


# Characterizing Active Noise Control Headphones with Psychoacoustics of Amplitude Modulation

*Thesis submitted in partial fulfillment*
*of the requirements for the degree of*
*Doctor of Philosophy in Cognitive Science*

*by*

Sohhom Bandyopadhyay
Roll No. 15510011

*Supervised by*

Nithin V. George
Associate Professor | Electrical Engineering

Indian Institute of Technology Gandhinagar

October 2022

# CERTIFICATE

October 2022

This is to certify that the work embodied in the thesis titled "**Characterizing Active Noise Control Headphones with Psychoacoustics of Amplitude Modulation**" by **Sohhom Bandyopadhyay** (Roll No. **15510011**) has been carried out under my supervision and this work has not been submitted elsewhere for a degree.

**Nithin V. George**

Associate Professor

Indian Institute of Technology Gandhinagar

© Sohhom Bandyopadhyay 2022

All Rights Reserved

# Acknowledgements

I would like to thank my supervisor Prof. Nithin V. George for providing support while carrying out the work described in this thesis. I would also like to thank the members of my doctoral studies committee (Prof. Madhu Vadali and Prof. Meera Sunny) for their suggestions and questions in each bi-annual presentation.

I would like to thank the colleagues and friends from departments of Cognitive Science and Electrical Engineering who have extended a helping hand in various matters throughout the research journey. In particular, I would like to thank Sankha Bhattacharjee (PhD alumnus), Nithin George (PhD alumnus), Kishore Jagini (PhD candidate), and Prof. Krishna Miyapuram (faculty) for the many insightful discussions on science, computational modeling and life at large. I would also like to acknowledge the support extended by Indian Institute of Technology Gandhinagar in the form of various laboratory and other infrastructure for conducting this research.

I would like to thank all the members of the AUDITORY email list (http://www.auditory.org/) who replied to my questions or provided useful materials related to various aspects of the research.

Last but not least, I would like to thank my parents and my sister for being sources of happiness and inspiration while I slogged through this PhD.

Sohhom Bandyopadhyay

October 2022

# Abstract

Active noise control (ANC) headphones work by electronically generating an 'anti-noise' that interferes with external sounds, thereby creating a zone of quiet at the listener's eardrum. Such headphones are supposed to deliver a superior audio experience because the desired i.e. the playback signal is less corrupted by external noise. However, despite being popular in the consumer electronics market and also as a research tool in a variety of disciplines, the effect of the ANC on the playback signal is not usually investigated. This thesis is an initial investigation into the possibility of the playback signal being physically improved (less degraded) due to the ANC, and any consequent psychoacoustic changes.

A signal can be distorted in many ways, and the corresponding space of psychological outcomes can be vast. Hence, this thesis is focused on only one aspect of the playback signal; its amplitude (a.k.a temporal) envelope. An acoustic signal can carry information for a brain in different ways, notably by amplitude and frequency modulations. Amplitude modulation is a mathematical process whereby a carrier signal's amplitude is changed according to an envelope signal. The amplitude envelope's role in human hearing has been studied in a large number of psychological and neuroscientific contexts; e.g. a speech signal can be understood by humans with only the information from its envelope.

This thesis carries out a perceptual evaluation of four commercially available ANC headphones as 'black boxes', i.e. without modifying the internal working of the ANC algorithm. The first two chapters elucidate the modulation transmission characteristics of headphones across the modulation frequency axis with modulated sinusoids, and the last two chapters do so by showing the variation of intelligibility and quality for speech with parametrically degraded envelopes. Overall, this work provides the first direct evidence that the playback signal produced by an ANC headphone has a relatively intact amplitude envelope when the ANC is enabled compared to when it is turned off. By quantifying this phenomenon with various acoustic and psychoacoustic models, and subsequently by studying the psychological effects on humans, we also provide evidence that ANC is helping the brain process the sounds more efficiently. Given the rising popularity of the ANC algorithm in 'hearables', headphones, hearing aids and potentially also cochlear implants, these findings may have implications for the audio industry as a whole, as well as for clinicians working with hearing impaired individuals. In addition, by linking these findings with the remainder of the knowledge of the human auditory system, ANC headphones could also be utilized as a tool for designing new kinds of psychophysical experiments to better understand the cochlear functions and phenomena such as masking.

