost	77.0 ± 5.0	77.3 ± 4.7

APPENDIX VI

INSTITUTIONAL REVIEW BOARD DOCUMENTATION

Projet H20200608

APPROVED July 21, 2021
Research Ethics Comittee – ÉTS

INFORMATION AND CONSENT FORM

TITLE OF THE RESEARCH PROJECT

“Listen to your heart'': Exploring the Link between In-Ear Audio and Emotions

RESEARCHER IN CHARGE OF THE PROJECT

Jérémie Voix, Professor in the Department of Mechanical Engineering – École de technologie supérieure (ÉTS)

Researcher in charge of the data bank CRITIAS-DB

CO-RESEARCHER

Rachel Bouserhal, CRITIAS Researcher

Vincent Pintat, CRITIAS Laboratory Manager

Cécile Le Cocq, CRITIAS Laboratory Coordinator

Sylvie Hébert, Professor in the Faculty of Medicine, School of Speech Pathology and Audiology, Université de Montréal

Alain Vinet, Professor in the Faculty of Medicine, Department of Pharmacology and Physiology, Université de Montréal

STUDENT

Danielle Benesch, Master’s student in the Department of Mechanical Engineering – ÉTS

Bérangère Villatte, Doctoral student in Biomedical Sciences, Université de Montréal

FUNDING

This project has received funding from the Natural Sciences and Engineering Research Council of Canada (NSERC) and EERS Global Technologies through CRITIAS industrial research chair, as well as from McGill University through the CIRMMT Agile Seed Funding, to carry out this research project.

INTRODUCTION

We are inviting you to take part in a research project. However, before accepting to participate in the project and signing this Information and Consent Form, please take the time to read, understand and carefully consider the following information.

This form contains words that you may not understand. We invite you to ask any questions that you feel may be useful to the researcher in charge of the project or to a member of the research team, and to ask them to explain any words or expressions that are unclear to you.

NATURE AND OBJECTIVES OF THE RESEARCH PROJECT

The following research project will study the relationship between sounds inside the ear and emotions. When wearing earpieces that block the ear canal opening, creating an acoustical seal, certain signals such as breathing and the heartbeat are amplified. This is called the occlusion effect. Based on previous research, we think that certain sounds originating in the body, such as a heartbeat, may change according to an individual's emotional state. To test this hypothesis, we will extract biosignals from audio recorded by a microphone inside the ear.

The research project also aims to contribute to a data bank established for the purpose of research and entitled CRITIAS-DB. The goal of this data bank is to group together all data collected by various projects carried out by Professor Jérémie Voix and to make them available to researchers at the École de technologie supérieure and other institutions in an effort to advance knowledge in the field of hearing protection and biomedical science.

To carry out the project, we intend to recruit 30 participants, male and female, aged 18 years or older.

EXECUTION OF THE RESEARCH PROJECT

1. Location and Duration of the Participation

This research project will take place in CRITIAS laboratories at ÉTS. Your participation in the project will last around 2 hours and require one single visit.

2. Nature of your Participation

Information consent form and demographic information (about 15 min)

After you sign the information form, you will be asked to provide information about your age, gender, whether you identify as someone with a physical disability, whether you identify as someone with a mental disability, any medical conditions that affect your autonomic nervous system or your ability to participate in the tasks safely, any known hearing impairments, and any auditory sensitivities. A member of the research team will review your answers. At that point, you may not be able to participate in this project for various reasons. The member of the research team will explain the reasons why you cannot be eligible.

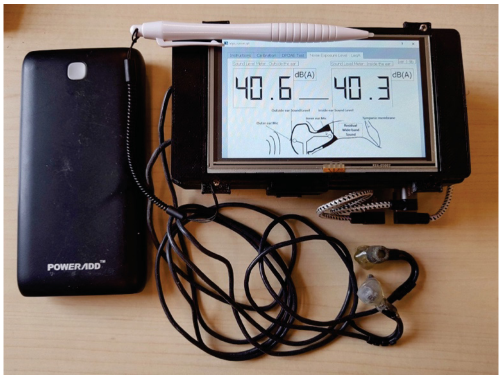
Set-up of the Bioharness (about 10 min)

Following its disinfection, you will be given a Bioharness belt (below) and instructed on how to wear it. This belt is used to obtain the heart and respiratory rate of the wearer.

You will go to a private area in order to put on the Bioharness (i.e. the restroom on the same floor as CRITIAS) and then when you return, a test recording will be done to ensure the quality of the recording.

Set-up and fit-test of the Auditory Research Platform (about 20 min)

You will be using the auditory research platform (ARP), shown below, throughout the test. The ARP earpieces contain an in-ear microphone (IEM), an outer ear microphone (OEM), and a miniature loudspeaker.



Once you enter the audiometric booth, you will be given an auditory research platform that you will be instructed to fit the 2 earpieces into your ears. The experimenter will give you instructions for how to insert the ear tips. After the ARP is fitted, you will be asked to sit still while pink noise is being played and recorded in your ear. The pink noise will be played at a level at 85 dB (A) (which corresponds to the level of a blender at 1m distance) for approximately 30 seconds at a time. This measure allows us to quantify the attenuation obtained with the earpieces. If the attenuation is too low, the experimenter will ask you to try with another type of ear tips. When a good acoustical seal is attained, you will then be instructed to put on several physiological sensors on your hands to monitor your heart rate.

Emotion induction (about 70 min)

First, you will be instructed to relax as the recording begins. In this test, some sound recordings will be captured using both microphones located into the earpieces of the ARP. First, there will be a short period of recording before any stimuli are presented.

Then, different emotions will be induced in several ways, with periods of rest between each of the following steps:

- You will hear different sounds in one ear, with short periods of silence between these sounds. The sounds that you hear will include generated sounds as well as natural sounds heard in daily life (such as a baby laughing, someone cooking, a thunderstorm, erotic sounds, music, etc.). The level of these sounds will not exceed 85 dB(A).
- You will be asked to place your hand in cold water (approximately 6.5 °C) for 3 minutes.
- You will have to complete a task on a computer under time pressure, under two conditions: in quiet and with sporadic noise. As part of the task, you will have to do mental arithmetic (e.g. subtracting a two-digit number from a four-digit number and having to start over with each error).

During the emotion induction sessions, you will be asked to speak so that your voice will be recorded under different emotional conditions. For example, you may be asked to explain the location of a place on a map. You will also be asked to rate your experience of emotion on several scales.

Fit variability check (about 5 min)

A final fit check of the earpiece attenuation will be performed for research purposes. The earpiece will not be adjusted.

At the end of the test, the research team will be available to answer your questions and then you will be escorted out of the lab.

3. Ownership of Data

You will remain the sole owner of your data at all times. It will only be used for the purpose of research and will never be sold.

The researcher in charge of the data bank will act as the trustee of all data in the data bank. He will be responsible for maintaining the data, for its stewardship and its security in accordance with the Data Bank Governance Framework. The researcher will also be responsible for sharing data with researchers who request it.

INCIDENTAL FINDINGS

Although they are not subject to a formal medical assessment since they will be part of a research project, the results of all tests, examinations and procedures performed as part of this research project may reveal problems that were unknown until now, which are referred to as incidental findings. As a result, if a peculiar observation is brought to light, the researcher in charge of the project will inform you to ensure follow-up.

ADVANTAGES ASSOCIATED WITH THE RESEARCH PROJECT

You may gain personal benefit from participating in this research project. However, we cannot ensure that this will be the case. The results obtained will nonetheless contribute to advancing scientific knowledge in this field of research.

INCONVENIENCES ASSOCIATED WITH THE RESEARCH PROJECT

Wearing hearing protectors may cause short-term discomfort for some people. In case of excessive discomfort, you can stop the tests at any time. Putting your hand in cold water can cause redness and pain. In case of excessive pain or skin reaction, you can stop the tests at any time.

The sounds and tasks used in the test are designed to trigger emotions. In case of too strong emotions, you can stop the tests at any time.

RISKS ASSOCIATED WITH THE RESEARCH PROJECT

The noise level used during the tests carried out as part of this study will not exceed 85 dB(A), which corresponds to the noise of a mixer placed at 1 metre. The total exposure time to this noise level for the entire study is less than 60 minutes. If you experience any discomfort during the experiment, the test will be stopped immediately and you can decide whether or not to continue the experiment.

VOLUNTARY PARTICIPATION AND RIGHT TO WITHDRAW

Your participation in this research project is voluntary. You are therefore free to refuse to participate. You may also withdraw from the project at any time without the need to provide any reason, by informing the research team.

Your decision not to participate in the research project or to withdraw from it will carry no consequences for you.

The researcher in charge of the project, the Research Ethics Committee of the École de technologie supérieure or the funding agencies may decide to end your participation without your consent. This can happen if new discoveries or information reveal that your participation in the project is no longer in your interest, or if you do not follow the research project instructions or if there are administrative reasons for abandoning the project.