# Contents

**Abstract** 5

**1 Introduction** 11
1.1 Motivation and background . . . 11
1.2 Active noise control and human hearing . . . 15
1.3 Acoustic amplitude modulation in human hearing . . . 22
1.4 Thesis objectives, scope and organization . . . 24
1.5 Common apparatus . . . 26

**2 Electroacoustic measurement of amplitude modulated signals through ANC** 28
2.1 Introduction . . . 28
2.1.1 Related work . . . 28
2.2 Comparing original and altered envelopes . . . 32
2.3 Materials and Methods . . . 37
2.3.1 Test signals and computations . . . 37
2.3.2 Recording environment . . . 39
2.4 Results and discussions . . . 39
2.5 Chapter summary . . . 46
2.5.1 Future work . . . 46

**3 Simulated neurons' response to amplitude modulated signals through ANC** 48
3.1 Introduction . . . 48
3.1.1 Related work . . . 49
3.1.2 Models of auditory neurons . . . 49
3.2 Materials and methods . . . 52
3.2.1 Simulation procedure and parameters . . . 53
3.2.2 Measures of AM representation . . . 55
3.2.3 Statistical analysis . . . 57
3.3 Results and discussions . . . 57
3.4 Chapter summary . . . 65
3.4.1 Future work . . . 66

**4 Intelligibility and listening-effort of shallow-envelope chimera through ANC** 68
4.1 Introduction . . . 68
4.1.1 Amplitude envelope in human speech perception . . . 69
4.2 Materials and methods . . . 72
4.2.1 Chimera algorithm . . . 72
4.2.2 Stimuli . . . 72
4.2.3 Experiment design . . . 73
4.2.4 Statistical analysis . . . 76
4.3 Results and discussions . . . 78
4.4 Chapter summary . . . 80
4.4.1 Future work . . . 83

**5 Instrumental audio quality of shallow-envelope chimera through ANC** **85**
5.1 Introduction . . . 85
5.1.1 Instrumental audio/speech metrics . . . 87
5.2 Materials and Methods . . . 91
5.2.1 Audio signals . . . 91
5.2.2 Statistical analysis . . . 92
5.3 Results and discussions . . . 92
5.4 Chapter summary . . . 110
5.4.1 Future work . . . 110

**6 Conclusions and Discussions** **112**
6.1 General comments . . . 113
6.2 Limitations and future work . . . 115

# List of Figures


1.1 Signals interacting in an ANC headphone chamber . . . 13
1.2 Thesis structure and chapter interconnections . . . 24
1.3 Setup for recording headphone output signals . . . 27

2.1 Modulation gain against modulation frequency . . . 41
2.2 Envelope Difference Index against modulation frequency . . . 42
2.3 Peak cross-correlation against modulation frequency . . . 43
2.4 Loudness gain against modulation frequency . . . 43
2.5 Sharpness gain against modulation frequency . . . 44
2.6 Kurtosis gain against modulation frequency . . . 44
2.7 Crest factor gain against modulation frequency . . . 45
2.8 Standard deviation gain against modulation frequency . . . 45

3.1 Simulated neural centers and inter-connections . . . 53
3.2 Examples of simulated neurons' spike timing vector in response to recorded AM signals . 54
3.3 rMTF of simulated inner hair cell . . . 57
3.4 rMTF of simulated auditory nerve . . . 58
3.5 MTF of simulated auditory nerve with coefficient of variation . . . 58
3.6 Sync-MTF of simulated auditory nerve with vector strength . . . 59
3.7 rMTF of simulated cochlear nucleus . . . 59
3.8 MTF of simulated cochlear nucleus with coefficient of variation . . . 60
3.9 Sync-MTF of simulated cochlear nucleus with vector strength . . . 60
3.10 rMTF of simulated inferior colliculus . . . 61
3.11 MTF of simulated inferior colliculus with coefficient of variation . . . 61
3.12 Sync-MTF of simulated inferior colliculus with vector strength . . . 62

4.1 Outline of chimera algorithm . . . 73
4.2 Plots of a speech signal and chimera corresponding to various modulation depths . . . 74
4.3 Response time against modulation depth in chimera experiment . . . 80
4.4 Intelligibility (rationalized) against modulation depth in chimera experiment . . . 81
4.5 Distribution of NASA-TLX ratings in chimera experiment . . . 82

5.1 Schematic diagram of ANC headphone evaluation with objective metrics . . . 91
5.2 Chimera audio quality with FES against modulation depth . . . 97
5.3 Chimera audio quality with GPSMq against modulation depth . . . 98
5.4 Chimera audio quality with HASPI against modulation depth . . . 99
5.5 Chimera audio quality with HASQI against modulation depth . . . 100
5.6 Chimera audio quality with PESQ against modulation depth . . . 101
5.7 Chimera audio quality with PEMO against modulation depth . . . 102
5.8 Chimera audio quality with STOI against modulation depth . . . 103
5.9 Chimera audio quality with VISQOL against modulation depth . . . 104
5.10 Chimera audio quality with wSTMI against modulation depth . . . 105
5.11 Chimera audio quality with SRMR against modulation depth . . . 106

# List of Tables


1.1 ANC headphones used in the present work . . . . . 26

2.1 Carrier frequency values . . . . . 38
2.2 Steps for Hilbert transform . . . . . 38
2.3 Mean and standard deviation of envelope metrics across headphones . . . . . 41
2.4 Statistical comparison (two-way ANOVAs) of envelope metrics . . . . . 42

3.1 Mean and standard deviation of simulated neurons' response across headphones . . . . . 64
3.2 Statistical comparison (two-way ANOVAs) of simulated neurons' response metrics . . . 65

4.1 Chimera algorithm . . . . . 72
4.2 Mean and standard deviation of response time and intelligibility across headphones . . . 80
4.3 Statistical comparison (two-way ANOVAs) of intelligibility and response time . . . . . 81
4.4 Statistical comparison (two-way ANOVAs) of NASA-TLX ratings . . . . . 83

5.1 Statistical comparison (two-way ANOVAs) of objective audio/speech metrics for CRM dataset . . . . . 107
5.2 Statistical comparison (two-way ANOVAs) of objective audio/speech metrics for NOIZEUS dataset . . . . . 108
5.3 Mean and standard deviation of objective audio/speech metrics across headphones and ANC conditions for CRM dataset . . . . . 109
5.4 Mean and standard deviation of objective audio/speech metrics across headphones and ANC conditions for NOIZEUS dataset . . . . . 109

# List of Abbreviations

| | |
|---|---|
| ANC | Active Noise Control |
| AM | Amplitude Modulation |
| AN | Auditory Nerve |
| ANOVA | Analysis Of Variance |
| BMF | Best Modulation Frequency |
| CF | Characteristic Frequency |
| CN | Cochlear Nucleus |
| CRM | Coordinate Response Measure |
| Cv | Coefficient of variation |
| EDI | Envelope Difference Index |
| EEG | Electroencephalography |
| ENV | (temporal) Envelope |
| FES | Fidelity of Envelope Shape |
| fMRI | functional Magnetic Resonance Imaging |
| GPSMq | Generalized Power Spectrum Model for Quality |
| HASPI | Hearing Aid Speech Perception Index |
| HASQI | Hearing Aid Speech Quality Index |
| IC | Inferior Colliculus |
| IHC | Inner Hair Cell |
| IQR | Inter-quartile Range |
| ITU | International Telecommunications Union |
| MTF | Modulation Transfer Function |
| NASA-TLX | National Aeronautics and Space Administration Task Load Index |
| PESQ | Perceptual Evaluation of Speech Quality |
| PEMO-Q | Perception Model for Quality |
| RAU | Rationalized Arcsine Units |
| rMTF | rate Modulation Transfer Function |
| SR | Spontaneous Rate |
| SRMR | Speech to Reverberation Modulation energy Ratio |
| STI | Speech Transmission Index |
| STOI | Short Time Objective Intelligibility |
| TFS | Temporal Fine Structure |
| VISQOL | Virtual Speech Quality Objective Listener |
| VS | Vector Strength |
| wSTMI | Weighted Spectro-Temporal Modulation Index |

# 1

# Introduction

## 1.1 Motivation and background

The concept of Active Noise Control (ANC) was proposed by Lueg (1936), and the technology has seen a lot of improvement due to the rapid advances in electronics industry. Unlike conventional passive noise control methods like mufflers and sound absorbers, ANC works on the principle of destructive interference. One of the most successful applications of ANC technology are in headphones. In addition to the circuitry and transducer to produce the acoustic signal from user's device, ANC headphones include a controller in the form of a digital signal processor that injects an out-of-phase acoustic signal ('anti-noise') in order to destructively interfere with external sounds. Usually these controllers are able to cancel frequencies up to 1 kHz (Hansen, 2002). This property makes these headphones an attractive option for listening to speech or music in noisy modern soundscapes, because these soundscapes' noise pollution tends to be dominated by low frequencies typically emitted from engines and other mechanical sources. A growing number of articles, guides and 'howto' sections in popular consumer electronics publications also indicate a growing popularity of ANC headphones among the inhabitants of today's noisy world, and ANC headphones are commonly advertised as the solution to ever present noise as well as a tool for better enjoying music (Bhattacharjya, 2019). The last few decades have seen significant progress in the development and diverse applications of ANC in various fields (Hansen et al., 2012; George & Panda, 2013; Lu et al., 2021a, 2021b).