If you withdraw or are withdrawn from the project, the information and material already collected as part of the project will still be held on file, analyzed or used to ensure the integrity of the project.

All new knowledge acquired during the project that may impact your decision to continue to participate will be shared with you in short order.

CONTRIBUTION, RETENTION, ACCESS TO THE DATA BANK AND CONFIDENTIALITY

During your participation in this project, the researcher in charge of the project and the members of the research team will collect and record the information about you in a research file. They will only collect the information required to achieve the scientific objectives of the project.

This information may include your name, your age, your history of ear surgery, along with the results of all tests and procedures carried out as part of this project.

All of the information collected will remain confidential, within the limits set out under the law. In order to protect your identity and the confidentiality of your information, you will be assigned a code number. The researcher in charge of the research project will keep the key code linking your name to the research file.

All of your information including the recording of your voice, all collected as research data, will be retained in a secure manner in the CRITIAS:DB data bank established for the purpose of research at the École de technologie supérieure, in accordance with the Data Bank Governance Framework.

As a reminder, the CRITIAS:DB data bank established for the purpose of research by Jérémie Voix will allow to develop a global resource to foster collaboration and develop new knowledge in the field of hearing protection and biomedical science.

The research data retained in the CRITIAS:DB data bank will be shared with other researchers. This means that your research data may be transferred in countries other than Canada. However, the researcher in charge of this data bank will follow the confidentiality regulations in effect in Quebec and in Canada, regardless of the country.

All research projects that use the data bank will be evaluated and approved by the Research Ethics Committee of the École de technologie supérieure prior to being carried out. The Committee will also follow up with these projects.

Your research data will be retained as long as it remains useful to the advancement of scientific knowledge. When it is no longer useful, your research data will be destroyed. Please note that at any time, you may request that your research data not be used by contacting the researcher in charge of the project. In such a case, your research data that has already been transferred to other researchers will still be maintained, analyzed or used to protect the integrity of research projects that are already underway and to comply with regulatory requirements. However, your data will not be used for any new projects.

The research data may also be published or used for scientific discussions. However, it will not be possible to identify you.

For the purpose of monitoring, control, protection and security, your research file may be consulted by individuals authorized by regulatory organizations, representatives of the funding agency, the École de technologie supérieure or the Research Ethics Committee. These individuals and organizations are bound by a confidentiality policy.

You have the right to consult your research file to verify the information collected and to modify it as needed.

COMPENSATION

As compensation for your participation in the project, the research team will provide you with a 30$ value pair of musicians' earplugs. If you withdraw from the project, or if your participation is terminated prior to its completion, you will not receive the compensation.

POSSIBILITY OF MARKETING

Your participation in this research project may lead to the creation of commercial products that may eventually be protected under a patent or other intellectual property rights. In such a case, you will not receive any resulting financial benefit.

IN CASE OF PREJUDICE

Should you suffer any prejudice due to your participation in the research project, you will receive all the care and services required by your state of health.

By accepting to participate in this research project, you are not waiving any of your legal rights or releasing the researcher in charge of the project, the École de technologie supérieure and the funding agency from their civil and professional responsibilities.

PROCEDURES IN THE EVENT OF A MEDICAL EMERGENCY

Please note that the École de technologie supérieure does not offer emergency services. Therefore, in the event of a medical condition that requires immediate care, first aid will be provided to you by the personnel on site and arrangements will be made to transfer you, if necessary, to the emergency room of a nearby hospital.

MONITORING OF ETHICAL ASPECTS

The Research Ethics Committee of the École de technologie supérieure approved this project and will ensure follow up.

CONTACT PERSONS

If you have any questions regarding the research project, you can contact the researcher in charge of the project at jvoix@critias.ca. You may also contact Danielle Benesch at Danielle.benesch.1@ens.etsmtl.ca.

For all questions regarding your rights as a participant in the research, you may contact the Research Ethics Committee Coordinator of the École de technologie supérieure by email at CER@etsmtl.ca or by telephone (514) 396-8800 poste 7129.

“Listen to your heart”: Exploring the Link between In-Ear Audio and Emotions

Consent

Participant

I have read this consent form and have been provided all the information and enough time to make my decision. Upon reflection, I voluntarily consent to participating in this research project in keeping with the conditions set out herein.

____

Name of the participant *Signature* *Date*

Person Obtaining Consent (if other than the researcher in charge of the research project)

I explained all relevant aspects of the research to the participant and answered all of the questions that he/she asked.

____

Name of the Individual Obtaining Consent *Signature* *Date*

Signature and Commitment of the Researcher in Charge of the Project

I hereby certify that we have explained this Information and Consent Form to the participant and answered all of their questions.

I agree, along with the research team, to respect everything that was agreed to in the information and consent form and to give a signed copy of this form to the participant.

____

Researcher in charge of the project *Signature* *Date*

LIST OF REFERENCES

Aazh, H., Knipper, M., Danesh, A. A., Cavanna, A. E., Andersson, L., Paulin, J., Schecklmann, M., Heinonen-Guzejev, M. & Moore, B. C. J. Insights from the Third International Conference on Hyperacusis: Causes, Evaluation, Diagnosis, and Treatment. 16.

Aazh, H., Landgrebe, M., Danesh, A. A. & Moore, B. C. (2019). Cognitive Behavioral Therapy For Alleviating The Distress Caused By Tinnitus, Hyperacusis And Misophonia: Current Perspectives. *Psychology Research and Behavior Management*, Volume 12, 991–1002. doi: 10.2147/PRBM.S179138.

Agarwal, V., Shetty, R. & Fritz, M. (2020). Towards Causal VQA: Revealing and Reducing Spurious Correlations by Invariant and Covariant Semantic Editing. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9687–9695. doi: 10.1109/CVPR42600.2020.00971.

Akmandor, A. O. & Jha, N. K. (2017). Keep the Stress Away with SoDA: Stress Detection and Alleviation System. *IEEE Transactions on Multi-Scale Computing Systems*, 3(4), 269–282. doi: 10.1109/TMSCS.2017.2703613.

Alexander, J. (2016). Hearing Aid Delay and Current Drain in Modern Digital Devices. *Canadian Audiologist*, 3(4).

Alexander, J. M., Clavier, O. & Audette, W. (2018). Audiologic Evaluation of the Tympan Open Source Hearing Aid. *The Journal of the Acoustical Society of America*, 143(3), 1736–1736. doi: 10.1121/1.5035665.

Anari, M., Axelsson, A., Eliasson, A. & Magnusson, L. (1999). Hypersensitivity to Sound: Questionnaire Data, Audiometry and Classification. *Scandinavian Audiology*, 28(4), 219–230. doi: 10.1080/010503999424653.

Arai, T. & Greenberg, S. (1998). Speech Intelligibility in the Presence of Cross-Channel Spectral Asynchrony. *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, 2, 933–936. doi: 10.1109/ICASSP.1998.675419.

Arnold, P. & Hill, F. (2001-05, 2001). Bisensory Augmentation: A Speechreading Advantage When Speech Is Clearly Audible and Intact. *British Journal of Psychology*, 92(2), 339–355. doi: 10.1348/000712601162220.

Arquilla, K., Webb, A. K. & Anderson, A. P. (2022). Utility of the Full ECG Waveform for Stress Classification. *Sensors*, 22(18), 7034. doi: 10.3390/s22187034.

Bach, D. R., Schachinger, H., Neuhoff, J. G., Esposito, F., Salle, F. D., Lehmann, C., Herdener, M., Scheffler, K. & Seifritz, E. (2008). Rising Sound Intensity: An Intrinsic Warning Cue Activating the Amygdala. *Cerebral Cortex*, 18(1), 145–150. doi: 10.1093/cercor/bhm040.

Bader, M. & Häussler, J. (2010-03, 2010). Word Order in German: A Corpus Study. *Lingua. International review of general linguistics. Revue internationale de linguistique générale*, 120(3), 717–762. doi: 10.1016/j.lingua.2009.05.007.

Baguley, D. M. & McFerran, D. J. (2011). Hyperacusis and Disorders of Loudness Perception. In Møller, A. R., Langguth, B., De Ridder, D. & Kleinjung, T. (Eds.), *Textbook of Tinnitus* (pp. 13–23). New York, NY: Springer New York. doi: 10.1007/978-1-60761-145-5_3.

Baigi, A., Oden, A., Almlid-Larsen, V., Barrenäs, M.-L. & Holgers, K.-M. (2011). Tinnitus in the General Population With a Focus on Noise and Stress: A Public Health Study. *Ear and Hearing*, 32(6), 787–789.

Balling, L. W., Mosgaard, L. D. & Helmink, D. (2022). Signal Processing and Sound Quality. *The Hearing Review*, 10.

Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. (2013). Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal. *Journal of Memory and Language*, 68(3), 255–278. doi: 10.1016/j.jml.2012.11.001.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheip, F., Grothendieck, G., Green, P. & Fox, J. (2018). Package 'Lme4'.