Apart from many different methods concerned with improving the controller, these research efforts also include some innovations and explorations around the concept of evaluating or assessing ANC headphones in consideration of aspects of human hearing, both physical (Guldenschuh et al., 2012) and psychological (Zacharov et al., 2010). According to some recent reports, the market for ANC and other smart headphones (dubbed 'hearables' [1]) is expected to reach 1.27 billion USD by the year 2025 (Hunn, 2016, 2020), and the acoustic performance evaluation of hearables has been identified to be a subject in its own right (Temme, 2019; Biberger et al., 2021). Currently, the common methods of evaluating or assessing ANC headphones are based around the concept of measuring the reduction of the external acoustic energy, using pure tones, narrowband noise, white noise or pink noise as the external sound field (e.g. (Pradhan, 2021)). However, human hearing is not uniformly sensitive to all sound frequencies, and a simple reduction in the acoustic energy may not correspond to the best overall impression of the ANC by a human listener. The discipline concerned with discovering the connections between properties of acoustic signals and the corresponding sensations in listeners is known as psychoacoustics (Zwicker & Fastl, 2013; Howard & Angus, 2017). In the context of ANC headphones, especially in evaluating them, the presence of psychoacoustics is somewhat sparse.

In the age of Internet audio the concept of evaluating sensory aspects of audio devices has become more relevant. With each new layer of technological development, the manufacturers need to thoroughly understand how their devices interact with human auditory perception so that they can distinguish themselves from competitors in the market (Zacharov, 2019, p. 2). The concept of using human listeners' judgment and opinion in assessing the quality of sound goes back decades, and has been applied to other classes of audio devices (Gabrielsson & Sjögren, 1979). Sound quality is characterized by many dimensions such as pleasantness, bass accuracy, sharpness and so on; the relation between physical aspects of the sound reproducing device and the corresponding evoked sensation or quality ratings given by listeners need to be worked out for each case (Temme et al., 2014). Thus, a significant amount of recent work is engaged with developing standardized methods of sensory or perceptual evaluation of audio technologies, such as concert halls, headphones, professional loudspeakers and so on (Bech & Zacharov, 2007; Lorho, 2010; Harris et al., 2018). To tackle this complexity, standardizing committees and organizations have put forward recommendations that any audio manufacturer can choose to abide by, in order to claim a certain level of performance. For instance, the International Telecommunications Union (ITU) has published a number

[1] 'Hearables' merges 'headphone' and 'wearables'; https://en.wikipedia.org/wiki/Hearables

of recommendations that are relevant for the topic of this work; Recommendation P.835 lays out detailed guidelines on the conducting and reporting of subjective ratings for evaluating audio systems that include noise suppression. This is done by collecting self-reported mean opinion scores from human listeners (e.g. on a 5-point Likert scale that goes from "not distorted" to "very distorted") and comparing it across conditions with statistical analysis (Mangiante & Canevet, 2007). Given the above, the problem of evaluating ANC enabled sound producing systems has remained at the level of an arbitrary mixture of methods from various disciplines but no comprehensive attempt have been made to unify these aspects.

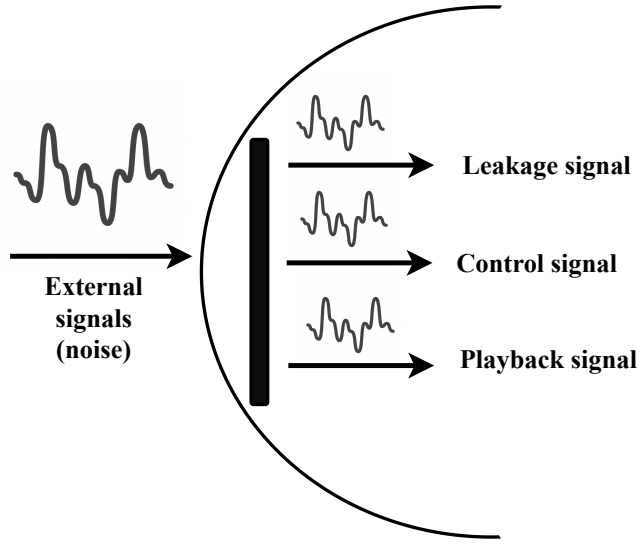



Figure 1.1: Signals interacting in an ANC headphone chamber

Figure 1.1, inspired by (Guldenschuh et al., 2012) and (Nyre et al., 2019), shows the different signals interacting in an ANC headphone chamber while the user is playing some audio. The playback signal refers to the audio signal which the user wishes to listen, the control signal is the 'anti-noise' signal generated by the ANC system and the leakage signal refers to part of the external noise which leaks through the headphone chamber. In theory, the leakage and control signals should exactly cancel each other and leave the playback signal intact. In practice, when evaluating or assessing an ANC headphone, the change (if any) in the playback signal is almost never studied. It has been shown that ANC systems have a complicated and non-linear relationship with the subjective experience such as 'sound quality' or 'acoustic comfort' experienced by human listeners; this relationship also depends on many factors including the spectral content