Bebko, J. M., Weiss, J. A., Demark, J. L. & Gomez, P. (2006). Discrimination of Temporal Synchrony in Intermodal Events by Children with Autism and Children with Developmental Disabilities without Autism. *Journal of Child Psychology and Psychiatry*, 47(1), 88–98. doi: 10.1111/j.1469-7610.2005.01443.x.

Benesch, D., Delain, C., Blain-Moraes, S. & Voix, J. (2020). Listen to Your Heart: Exploring the Link between in-Ear Audio and Emotions. Online.

Benesch, D., Bouserhal, R. & Voix, J. (2021a). Hearables for Managing Auditory Sensitivities Using Heart Rate Variability. *The Journal of the Acoustical Society of America*, 149(4), A19-A19. doi: 10.1121/10.0004400.

Benesch, D., Bouserhal, R. E., Blain-Moraes, S. & Voix, J. (2021b). Managing Auditory Sensitivities in Autism: The Potential of Smart Hearing Protection. Online.

Benesch, D., Raj, K. N., Bouserhal, R. & Voix, J. (2021c). Interfacing the Tympan Open-Source Hearing Aid with an External Computer for Research on Decreased Sound Tolerance. *181st Meeting of the Acoustical Society of America*, pp. 050005. doi: 10.1121/2.0001616.

Benesch, D., Raj, K. N. & Voix, J. (2021d). A General-Purpose Pipeline to Interface the Tympan Hardware with an External Computer. *The Journal of the Acoustical Society of America*, 150(4), A266-A266. doi: 10.1121/10.0008241.

Benesch, D., Orloff, D. & Hansen, H. (2022a). FOAMS: Processed Audio Files (Version v1.0.0). Zenodo.

Benesch, D., Raj, K. N., Bouserhal, R. E. & Voix, J. (2022b). Tympan-PC Pipeline.

Benesch, D., Schwab, J., Voix, J. & Bouserhal, R. (2022c). Open Science Foundation Data Repository: Evaluating the Effects of Audiovisual Delays on Speech Understanding with Hearables.

Berger, E. H. & Voix, J. (2019). Hearing Protection Devices. In Meinke, DK, Berger, EH, Neitzel, R, Driscoll, DP & Hager, LD (Eds.), *The Noise Manual* (ed. 6th Edition). American Industrial Hygiene Association.

Bernal, G., Montgomery, S. M. & Maes, P. (2021). Brain-Computer Interfaces, Open-Source, and Democratizing the Future of Augmented Consciousness. *Frontiers in Computer Science*, 3, 661300. doi: 10.3389/fcomp.2021.661300.

Bhat, G. S., Shankar, N. & Panahi, I. M. S. (2020). Design and Integration of Alert Signal Detector and Separator for Hearing Aid Applications. *IEEE Access*, 8, 106296–106309. doi: 10.1109/ACCESS.2020.2999546.

Bolea, J., Laguna, P., Remartínez, J. M., Rovira, E., Navarro, A. & Bailón, R. (2014). Methodological Framework for Estimating the Correlation Dimension in HRV Signals. *Computational and Mathematical Methods in Medicine*, 2014, 1–11. doi: 10.1155/2014/129248.

Boon, M. Y., Henry, B. I., Suttle, C. M. & Dain, S. J. (2008). The Correlation Dimension: A Useful Objective Measure of the Transient Visual Evoked Potential? *Journal of Vision*, 8(1), 6. doi: 10.1167/8.1.6.

Bou Serhal, R. E., Falk, T. H. & Voix, J. (2013). Integration of a Distance Sensitive Wireless Communication Protocol to Hearing Protectors Equipped with In-Ear Microphones. *Proceedings of Meetings on Acoustics ICA2013*, 19, 040013.

Bouserhal, R. E., Falk, T. H. & Voix, J. (2017). In-Ear Microphone Speech Quality Enhancement via Adaptive Filtering and Artificial Bandwidth Extension. *The Journal of the Acoustical Society of America*, 141(3), 1321–1331. doi: 10.1121/1.4976051.

Bridges, D., Pitiot, A., MacAskill, M. R. & Peirce, J. W. (2020). The Timing Mega-Study: Comparing a Range of Experiment Generators, Both Lab-Based and Online. *PeerJ*, 8, e9414. doi: 10.7717/peerj.9414.

Brindle, R. C., Ginty, A. T., Phillips, A. C., Fisher, J. P., McIntyre, D. & Carroll, D. (2016). Heart Rate Complexity: A Novel Approach to Assessing Cardiac Stress Reactivity: Cardiac Stress Reactivity and Heart Rate Complexity. *Psychophysiology*, 53(4), 465–472. doi: 10.1111/psyp.12576.

Brout, J. J., Edelstein, M., Erfanian, M., Mannino, M., Miller, L. J., Rouw, R., Kumar, S. & Rosenthal, M. Z. (2018). Investigating Misophonia: A Review of the Empirical Literature, Clinical Implications, and a Research Agenda. *Frontiers in Neuroscience*, 12, 36. doi: 10.3389/fnins.2018.00036.

Brugnera, A., Zarbo, C., Tarvainen, M. P., Marchettini, P., Adorni, R. & Compare, A. (2018). Heart Rate Variability during Acute Psychosocial Stress: A Randomized Cross-over Trial of Verbal and Non-Verbal Laboratory Stressors. *International Journal of Psychophysiology*, 127, 17–25. doi: 10.1016/j.ijpsycho.2018.02.016.

Cajal, D., Hernando, D., Lázaro, J., Laguna, P., Gil, E. & Bailón, R. (2022). Effects of Missing Data on Heart Rate Variability Metrics. *Sensors*, 22(15), 5774. doi: 10.3390/s22155774.

Carroll, R. & Ruigendijk, E. (2013). The Effects of Syntactic Complexity on Processing Sentences in Noise. *Journal of Psycholinguistic Research*, 42(2), 139–159. doi: 10.1007/s10936-012-9213-7.

Cavanna, A. E. & Seri, S. (2015). Misophonia: Current Perspectives. *Neuropsychiatric Disease and Treatment*, 2117. doi: 10.2147/NDT.S81438.

Chabot, P., Bouserhal, R. E., Cardinal, P. & Voix, J. (2021). Detection and Classification of Human-Produced Nonverbal Audio Events. *Applied Acoustics*, 171, 107643. doi: 10.1016/j.apacoust.2020.107643.

Chan, Y. M., Pianta, M. J. & McKendrick, A. M. (2014). Reduced Audiovisual Recalibration in the Elderly. *Frontiers in Aging Neuroscience*, 6. doi: 10.3389/fnagi.2014.00226.

Chang, C.-H., Adam, G. A. & Goldenberg, A. (2021). Towards Robust Classification Model by Counterfactual and Invariant Data Generation. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15207–15216. doi: 10.1109/CVPR46437.2021.01496.

Chen, T. & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. doi: 10.1145/2939672.2939785.

Cho, H.-M., Park, H., Dong, S.-Y. & Youn, I. (2019). Ambulatory and Laboratory Stress Detection Based on Raw Electrocardiogram Signals Using a Convolutional Neural Network. *Sensors*, 19(20), 4408. doi: 10.3390/s19204408.

Choi, E. J., Yun, Y., Yoo, S., Kim, K. S., Park, J.-S. & Choi, I. (2013). Autonomic Conditions in Tinnitus and Implications for Korean Medicine. *Evidence-Based Complementary and Alternative Medicine*, 2013, 1–5. doi: 10.1155/2013/402585.

Claiborn, J. M., Dozier, T. H., Hart, S. L. & Lee, J. (2020). Self-Identified Misophonia Phenomenology, Impact, and Clinical Correlates. *Psychological Thought*, 13(2), 349–375. doi: 10.37708/psyct.v13i2.454.

Colagrosso, E. M., Fournier, P., Fitzpatrick, E. M. & Hébert, S. (2019). A Qualitative Study on Factors Modulating Tinnitus Experience. *Ear and hearing*, 40(3), 636–644.

Costa, M., Goldberger, A. L. & Peng, C.-K. (2002). Multiscale Entropy Analysis of Complex Physiologic Time Series. *Physical Review Letters*, 89(6), 068102. doi: 10.1103/PhysRevLett.89.068102.

de Boer-Schellekens, L., Eussen, M. & Vroomen, J. (2013). Diminished Sensitivity of Audiovisual Temporal Order in Autism Spectrum Disorder. *Frontiers in Integrative Neuroscience*, 7. doi: 10.3389/fnint.2013.00008.

Denk, F., Schepker, H., Doclo, S. & Kollmeier, B. (2020). Acoustic Transparency in Hearables - Technical Evaluation. *Journal of the Audio Engineering Society*, 68(7/8), 508–521. doi: 10.17743/jaes.2020.0042.