and intensity of the noise to be reduced (Canevet & Mangiante, 2004; Mangiante, 2004; Zacharov et al., 2010). Interestingly, discussions around the psychoacoustics of ANC frequently gloss over the details of how the playback signal might be affected by the noise control system. For instance, Howard and Angus (2017, p. 440) states that although ANC works quite well, the authors prefer only the passive noise cancellation "…because only the wanted sound is being presented to the ear and the sound itself is not being modified in any way"; they go on to clarify that in an ANC headphone, "…the wanted signal is having noise added to it — not something one really wants to be doing if one can avoid it. If the phase shift is not absolutely correct, complete noise cancelation will not occur, with the result that the wanted signal is then being further contaminated". Pulkki and Karjalainen (2015, p. 278) mentions ANC headphones under 'enhancement of acoustics' where "sound is captured with a microphone and immediately reproduced after possible processing, resulting in a change in the properties of the sound field". Owen (2019) goes into elaborate detail on how headphones with adjustable sound quality control (including ANC) are often considered more desirable than hearing aids by hearing-impaired individuals. These examples show a definite need for a better characterization of the ANC headphones in terms of modifications (improvement or otherwise) of the playback signal, and to do so in terms of the human hearing. Moreover, active noise control has been said to have entered a 'golden age' of applications in the late 2010s (Qiu et al., 2014); consequently, research into its effects and qualitative aspects are also expected to pick up.

This thesis is focused on shedding some light on the changes that may happen to a playback signal when it is played through an ANC headphone, and also to explore how such change(s) may affect the perception for the listener. To tackle the complexity of the phenomenon, we choose to narrow our focus only on one aspect, namely the amplitude envelope of the playback signal. Importantly, this work is not concerned with developing or fine-tuning any ANC algorithm; we use off-the-shelf ANC headphones as 'black boxes', i.e. the goal is to describe these headphones' characteristics in terms of the psychoacoustics of the amplitude envelope across the 'ANC on' and 'ANC off' conditions.

The remainder of this chapter is organized as follows. Section 1.2 reviews the relevant literature regarding the known effects of ANC on human listeners, design and evaluation of ANC algorithms with human listeners or models thereof. Section 1.3 lays out the significance of amplitude modulation in human hearing, as a justification of its use as a characterization for ANC headphones. Section 1.4 provides a brief outline of the chapters of the thesis. Section 1.5 provides details on some common equipment which are used in multiple chapters.

## 1.2 Active noise control and human hearing

As mentioned before, psychoacoustics is the science of understanding the mappings between physical properties of sound signals and the corresponding sensations they evoke in listeners. ANC headphones have occasionally been studied in consideration of the human hearing system. In this section, these studies are briefly reviewed, grouped into three related but distinct categories.

**A. ANC design with psychoacoustics**

ANC works by sampling the environmental sound with a 'reference microphone', estimating the transfer function of the passive absorption through the headphone enclosure, and using a digital signal processor to produce an appropriate anti-sound to interfere with the 'leakage signal' (figure 1.1) just-in-time as it arrives. This concept is called 'feedforward ANC'. Then there is also 'feedback ANC' where the resulting acoustic signal (after the interference) is recorded with an 'error microphone' and is used to update the digital filter used for producing the anti-sound (Snyder, 2000; Hansen, 2002). ANC researchers have taken note of the fact that human hearing has complex properties, hence 'performance' of an ANC measured with purely spectrum based methods may not correspond to the subjective or perceptual judgment by human listeners (de Oliveira, 2010); i.e., "…it is often possible to implement an active control system where the performance function is attenuated substantially, but in which human observers have noted little difference between the uncontrolled and controlled states" (Sommerfeldt & Samuels, 2001). Hansen et al. (2012, p. 2) cautioned that the harmonic distortion due to the ANC actuator would result in additional, higher frequencies being produced (those which were not present in the external noise). Since these frequencies are multiples of the lower frequencies to which the ear is more sensitive, the net result can be that the result of ANC may sound *worse* than the condition with no ANC at all. Keeping these concerns in mind, many researchers have proposed 'psychoacoustic ANC' algorithms, which are usually in the form of: a) using an auditory-inspired weighting function on the residual noise from error microphone, or b) auditory-inspired weighting function on the input noise from reference microphone, or c) use a psychoacoustic metric as a cost function of the adaptive algorithm (reviewed by (Guicking, 2007; Jiang & Li, 2018)). Such proposed algorithms/systems then typically use a psychoacoustic 'dimension' such as loudness, roughness, sharpness or tonality in the evaluation of the proposed ANC system; this evaluation is either done with rating-scales or mean-opinion-scores from human listeners or with objective models of these perceptual quantities. de Diego et al. (2001) used a jury test and objective models of roughness, loudness, tonality and sharpness to

assess the goodness of an ANC against car engine noise, and concluded that reduction of overall acoustic energy does not necessarily correspond to reduced annoyance of car engine noise, and specific spectral weighting must be applied to account for this fact. Similarly, a number of other researchers also implemented various filters to mimic the perceived loudness contour of human hearing with respect to audio frequencies, to shape the input noise signal or the after-cancellation error signal of an ANC system (Sommerfeldt & Samuels, 2001; Bao & Panahi, 2010a, 2010b; Belyi & Gan, 2017; Benois et al., 2018; Belyi & Gan, 2019; Wang et al., 2019). This human frequency sensitivity curve has also been standardized for these kinds of purposes by ITU (ITU-R, 1986). de Oliveira et al. (2010) proposed a fast converging algorithm where the focus is expanded from 'noise control' to 'sound quality control', where similar ideas were used to shape the harmonics of the noise. Hasegawa et al. (2018) combined these ideas, and used an auditory-based frequency weighting filter for both the input noise and the residual noise (using both A-weighting and ITU-R 468 weighting). For hearing-impaired listeners, Durant et al. (2013) proposed modifying the cost function of the adaptive algorithm itself in order to make it take into account the non-linear loudness sensitivities experienced by such listeners. Brammer et al. (2005) suggested that the ANC controller architecture has a close relation to the speech intelligibility improvement of the human users of the headphone, and later developed an ANC headphone based on frequency 'sub-band' processing combined with a method for optimizing the playback signal's SNR, and found that word recognition scores improved in human listeners (Brammer et al., 2014). Schlieper et al. (2021) proposed a genetic-algorithm based 'mixed sensitivity' ANC system, which means it takes into account real world constraints such as headphone placement variability and other external factors; they evaluated this system in terms of loudness and sharpness and concluded that optimizing for these perceptual dimensions may result in improved controller. In terms of modulation characteristics, there has been almost no ANC proposed to account for modulation frequency sensitivity of human listeners (reviewed in section 1.3). de Oliveira et al. (2009) did touch upon the concept of modulation index while studying ANC sound quality of a car engine noise using roughness and other perceptual metrics, but did not compute the change in the modulation characteristics.

**B. ANC evaluation with psychoacoustics**

In an ANC headphone, a number of factors come into play that affect the overall perceptual outcome. These factors include the nature and design of the controller algorithm, the quality of the acoustic seal of the device around the listener's ears, the ambient noise environment where it is being used, specific aspects of the waveform of playback signal, idiosyncrasies of the listeners themselves, and more. Recognizing

the complexity of the acoustical situation in an ANC headphone, a few papers have attempted only to evaluate ANC headphones or other devices using psychoacoustic principles. Brammer et al. (2005) used a modulation transmission concept called 'speech transmission index' and used it to characterize the differences caused by feedforward and feedback architectures in ANC headphones. Darlington and Guiu (2010) proposed ANC headphones be assessed using an adaptive algorithm coupled with pre-determined sound intensity levels to find the preferred listening level by human listeners. Zacharov et al. (2010) set out to establish the dimensions of perception that may feel different while using an ANC headset using 'multidimensional scaling', which basically amounts to asking listeners to give ratings along pre-established scales like 'smallness' of stereo space, 'clarity' of noise, perceived 'linearity' of attenuation and so on; their methods are actually an instance of the concept of perceptual evaluation which has a much richer history in other areas of acoustics such as loudspeaker evaluation (Bech & Zacharov, 2007; Lorho, 2010; Zacharov, 2019). Watanabe and Hamada (2011) provided some results using ANC earphones regarding the binaural effects of ANC in objective and subjective measurements; they found that the ANC correlates between the spatial information of the noise source and its subjective performance. Guldenschuh et al. (2012) investigated the effect of leakage while conducting objective ANC evaluations, and also measured the correlation between objective noise attenuation and subjective ratings of loudness and noise reduction. Behn (2013) evaluated ANC headphones with a semantic differential, which similarly uses ratings on contrasting perceptual dimensions like 'soothing-exciting' and 'strong-weak' etc. Priese et al. (2013) studied preferences in terms of pairwise comparisons between different ANC algorithms with ratings of loudness and pleasantness from listeners, and concluded that in order to substantiate ANC manufacturers' claims regarding hearing comfort etc. these devices should be evaluated with properly designed psychoacoustic studies, and that "…deeper research connecting psychoacoustics and control engineering is needed". Li et al. (2018) studied the effect of ANC on perceptual dimensions of loudness, harmonic noisiness, fluctuation strength and other metrics using a listener experiment and proposed a neural network model for combining these separate dimensions into an overall model for predicting sound quality. Their results showed that the ANC was able to substantially improve most of the annoyance-causing factors in a specific kind of noise environment. Reimes (2019) utilized the modulation filterbank model of auditory perception (discussed in section 1.3) of auditory perception, and attempted to build an automated framework to predict 'listening effort' while using ANC headphones. Basu et al. (2021) devised an objective method based on audio spectrum to quantify the relative lack of degradation of music signals when played through ANC.