Dillon, H. (2012). *Hearing Aids* (ed. 2. ed). Sydney: Boomerang Press [u.a.].

Dixon, N. F. & Spitz, L. (1980). The Detection of Auditory Visual Desynchrony. *Perception*, 9(6), 719–721. doi: 10.1068/p090719.

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E. & Brysbaert, M. (2018). MultiPic: A Standardized Set of 750 Drawings with Norms for Six European Languages. *Quarterly Journal of Experimental Psychology*, 71(4), 808–816. doi: 10.1080/17470218.2017.1310261.

Eddins, D. A., Formby, C. C. & Armstrong, S. W. *(2020).*

Edelstein, M., Brang, D., Rouw, R. & Ramachandran, V. S. (2013). Misophonia: Physiological Investigations and Case Descriptions. *Frontiers in Human Neuroscience*, 7. doi: 10.3389/fnhum.2013.00296.

Egger, S., Schatz, R. & Scherer, S. (2010). It Takes Two to Tango — Assessing the Impact of Delay on Conversational Interactivity on Perceived Speech Quality. *Interspeech 2010*, pp. 4.

Enzler, F., Fournier, P. & Noreña, A. J. (2021a). A Psychoacoustic Test for Diagnosing Hyperacusis Based on Ratings of Natural Sounds. *Hearing Research*, 400, 108124. doi: 10.1016/j.heares.2020.108124.

Enzler, F., Loriot, C., Fournier, P. & Noreña, A. J. (2021b). A Psychoacoustic Test for Misophonia Assessment. *Scientific Reports*, 11(1), 11044. doi: 10.1038/s41598-021-90355-8.

Fackrell, K., Stratmann, L., Kennedy, V., MacDonald, C., Hodgson, H., Wray, N., Farrell, C., Meadows, M., Sheldrake, J., Byrom, P., Baguley, D. M., Kentish, R., Chapman, S., Marriage, J., Phillips, J., Pollard, T., Henshaw, H., Gronlund, T. A. & Hoare, D. J. (2019). Identifying and Prioritising Unanswered Research Questions for People with Hyperacusis: James Lind Alliance Hyperacusis Priority Setting Partnership. *BMJ Open*, 9(11), e032178. doi: 10.1136/bmjopen-2019-032178.

Ferdinando, H., Ye, L., Seppänen, T. & Alasaarela, E. (2014). Emotion Recognition by Heart Rate Variability. *Australian Journal of Basic and Applied Sciences*, 7.

Ferrer-Torres, A. & Giménez-Llort, L. (2021). Sounds of Silence in Times of COVID-19: Distress and Loss of Cardiac Coherence in People With Misophonia Caused by Real, Imagined or Evoked Triggering Sounds. *Frontiers in Psychiatry*, 12, 638949. doi: 10.3389/fpsyt.2021.638949.

Fodstad, J. C., Kerswill, S. A., Kirsch, A. C., Lagges, A. & Schmidt, J. (2021). Assessment and Treatment of Noise Hypersensitivity in a Teenager with Autism Spectrum Disorder: A Case Study. *Journal of Autism and Developmental Disorders*, 51(6), 1811–1822. doi: 10.1007/s10803-020-04650-w.

Frank, B. & McKay, D. (2019). The Suitability of an Inhibitory Learning Approach in Exposure When Habituation Fails: A Clinical Application to Misophonia. *Cognitive and Behavioral Practice*, 26(1), 130–142. doi: 10.1016/j.cbpra.2018.04.003.

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A. & Tsiknakis, M. (2019a). Review on Psychological Stress Detection Using Biosignals. *IEEE Transactions on Affective Computing*, 1–1. doi: 10.1109/TAFFC.2019.2927337.

Giannakakis, G., Marias, K. & Tsiknakis, M. (2019b). A Stress Recognition System Using HRV Parameters and Machine Learning Techniques. *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 269–272. doi: 10.1109/ACIIW.2019.8925142.

Goehring, T., Chapman, J. L., Bleeck, S. & Monaghan, J. J. M. (2018). Tolerable Delay for Speech Production and Perception: Effects of Hearing Ability and Experience with Hearing Aids. *International Journal of Audiology*, 57(1), 61–68. doi: 10.1080/14992027.2017.1367848.

Goodson, S. & Hull, R. (2015). Hyperacusis. *American Speech-Language-Hearing Association. Audiology Information Series: Hyperacusis.*

Gordon, P. C., Hendrick, R. & Levine, W. H. (2002). Memory-Load Interference in Syntactic Processing. *Psychological Science*, 13(5), 6.

Gordon-Salant, S., Schwartz, M. S., Oppler, K. A. & Yeni-Komshian, G. H. (2022). Detection and Recognition of Asynchronous Auditory/Visual Speech: Effects of Age, Hearing Loss, and Talker Accent. *Frontiers in Psychology*, 12, 772867. doi: 10.3389/fpsyg.2021.772867.

Grant, K. W. & Bernstein, J. G. W. (2019). Toward a Model of Auditory-Visual Speech Intelligibility. In Lee, A. K. C., Wallace, M. T., Coffin, A. B., Popper, A. N. & Fay, R. R. (Eds.), *Multisensory Processes* (vol. 68, pp. 33–57). Cham: Springer International Publishing. doi: 10.1007/978-3-030-10461-0_3.

Grant, K. W. & Seitz, P. F. (1998). Measures of Auditory–Visual Integration in Nonsense Syllables and Sentences. *The Journal of the Acoustical Society of America*, 104(4), 2438–2450. doi: 10.1121/1.423751.

Grossini, E., Stecco, A., Gramaglia, C., De Zanet, D., Cantello, R., Gori, B., Negroni, D., Azzolina, D., Ferrante, D., Feggi, A., Carriero, A. & Zeppegno, P. (2022). Misophonia: Analysis of the Neuroanatomic Patterns at the Basis of Psychiatric Symptoms and Changes of the Orthosympathetic/ Parasympathetic Balance. *Frontiers in Neuroscience*, 16, 827998. doi: 10.3389/fnins.2022.827998.

Guo, J., Dai, Y., Wang, C., Wu, H., Xu, T. & Lin, K. (2020). A Physiological Data-driven Model for Learners' Cognitive Load Detection Using HRV-PRV Feature Fusion and Optimized XGBoost Classification. *Software: Practice and Experience*, 50(11), 2046–2064. doi: 10.1002/spe.2730.

Gupta, A., Jain, J., Poundrik, S., Shetty, M. K., Girish, M. P. & Gupta, M. D. (2021). Interpretable AI Model-Based Predictions of ECG Changes in COVID-recovered Patients. *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*, pp. 1–5. doi: 10.1109/BioSMART54244.2021.9677747.

Han, L., Zhang, Q., Chen, X., Zhan, Q., Yang, T. & Zhao, Z. (2017). Detecting Work-Related Stress with a Wearable Device. *Computers in Industry*, 90, 42–49. doi: 10.1016/j.compind.2017.05.004.

Hansen, H. A., Leber, A. B. & Saygin, Z. M. (2021). What Sound Sources Trigger Misophonia? Not Just Chewing and Breathing. *Journal of Clinical Psychology*, 77(11), 2609–2625. doi: 10.1002/jclp.23196.

Hansen, J. H., Ali, H., Saba, J. N., Charan, M. C. R., Mamun, N., Ghosh, R. & Brueggeman, A. (2019). CCi-MOBILE: Design and Evaluation of a Cochlear Implant and Hearing Aid Research Platform for Speech Scientists and Engineers. *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 1–4. doi: 10.1109/BHI.2019.8834652.

Hasson, D., Theorell, T., Bergquist, J. & Canlon, B. (2013). Acute Stress Induces Hyperacusis in Women with High Levels of Emotional Exhaustion. *PLoS ONE*, 8(1), e52945. doi: 10.1371/journal.pone.0052945.

Hébert, S. & Lupien, S. J. (2009). Salivary Cortisol Levels, Subjective Stress, and Tinnitus Intensity in Tinnitus Sufferers during Noise Exposure in the Laboratory. *International Journal of Hygiene and Environmental Health*, 212(1), 37–44. doi: 10.1016/j.ijheh.2007.11.005.

Hébert, S., Canlon, B. & Hasson, D. (2012). Emotional Exhaustion as a Predictor of Tinnitus. *Psychotherapy and Psychosomatics*, 81(5), 324–326. doi: 10.1159/000335043.

Herzke, T., Kayser, H., Loshaj, F., Grimm, G. & Hohmann, V. Open Signal Processing Software Platform for Hearing Aid Research (openMHA). 8.

Higuchi, T. (1988). Approach to an Irregular Time Series on the Basis of the Fractal Theory. *Physica D: Nonlinear Phenomena*, 31(2), 277–283. doi: 10.1016/0167-2789(88)90081-4.

Hoffman, M. A New Generational Approach: Integrating Psychology and Audiology Care. 14.