### C. Practical effects of ANC on human listeners

Passive noise cancellation methods are quite effective for high frequency noise cancellation, but ineffective or impractically bulky and/or costly for low frequency noise cancellation. ANC aims to fill this gap and have been reported to be an efficient solution for low frequency noise cancellation. Hence, ANC has emerged as a popular tool for mitigation of the noise problem in a variety of real world settings. Researchers working in fields as diverse as pediatric medical care to gait analysis have attempted to use ANC for the benefit of human participants. In this section, some of these studies are reviewed.

Due to certain peculiarities of the human cochlea, low frequencies are more effective at masking higher frequencies than vice versa; this phenomenon is dubbed 'upward spread of masking' (Moore, 2012, chap. 3). Given the ubiquity of upward spread of masking, and the fact that ANC is most effective against low frequencies, ANC systems included as part of hearing protectors or communication headsets used by pilots have attracted considerable interest since the early 1990s (reviewed by (Brammer et al., 2008)). Most of these studies are concerned with measuring changes in speech intelligibility under realistic noises such as various aircrafts, tanks, artillery, traffic, and so on. It was frequently found in these studies that although there was a large drop in acoustic sound pressure level and a corresponding increase in users' ratings on various opinion scales (such as preference in pairwise comparisons), the objective measurement of intelligibility showed little to no improvement (Casali & Gower, 1993; Gower & Casali, 1994; Robinson & Casali, 2000; Powell et al., 2003). In a 1997 conference titled 'Audio Effectiveness in Aviation', a number of reports laid out properties of ANC (then referred to as 'ANR' Active Noise Reduction) in a human context. Wagstaff and Woxen (1997) measured the level of noise reaching the eardrum after the ANC did its job; they hinted at an interesting possibility of ear canal resonance interacting with the ANC-generated anti-noise and cause unknown distortions in the audio signal reaching the inner ear. Simpson and King (1997) conducted an in-depth investigation with acoustic and communication effectiveness of ANCs and found that ANC-enabled hearing devices are generally helpful for military pilots because they "…substantially increases speech intelligibility", reduce demands on attention, are effective against the typical acoustical environment (weapons), and are rated highly by pilots. Anderson and Garinther (1997) used consonant-vowel-consonant word identification test and found that ANC increases speech intelligibility from 68% to 89%, and also that this led to greater proportion of missions completed successfully by the soldiers. Crabtree (1997) went a step beyond the usual play-and-test approach and devised their own ANC testing methodology, using specific external signals to stress test the ANC systems such as high intensity and low-frequency

pure tones.

A number of researchers have also attempted to find precise evaluation methodology for ANC headphones or hearing protection devices, stressing the importance of careful experimental design incorporating the variety of factors that may change the outcome (McKinley et al., 1997; Steeneken & Verhave, 1997; Casali et al., 2004; Bolia et al., 2005; Valimont, 2006; Bockstael et al., 2011). Abel et al. (2007) studied localization abilities in addition to speech understanding in specific noises and found improvement with ANC compared to passive hearing protectors. Bolia (2003) found that binaural hearing has an interaction with ANC in airforce noise environments, specifically that speech intelligibility increases when the speech signal is presented in the other ear to the noise and this effect is enhanced with an ANC headphone. Valimont et al. (2006) investigated the effect of ANC headsets among trained pilots and found significant improvement in speech intelligibility and mental workload, and also some flight performance measures like magnetic heading, altitude, indicated airspeed etc., and suggested the possibility of a "synergistic interaction" between the acoustic and cognitive outcomes. A number of studies have investigated the possible changes caused by the use of ANC headphones in functional magnetic resonance imaging (fMRI). An fMRI machine emits a great deal of acoustic noise so ANC may lead to a better brain activity signal of the participants. Such evaluations have been reported to be beneficial both in terms of detectable differences among brain regions engaged in auditory tasks, and also behavioral measures such as threshold of tone-in-noise detection (McJury et al., 1997; Hall et al., 2009; Peelle, 2014).

The reviews on music listening with ANC headphones are varied. Some users are convinced that ANC makes the music sound better while others are certain of the opposite; researchers have found evidence of both sides. Liang et al. (2012) studied music playback habits and found that ANCs are beneficial for users listening to music through personal music players. Bagha et al. (2018) studied music perception under ANC with machine learning. They found that computational prediction of the music genre from brain signals (using electroencephalography) was easier when the ANC was enabled, which was taken as evidence that ANC improves the music perception. Butterbaugh and Gordon-Pershey (2019) found that ANC may improve the music experience in some cases, but the participants nevertheless used higher playback volumes when listening in a noisy environment; moreover their results indicated that ANC might actually cause users to listen to music at "unsafe listening levels".