Hosseini, M., Powell, M., Collins, J., Callahan-Flintoft, C., Jones, W., Bowman, H. & Wyble, B. (2020). I Tried a Bunch of Things: The Dangers of Unexpected Overfitting in Classification of Brain Data. *Neuroscience & Biobehavioral Reviews*, 119, 456–467. doi: 10.1016/j.neubiorev.2020.09.036.

Hovsepian, K., al'Absi, M., Ertin, E., Kamarck, T., Nakajima, M. & Kumar, S. (2015). cStress: Towards a Gold Standard for Continuous Stress Assessment in the Mobile Environment. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, pp. 493–504. doi: 10.1145/2750858.2807526.

Ishikawa, A. & Mieno, H. (1979). The Fuzzy Entropy Concept and Its Application. *Fuzzy Sets and Systems*, 2(2), 113–123. doi: 10.1016/0165-0114(79)90020-4.

Jager, I., de Koning, P., Bost, T., Denys, D. & Vulink, N. (2020). Misophonia: Phenomenology, Comorbidity and Demographics in a Large Sample. *PLOS ONE*, 15(4), e0231390. doi: 10.1371/journal.pone.0231390.

Jastreboff, P. & Jastreboff, M. (2014). Treatments for Decreased Sound Tolerance (Hyperacusis and Misophonia). *Seminars in Hearing*, 35(02), 105–120. doi: 10.1055/s-0034-1372527.

Jastreboff, P. J. (2004). The Neurophysiological Model of Tinnitus. In *Tinnitus: Theory and Management* (pp. 15). B. C. Decker.

Jastreboff, P. J. & Jastreboff, M. M. (2015). Decreased Sound Tolerance. In *Handbook of Clinical Neurology* (vol. 129, pp. 375–387). Elsevier. doi: 10.1016/B978-0-444-62630-1.00021-4.

Jeppesen, J., Beniczky, S., Johansen, P., Sidenius, P. & Fuglsang-Frederiksen, A. (2014). Using Lorenz Plot and Cardiac Sympathetic Index of Heart Rate Variability for Detecting Seizures for Patients with Epilepsy. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4563–4566. doi: 10.1109/EMBC.2014.6944639.

Johansen, B., Flet-Berliac, Y. P. R., Korzepa, M. J., Sandholm, P., Pontoppidan, N. H., Petersen, M. K. & Larsen, J. E. (2017). Hearables in Hearing Care: Discovering Usage Patterns Through IoT Devices. In Antona, M. & Stephanidis, C. (Eds.), *Universal Access in Human–Computer Interaction. Human and Technological Environments* (vol. 10279, pp. 39–49). Cham: Springer International Publishing. doi: 10.1007/978-3-319-58700-4_4.

Jongyoon Choi, Ahmed, B. & Gutierrez-Osuna, R. (2012). Development and Evaluation of an Ambulatory Stress Monitor Based on Wearable Sensors. *IEEE Transactions on Information Technology in Biomedicine*, 16(2), 279–286. doi: 10.1109/TITB.2011.2169804.

Katz, M. J. (1988). Fractals and the Analysis of Waveforms. *Computers in biology and medicine*, 18(3), 145–156.

Kaur, B., Durek, J. J., O'Kane, B. L., Tran, N., Moses, S., Luthra, M. & Ikonomidou, V. N. (2014). Heart Rate Variability (HRV): An Indicator of Stress. *SPIE Sensing Technology + Applications*, pp. 91180V. doi: 10.1117/12.2051148.

Keith, J. M., Jamieson, J. P. & Bennetto, L. (2019). The Influence of Noise on Autonomic Arousal and Cognitive Performance in Adolescents with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 49(1), 113–126. doi: 10.1007/s10803-018-3685-8.

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G. & King, D. (2019). Key Challenges for Delivering Clinical Impact with Artificial Intelligence. *BMC Medicine*, 17(1), 195. doi: 10.1186/s12916-019-1426-2.

Kemeny, M. E. (2003). The Psychobiology of Stress. *Current Directions in Psychological Science*, 12(4), 124–129. doi: 10.1111/1467-8721.01246.

Khalfa, S., Dubal, S., Veuillet, E., Perez-Diaz, F., Jouvent, R. & Collet, L. (2002). Psychometric Normalization of a Hyperacusis Questionnaire. *ORL*, 64(6), 436–442. doi: 10.1159/000067570.

Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H. & Koo, B.-H. (2018). Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investigation*, 15(3), 235–245. doi: 10.30773/pi.2017.08.17.

Kirschbaum, C., Pirke, K.-M. & Hellhammer, D. (1993). The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, 28(1-2), 76–81. doi: 10.1159/000119004.

Kuiper, M. W. M., Verhoeven, E. W. M. & Geurts, H. M. (2019). Stop Making Noise! Auditory Sensitivity in Adults with an Autism Spectrum Disorder Diagnosis: Physiological Habituation and Subjective Detection Thresholds. *Journal of Autism and Developmental Disorders*, 49(5), 2116–2128. doi: 10.1007/s10803-019-03890-9.

Kumar, S., Tansley-Hancock, O., Sedley, W., Winston, J. S., Callaghan, M. F., Allen, M., Cope, T. E., Gander, P. E., Bamiou, D.-E. & Griffiths, T. D. (2017). The Brain Basis for Misophonia. *Current Biology*, 27(4), 527–533. doi: 10.1016/j.cub.2016.12.048.

Kyle, F. E., Campbell, R., Mohammed, T., Coleman, M. & MacSweeney, M. (2013). Speechreading Development in Deaf and Hearing Children: Introducing the Test of Child Speechreading. *Journal of Speech, Language, and Hearing Research*, 56(2), 416–426. doi: 10.1044/1092-4388(2012/12-0039).

Laborde, S., Mosley, E. & Thayer, J. F. (2017). Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research – Recommendations for Experiment Planning, Data Analysis, and Data Reporting. *Frontiers in Psychology*, 08. doi: 10.3389/fpsyg.2017.00213.

Lempel, A. & Ziv, J. (1976). On the Complexity of Finite Sequences. *IEEE Transactions on Information Theory*, 22(1), 75–81. doi: 10.1109/TIT.1976.1055501.

Lentz, J., Yun, D. & Smiley, S. (2021). Measuring Hearing Aid Compression Algorithm Preference with the Tympan. *The Journal of the Acoustical Society of America*, 150(4), A265–A266.

Lezzoum, N., Gagnon, G. & Voix, J. (2013). A Low-Complexity Voice Activity Detector for Smart Hearing Protection of Hyperacusic Persons. *Interspeech*, pp. 4–8.

Lezzoum, N., Gagnon, G. & Voix, J. (2016). Echo Threshold between Passive and Electro-Acoustic Transmission Paths in Digital Hearing Protection Devices. *International Journal of Industrial Ergonomics*, 53, 372–379. doi: http://dx.doi.org/10.1016/j.ergon.2016.04.004.

Liew, W. S., Loo, C. K. & Wermter, S. (2021). Emotion Recognition Using Explainable Genetically Optimized Fuzzy ART Ensembles. *IEEE Access*, 9, 61513–61531. doi: 10.1109/ACCESS.2021.3072120.

Lovallo, W. (1975). The Cold Pressor Test and Autonomic Function: A Review and Integration. *Psychophysiology*, 12(3), 268–282. doi: 10.1111/j.1469-8986.1975.tb01289.x.

Lusk, S. L., Gillespie, B., Hagerty, B. M. & Ziemba, R. A. (2004). Acute Effects of Noise on Blood Pressure and Heart Rate. *Archives of Environmental Health: An International Journal*, 59(8), 392–399. doi: 10.3200/AEOH.59.8.392-399.

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C. & Chen, S. H. A. (2021). NeuroKit2: A Python Toolbox for Neurophysiological Signal Processing. *Behavior Research Methods*, 53(4), 1689–1696. doi: 10.3758/s13428-020-01516-y.

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J. & Schwartz, P. J. (1996). Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. *European heart journal*, 17(3), 354–381.

Martin, A. & Voix, J. (2018). In-Ear Audio Wearable: Measurement of Heart and Breathing Rates for Health and Safety Monitoring. *IEEE Transactions on Biomedical Engineering*, 65(6), 1256–1263. doi: 10.1109/TBME.2017.2720463.

Masino, A. J., Forsyth, D., Nuske, H., Herrington, J., Pennington, J., Kushleyeva, Y. & Bonafide, C. P. (2019). M-Health and Autism: Recognizing Stress and Anxiety with Machine Learning and Wearables Data. *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 714–719. doi: 10.1109/CBMS.2019.00144.

MATLAB. (2019). *Version 9.7 (R2019b)*. Natick, Massachusetts: The MathWorks Inc.

Mazurek, B., Szczepek, A. & Hebert, S. (2015). Stress and Tinnitus. *HNO*, 63(4), 258–265. doi: 10.1007/s00106-014-2973-7.

Mazurek, B., Boecking, B. & Brueggemann, P. (2019). Association Between Stress and Tinnitus—New Aspects. *Otology & Neurotology*, 40(4), e467-e473. doi: 10.1097/MAO.0000000000002180.

Melillo, P., Bracale, M. & Pecchia, L. (2011). Nonlinear Heart Rate Variability Features for Real-Life Stress Detection. Case Study: Students under Stress Due to University Examination. *BioMedical Engineering OnLine*, 10(1), 96. doi: 10.1186/1475-925X-10-96.

Miller, R. L. & Donahue, A. (2014). *Open Speech Signal Processing Platform Workshop | NIDCD*. Bethesda, MD, USA.

Mishra, V., Pope, G., Lord, S., Lewia, S., Lowens, B., Caine, K., Sen, S., Halter, R. & Kotz, D. (2018). The Case for a Commodity Hardware Solution for Stress Detection. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pp. 1717–1728. doi: 10.1145/3267305.3267538.

Mishra, V., Sen, S., Chen, G., Hao, T., Rogers, J., Chen, C.-H. & Kotz, D. (2020). Evaluating the Reproducibility of Physiological Stress Detection Models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4), 1–29. doi: 10.1145/3432220.

Molnar, C. (2020). *Interpretable Machine Learning*. Lulu. com.

Momeni, N., Arza, A., Rodrigues, J., Sandi, C. & Atienza, D. (2021). CAFS: Cost-Aware Features Selection Method for Multimodal Stress Monitoring on Wearable Devices. *IEEE Transactions on Biomedical Engineering*, 1–1. doi: 10.1109/TBME.2021.3113593.

Monaco, A., Cattaneo, R., Ortu, E., Constantinescu, M. V. & Pietropaoli, D. (2017). Sensory Trigeminal ULF-TENS Stimulation Reduces HRV Response to Experimentally Induced Arithmetic Stress: A Randomized Clinical Trial. *Physiology & Behavior*, 173, 209–215. doi: 10.1016/j.physbeh.2017.02.014.

Moshgelani, F., Parsa, V., Allan, C., Veeranna, S. A. & Allen, P. (2020). Perceptual and Objective Assessment of Envelope Enhancement for Children With Auditory Processing Disorder. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(1), 143–151. doi: 10.1109/TNSRE.2019.2957230.

Mourot, L., Bouhaddi, M. & Regnard, J. (2009). Effects of the Cold Pressor Test on Cardiac Autonomic Control in Normal Subjects. *Physiological Research*, 83–91. doi: 10.33549/physiolres.931360.

Nadon, V., Bonnet, F., Bouserhal, R. E., Bernier, A. & Voix, J. (2021). Method for Protected Noise Exposure Level Assessment under an In-Ear Hearing Protection Device: A Pilot Study. *International Journal of Audiology*, 60(1), 60–69. doi: 10.1080/14992027.2020.1799082.

Nardelli, M., Valenza, G., Greco, A., Lanata, A. & Scilingo, E. P. (2015). Recognizing Emotions Induced by Affective Sounds through Heart Rate Variability. *IEEE Transactions on Affective Computing*, 6(4), 385–394. doi: 10.1109/TAFFC.2015.2432810.

Navarra, J., Alsius, A., Velasco, I., Soto-Faraco, S. & Spence, C. (2010). Perception of Audiovisual Speech Synchrony for Native and Non-Native Language. *Brain Research*, 1323, 84–93. doi: 10.1016/j.brainres.2010.01.059.

Neacsiu, A. D., Szymkiewicz, V., Galla, J. T., Li, B., Kulkarni, Y. & Spector, C. W. (2022). The Neurobiology of Misophonia and Implications for Novel, Neuroscience-Driven Interventions. *Frontiers in Neuroscience*, 16, 893903. doi: 10.3389/fnins.2022.893903.

Neave-DiToro, D., Fuse, A. & Bergen, M. (2021). Knowledge and Awareness of Ear Protection Devices for Sound Sensitivity by Individuals With Autism Spectrum Disorders. *Language, Speech, and Hearing Services in Schools*, 52(1), 409–425. doi: 10.1044/2020_LSHSS-19-00119.

Nikolic-Popovic, J. & Goubran, R. (2013). Impact of Motion Artifacts on Heart Rate Variability Measurements and Classification Performance. *2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 156–159. doi: 10.1109/MeMeA.2013.6549726.

Noel, J.-P., De Niear, M. A., Stevenson, R., Alais, D. & Wallace, M. T. (2017). Atypical Rapid Audio-Visual Temporal Recalibration in Autism Spectrum Disorders: Audiovisual Temporal Recalibration in ASD. *Autism Research*, 10(1), 121–129. doi: 10.1002/aur.1633.

Noreña, A. J. & Chery-Croze, S. (2007). Enriched Acoustic Environment Rescales Auditory Sensitivity. *NeuroReport*, 18(12).

Ollander, S., Godin, C., Campagne, A. & Charbonnier, S. (2016). A Comparison of Wearable and Stationary Sensors for Stress Detection. *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 004362–004366. doi: 10.1109/SMC.2016.7844917.

Opoku Asare, K., Moshe, I., Terhorst, Y., Vega, J., Hosio, S., Baumeister, H., Pulkki-Råback, L. & Ferreira, D. (2022). Mood Ratings and Digital Biomarkers from Smartphone and Wearable Data Differentiates and Predicts Depression Status: A Longitudinal Data Analysis. *Pervasive and Mobile Computing*, 83, 101621. doi: 10.1016/j.pmcj.2022.101621.

Orloff, D., Benesch, D. & Hansen, H. (2022). Curation of FOAMS: A Free Open-Access Misophonia Stimuli Database. Zenodo.

Orozco-Duque, A., Novak, D., Kremen, V. & Bustamante, J. (2015). Multifractal Analysis for Grading Complex Fractionated Electrograms in Atrial Fibrillation. *Physiological Measurement*, 36(11), 2269–2284. doi: 10.1088/0967-3334/36/11/2269.

Pagé, M. G., Dassieu, L., Develay, E., Roy, M., Vachon-Presseau, E., Lupien, S. & Rainville,PhD, P. (2021). The Stressful Characteristics of Pain That Drive You NUTS: A Qualitative Exploration of a Stress Model to Understand the Chronic Pain Experience. *Pain Medicine*, 22(5), 1095–1108. doi: 10.1093/pm/pnaa370.

Pandey, P. C., Kunov, H. & Abel, S. M. (1986). Disruptive Effects of Auditory Signal Delay on Speech Perception with Lipreading. *Journal of Auditory Research*.

Parent, M., Tiwari, A., Albuquerque, I., Gagnon, J.-F., Lafond, D., Tremblay, S. & Falk, T. H. (2019). A Multimodal Approach to Improve the Robustness of Physiological Stress Prediction During Physical Activity. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 4131–4136. doi: 10.1109/SMC.2019.8914254.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A. & Cournapeau, D. (2011). Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, 6.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E. & Lindeløv, J. K. (2019). PsychoPy2: Experiments in Behavior Made Easy. *Behavior research methods*, 51(1), 195–203. doi: 10.3758/s13428-018-01193-y.

Pfeiffer, B., Erb, S. R. & Slugg, L. (2019a). Impact of Noise-Attenuating Headphones on Participation in the Home, Community, and School for Children with Autism Spectrum Disorder. *Physical & Occupational Therapy In Pediatrics*, 39(1), 60–76. doi: 10.1080/01942638.2018.1496963.

Pfeiffer, B., Stein Duker, L., Murphy, A. & Shui, C. (2019b). Effectiveness of Noise-Attenuating Headphones on Physiological Responses for Children With Autism Spectrum Disorders. *Frontiers in Integrative Neuroscience*, 13, 65. doi: 10.3389/fnint.2019.00065.

Pham, T., Lau, Z. J., Chen, S. H. A. & Makowski, D. (2021a). Heart Rate Variability in Psychology: A Review of HRV Indices and an Analysis Tutorial. *Sensors*, 21(12), 3998. doi: 10.3390/s21123998.

Pham, T., Lau, Z. J., Chen, S. A. & Makowski, D. (2021b). *Unveiling the Structure of Heart Rate Variability (HRV) Indices: A Data-driven Meta-clustering Approach*.

Pienkowski, M., Tyler, R. S., Roncancio, E. R., Jun, H. J., Brozoski, T., Dauman, N., Coelho, C. B., Andersson, G., Keiner, A. J., Cacace, A. T., Martin, N. & Moore, B. C. J. (2014). A Review of Hyperacusis and Future Directions: Part II. Measurement, Mechanisms, and Treatment. *American Journal of Audiology*, 23(4), 420–436. doi: 10.1044/2014_AJA-13-0037.

Pisha, L., Warchall, J., Zubatiy, T., Hamilton, S., Lee, C.-H., Chockalingam, G., Mercier, P. P., Gupta, R., Rao, B. D. & Garudadri, H. (2019). A Wearable, Extensible, Open-Source Platform for Hearing Healthcare Research. *IEEE Access*, 7, 162083–162101. doi: 10.1109/ACCESS.2019.2951145.

Piskorski, J. & Guzik, P. (2011). Asymmetric Properties of Long-Term and Total Heart Rate Variability. *Medical & Biological Engineering & Computing*, 49(11), 1289–1297. doi: 10.1007/s11517-011-0834-z.

Plesa Skwerer, D., Jordan, S. E., Brukilacchio, B. H. & Tager-Flusberg, H. (2016). Comparing Methods for Assessing Receptive Language Skills in Minimally Verbal Children and Adolescents with Autism Spectrum Disorders. *Autism*, 20(5), 591–604. doi: 10.1177/1362361315600146.

Poore-Pariseau, C. (2019). Misophonia: A Connection between Sounds and Emotions? *Disability Compliance for Higher Education*, 25(1), 4–5. doi: 10.1002/dhe.30687.

Potgieter, I., MacDonald, C., Partridge, L., Cima, R., Sheldrake, J. & Hoare, D. J. (2019). Misophonia: A Scoping Review of Research. *Journal of Clinical Psychology*, 75(7), 1203–1218. doi: 10.1002/jclp.22771.

Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Wortman Vaughan, J. W. & Wallach, H. (2021). Manipulating and Measuring Model Interpretability. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–52. doi: 10.1145/3411764.3445315.

Prajod, P. & André, E. *(2022)*. On the Generalizability of ECG-based Stress Detection Models. arXiv.

Pulopulos, M. M., Vanderhasselt, M.-A. & De Raedt, R. (2018). Association between Changes in Heart Rate Variability during the Anticipation of a Stressful Situation and the Stress-Induced Cortisol Response. *Psychoneuroendocrinology*, 94, 63–71. doi: 10.1016/j.psyneuen.2018.05.004.

R Core Team. (2019). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Raj-Koziak, D., Gos, E., Kutyba, J., Skarzynski, H. & Skarzynski, P. H. (2021). Decreased Sound Tolerance in Tinnitus Patients. 12.

Reinhart, P., Griffin, K. & Micheyl, C. (2021). Changes in Heart Rate Variability Following Acoustic Therapy in Individuals With Tinnitus. *Journal of Speech, Language, and Hearing Research*, 1–7. doi: 10.1044/2021_JSLHR-20-00596.

Reisberg, D., Mclean, J. & Goldfield, A. (1987). Easy to Hear but Hard to Understand: A Lip-Reading Advantage with Intact Auditory Stimuli. In *Hearing by Eye: The Psychology of Lip-Reading* (pp. 97–113). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Rosemann, S. & Thiel, C. M. (2018). Audio-Visual Speech Processing in Age-Related Hearing Loss: Stronger Integration and Increased Frontal Lobe Recruitment. *NeuroImage*, 175, 425–437. doi: 10.1016/j.neuroimage.2018.04.023.

Rowe, C., Candler, C. & Neville, M. (2011). Noise Reduction Headphones and Autism: A Single Case Study. *Journal of Occupational Therapy, Schools, & Early Intervention*, 4(3-4), 229–235. doi: 10.1080/19411243.2011.629551.

Sabeti, M., Katebi, S. & Boostani, R. (2009). Entropy and Complexity Measures for EEG Signal Classification of Schizophrenic and Control Participants. *Artificial Intelligence in Medicine*, 47(3), 263–274. doi: 10.1016/j.artmed.2009.03.003.

Sadeghi, M., McDonald, A. D. & Sasangohar, F. (2021). Posttraumatic Stress Disorder Hyperarousal Event Detection Using Smartwatch Physiological and Activity Data. *arXiv:2109.14743 [cs]*.

Sammeth, C. A., Preves, D. A. & Brandy, W. T. (2000). Hyperacusis: Case Studies and Evaluation of Electronic Loudness Suppression Devices as a Treatment Approach. *Scandinavian Audiology*, 29(1), 28–36. doi: 10.1080/010503900424570.

Sander, E. L. J., Marques, C., Birt, J., Stead, M. & Baumann, O. (2021). Open-Plan Office Noise Is Stressful: Multimodal Stress Detection in a Simulated Work Environment. *Journal of Management & Organization*, 1–17. doi: 10.1017/jmo.2021.17.

Schaaf, H., Klofat, B. & Hesse, G. (2003). Hyperakusis, Phonophobie und Recruitment. *HNO*, 51(12), 1005–1011. doi: 10.1007/s00106-003-0967-y.

Schlee, W., Kraft, R., Schobel, J., Langguth, B., Probst, T., Lourenco, M. P., Simoes, J., Neff, P., Hannemann, R. & Reichert, M. (2023). Momentary Assessment of Tinnitus—How Smart Mobile Applications Advance Our Understanding of Tinnitus. In *Digital Phenotyping and Mobile Sensing* (pp. 285–303). Springer.

Schmithorst, V. J., Holland, S. K. & Plante, E. (2007). Object Identification and Lexical/Semantic Access in Children: A Functional Magnetic Resonance Imaging Study of Word-Picture Matching. *Human Brain Mapping*, 28(10), 1060–1074. doi: 10.1002/hbm.20328.

Schröder, A., Vulink, N. & Denys, D. (2013). Misophonia: Diagnostic Criteria for a New Psychiatric Disorder. *PLoS ONE*, 8(1), e54706. doi: 10.1371/journal.pone.0054706.

Schröder, A., van Wingen, G., Eijsker, N., San Giorgi, R., Vulink, N. C., Turbyne, C. & Denys, D. (2019). Misophonia Is Associated with Altered Brain Activity in the Auditory Cortex and Salience Network. *Scientific Reports*, 9(1), 7542. doi: 10.1038/s41598-019-44084-8.

Schröder, A. E., Vulink, N. C., van Loon, A. J. & Denys, D. A. (2017). Cognitive Behavioral Therapy Is Effective in Misophonia: An Open Trial. *Journal of Affective Disorders*, 217, 289–294. doi: 10.1016/j.jad.2017.04.017.

Searchfield, G. D., Sanders, P. J., Doborjeh, Z., Doborjeh, M., Boldu, R., Sun, K. & Barde, A. (2021). A State-of-Art Review of Digital Technologies for the Next Generation of Tinnitus Therapeutics. *Frontiers in Digital Health*, 3, 724370. doi: 10.3389/fdgth.2021.724370.

Shaffer, F. & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, 5, 258. doi: 10.3389/fpubh.2017.00258.

Sheldrake, J., Diehl, P. U. & Schaette, R. (2015). Audiometric Characteristics of Hyperacusis Patients. *Frontiers in Neurology*, 6. doi: 10.3389/fneur.2015.00105.

Shepherd, D. (2016). Electrophysiological Approaches to Noise Sensitivity. *Journal of Clinical and Experimental Neuropsychology*, 14.

Shi, B., Zhang, Y., Yuan, C., Wang, S. & Li, P. (2017). Entropy Analysis of Short-Term Heartbeat Interval Time Series during Regular Walking. *Entropy*, 19(10), 568. doi: 10.3390/e19100568.

Shi, K., Steigleder, T., Schellenberger, S., Michler, F., Malessa, A., Lurz, F., Rohleder, N., Ostgathe, C., Weigel, R. & Koelpin, A. (2021). Contactless Analysis of Heart Rate Variability during Cold Pressor Test Using Radar Interferometry and Bidirectional LSTM Networks. *Scientific Reports*, 11(1), 3025. doi: 10.1038/s41598-021-81101-1.

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X. & Yang, X. (2018). A Review of Emotion Recognition Using Physiological Signals. *Sensors*, 18(7), 2074. doi: 10.3390/s18072074.

Smith, E. E., Guzick, A. G., Draper, I. A., Clinger, J., Schneider, S. C., Goodman, W. K., Brout, J. J., Lijffijt, M. & Storch, E. A. (2022). Perceptions of Various Treatment Approaches for Adults and Children with Misophonia. *Journal of Affective Disorders*, 316, 76–82. doi: 10.1016/j.jad.2022.08.020.

Smith, G. W. & Riccomini, P. J. (2013). The Effect of a Noise Reducing Test Accommodation on Elementary Students with Learning Disabilities. *Learning Disabilities Research & Practice*, 28(2), 89–95. doi: 10.1111/ldrp.12010.

Smith, L. M., Wang, L., Mazur, K., Carchia, M., DePalma, G., Azimi, R., Mravca, S. & Neitzel, R. L. (2020). Impacts of COVID-19-related Social Distancing Measures on Personal Environmental Sound Exposures. *Environmental Research Letters*, 15(10), 104094. doi: 10.1088/1748-9326/abb494.

Stiegler, L. N. & Davis, R. (2010). Understanding Sound Sensitivity in Individuals with Autism Spectrum Disorders. *Focus on Autism and Other Developmental Disabilities*, 25(2), 67–75. doi: 10.1177/1088357610364530.

Stone, M. A. & Moore, B. C. J. (2003). Tolerable Hearing Aid Delays. III. Effects on Speech Production and Perception of Across-Frequency Variation in Delay:. *Ear and Hearing*, 24(2), 175–183. doi: 10.1097/01.AUD.0000058106.68049.9C.

Subramaniam, S. D. & Dass, B. (2022). An Efficient Convolutional Neural Network for Acute Pain Recognition Using HRV Features. *Proceedings of the International E-Conference on Intelligent Systems and Signal Processing*, pp. 119–132.

Sumby, W. H. & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, 26(2), 212–215. doi: 10.1121/1.1907309.

Summerfield, Q. (1992). Lipreading and Audio-Visual Speech Perception. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 335(1273), 71–78. doi: 10.1098/rstb.1992.0009.

Sun, F.-T., Kuo, C., Cheng, H.-T., Buthpitiya, S., Collins, P. & Griss, M. L. (2012). Activity-Aware Mental Stress Detection Using Physiological Sensors. *International Conference on Mobile Computing, Applications, and Services*, pp. 20.

Suzuki, K., Laohakangvalvit, T., Matsubara, R. & Sugaya, M. (2021). Constructing an Emotion Estimation Model Based on EEG/HRV Indexes Using Feature Extraction and Feature Selection Algorithms. *Sensors*, 21(9), 2910. doi: 10.3390/s21092910.

Swedo, S., Baguley, D. M., Denys, D., Dixon, L. J., Erfanian, M., Fioretti, A., Jastreboff, P. J., Kumar, S., Rosenthal, M. Z., Rouw, R., Schiller, D., Simner, J., Storch, E. A., Taylor, S., Werff, K. R. V. & Raver, S. M. (2021). A Consensus Definition of Misophonia: Using a Delphi Process to Reach Expert Agreement. doi: 10.1101/2021.04.05.21254951.

Szakonyi, B., Vassányi, I., Schumacher, E. & Kósa, I. (2021). Efficient Methods for Acute Stress Detection Using Heart Rate Variability Data from Ambient Assisted Living Sensors. *BioMedical Engineering OnLine*, 20(1), 73. doi: 10.1186/s12938-021-00911-6.

Talay-Ongan, A. & Wood, K. (2000). Unusual Sensory Sensitivities in Autism: A Possible Crossroads. *International Journal of Disability, Development and Education*, 47(2), 201–212. doi: 10.1080/713671112.

Toichi, M., Sugiura, T., Murai, T. & Sengoku, A. (1997). A New Method of Assessing Cardiac Autonomic Function and Its Comparison with Spectral Analysis and Coefficient of Variation of R–R Interval. *Journal of the Autonomic Nervous System*, 62(1-2), 79–84. doi: 10.1016/S0165-1838(96)00112-9.

Tomar, S. (2006). Converting Video Formats with FFmpeg. *Linux Journal*, 2006(146), 10.

Traynor, R. (2021). Reducing Dependence on Attenuation for Decreased Sound Tolerance. *Canadian Audiologist*, 8(5), 4.

Tsutsui da Silva, L., Souza, V. M. A. & Batista, G. E. A. P. A. (2022). An Open-Source Tool for Classification Models in Resource-Constrained Hardware. *IEEE Sensors Journal*, 22(1), 544–554. doi: 10.1109/JSEN.2021.3128130.

Uslar, V. N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Brand, T. & Kollmeier, B. (2013). Development and Evaluation of a Linguistically and Audiologically Controlled Sentence Intelligibility Test. *The Journal of the Acoustical Society of America*, 134(4), 3039–3056. doi: 10.1121/1.4818760.

Valente, M., Goebel, J., Duddy, D., Sinks, B. & Peterin, J. (2000). Evaluation and Treatment of Severe Hyperacusis. *Journal of the American Academy of Audiology*, 11(6), 6.

Villescas, M. R., de Vries, B., Stuijk, S. & Corporaal, H. (2020). Real-Time Audio Processing for Hearing Aids Using a Model-Based Bayesian Inference Framework. *Proceedings of the 23th International Workshop on Software and Compilers for Embedded Systems*, pp. 82–85. doi: 10.1145/3378678.3397528.

Visnovcova, Z., Mestanik, M., Javorka, M., Mokra, D., Gala, M., Jurko, A., Calkovska, A. & Tonhajzerova, I. (2014). Complexity and Time Asymmetry of Heart Rate Variability Are Altered in Acute Mental Stress. *Physiological Measurement*, 35(7), 1319–1334. doi: 10.1088/0967-3334/35/7/1319.

Voix, J. (2014). Did You Say "Bionic" Ear? *Canadian Acoustics*, 42(3).

Voix, J. & Laville, F. (2002). Expandable Earplug with Smart Custom Fitting Capabilities. *2002*, Vol. 111,, pp. 833–841.

Voix, J. & Laville, F. (2009). The Objective Measurement of Individual Earplug Field Performance. *The Journal of the Acoustical Society of America*, 125(6), 3722–3732. doi: 10.1121/1.3125769.

Wagenknecht, A. (2018). Measuring Audio Latency.

Wang, Y., Zhang, X., Chakalasiya, J. M., Xu, X., Jiang, Y., Li, Y., Patel, S. & Shi, Y. (2022). HearCough: Enabling Continuous Cough Event Detection on Edge Computing Hearables. *Methods*, 205, 53–62. doi: 10.1016/j.ymeth.2022.05.002.

Waye, K. P., Bengtsson, J., Rylander, R., Hucklebridge, F., Evans, P. & Clow, A. (2002). Low Frequency Noise Enhances Cortisol among Noise Sensitive Subjects during Work Performance. *Life Sciences*, 70(7), 745–758. doi: 10.1016/S0024-3205(01)01450-3.

Weatherhead, D., Arredondo, M. M., Nácar Garcia, L. & Werker, J. F. (2021). The Role of Audiovisual Speech in Fast-Mapping and Novel Word Retention in Monolingual and Bilingual 24-Month-Olds. *Brain Sciences*, 11(1), 114.

Wilhelm, F. H. & Grossman, P. (2010). Emotions beyond the Laboratory: Theoretical Fundaments, Study Design, and Analytic Strategies for Advanced Ambulatory Assessment. *Biological Psychology*, 84(3), 552–569. doi: 10.1016/j.biopsycho.2010.01.017.

Williams, Z. J., He, J. L., Cascio, C. J. & Woynaroski, T. G. (2021). A Review of Decreased Sound Tolerance in Autism: Definitions, Phenomenology, and Potential Mechanisms. *Neuroscience & Biobehavioral Reviews*, 121, 1–17. doi: 10.1016/j.neubiorev.2020.11.030.

Wu, M. S., Lewin, A. B., Murphy, T. K. & Storch, E. A. (2014). Misophonia: Incidence, Phenomenology, and Clinical Correlates in an Undergraduate Student Sample: Misophonia. *Journal of Clinical Psychology*, 70(10), 994–1007. doi: 10.1002/jclp.22098.

Wu, S.-D., Wu, C.-W., Lin, S.-G., Wang, C.-C. & Lee, K.-Y. (2013). Time Series Analysis Using Composite Multiscale Entropy. *Entropy*, 15(3), 1069–1084. doi: 10.3390/e15031069.

Xiefeng, C., Wang, Y., Dai, S., Zhao, P. & Liu, Q. (2019). Heart Sound Signals Can Be Used for Emotion Recognition. *Scientific Reports*, 9(1), 6486. doi: 10.1038/s41598-019-42826-2.

Yu, B., Funk, M., Hu, J., Wang, Q. & Feijs, L. (2018). Biofeedback for Everyday Stress Management: A Systematic Review. *Frontiers in ICT*, 5, 23. doi: 10.3389/fict.2018.00023.

Yu, H. & Sano, A. *(2022)*. Semi-Supervised Learning and Data Augmentation in Wearable-based Momentary Stress Detection in the Wild. arXiv.

Yuda, E., Shibata, M., Ogata, Y., Ueda, N., Yambe, T., Yoshizawa, M. & Hayano, J. (2020). Pulse Rate Variability: A New Biomarker, Not a Surrogate for Heart Rate Variability. *Journal of Physiological Anthropology*, 39(1), 21. doi: 10.1186/s40101-020-00233-x.

Zadeh, L. A., Klir, G. J. & Yuan, B. (1996). *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers*. World Scientific.

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J. & Oermann, E. K. (2018). Variable Generalization Performance of a Deep Learning Model to Detect Pneumonia in Chest Radiographs: A Cross-Sectional Study. *PLOS Medicine*, 15(11), e1002683. doi: 10.1371/journal.pmed.1002683.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V. & Chang, K.-W. *(2018)*. Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods. arXiv.