Audiometry refers to the process of clinically assessing an individual's hearing abilities (typically detection thresholds for pure tones at pre-determined frequencies), and this process requires carefully calibrated

equipment and sound-isolated room for conducting the test. ANC headphones have attracted interest of researchers working on developing novel methods of audiometry that can be deployed in non-laboratory settings which would have poor acoustic isolation. Abel and Spencer (1995) measured listeners audiometry thresholds and speech-in-noise performance, and failed to find any improvement or degradation due to ANC; in a follow-up study, they investigated this issue with hearing-impaired as well as normal-hearing subjects and found that for normal hearing group the word recognition performance increased, but for impaired group the performance actually dropped due to ANC (Abel & Spencer, 1997). However, in later years this trend reversed and ANC began to become a viable tool for audiometry. The general conclusion from these later studies is that ANC provides a convenient alternative to the lab-based hearing tests, and produces reliable results (Bromwich et al., 2008; Chang et al., 2019; Sun et al., 2019; Lelo de Larrea-Mancera et al., 2020). Clark et al. (2017) extended this by including bone-conduction in their analysis; they concluded that "audiometric bone-conduction tests can be performed on most clinical patients via noise cancellation earphones in lieu of an audiometric sound enclosure". Nielsen et al. (2011) found that ANC headphones were more effective in assessing distortion-product otoacoustic emissions, which is a standard protocol for checking normal cochlear function in newborns and individuals who cannot respond to verbal instructions.

Molesworth and colleagues got inspired by the classification and consequent prohibition of ANC headphones as personal electronic device in commercial flights due to a complaint to 'Asia Pacific Flight Cabin Safety Working Group' (Burgess & Molesworth, 2013). They then set out to investigate this issue from a safety and ergonomics perspective, and investigated the potential *beneficial* effect(s) of ANC headphones on human listeners in a civilian aircraft, particularly relating to understanding and remembering safety related instructions (for passengers) and for carrying out aviation related tasks (for crew and pilots). They conducted a number of experiments focused on civil aviation environments, such as finding the patterns of mis-communication errors in spoken instructions (Burgess & Molesworth, 2012), improvement in a memory recall task involving complex information delivered via audio alongside a concurrent mathematical task (Molesworth, Burgess, & Kwon, 2013), effects on memory (recall and recognition) of complex information (Molesworth & Burgess, 2013; Molesworth, Burgess, & Chung, 2013); they found that ANC does assist in these types of memory situations but only when the amount of noise is moderate. Then they extended the study designs to multi-sensory effect (pictures of safety related instructions in addition to audio delivery) (Molesworth, Burgess, Gunnell, & Duong, 2014); they found that while ANC did improve the perfor-

mance in pictorial task, more experienced speakers of the language did better than non-native speakers regardless of ANC or conditions. Finally, they investigated the relative improvements caused by ANC across native and non-native speakers of English, and also among trained pilots vs. non-pilots; here they found that the effect of ANC was more in non-native speakers compared to native speakers, and non-pilots had greater benefit from the ANC than trained pilots with recognizing the aviation phrases (Jang et al., 2014; Molesworth, Burgess, Gunnell, Löffler, et al., 2014).

ANC has been deployed in a number of medical contexts to assess potential improvements in patients or medical staff or other caregivers. Akhtar et al. (2000) studied the application of ANC to reduce the noise burden in patients and caregivers in medical pediatric intensive care unit, and found a significant reduction in subjectively experienced noise level for nurses, parents, and other personnel. Gallacher et al. (2017) used acoustic recordings to study the exposure level of intensive care units in a hospital and whether ANC could reduce exposure to stressful noises; they found that the ANC does reduce the overall noise exposure level but did not speculate about actual benefits to human participants. Tran et al. (2020) investigated whether patients undergoing an operation ('primary total knee arthroplasty') requested more sedatives when they were allowed to listen to music via ANC headphones as opposed to simply wear the ANC headphones; they found no significant difference between the two groups. Karalar et al. (2016) found that ANC headphones reduced pain and anxiety in patients undergoing shockwave lithotripsy, a process of kidney stone removal. Tsivian et al. (2012) however found that although music improves pain and anxiety, ANC alone did not make a significant difference in patients undergoing transrectal prostate biopsy. Sociologists, philosophers, urban scientists and so on have also taken note of the ANC phenomenon; Hagood (2011) complained about some of the cultural, emotional and socio-economic implications of the use of ANC headphones among flyers of airplanes. Nyre et al. (2019) put forth a formulation of ANC headphones called 'the theory of perceptual mechanics', which is essentially a set-theoretic description of the unwanted noise and desired sound entities interfering inside the headphone chamber (figure 1.1), and its potential impact on the overall multi-sensory perception of the user in a noisy city. Sosna (2019) explored the overall experiences of sonic interaction between a listeners body, environment and ANC from a philosophical perspective. Hamacher et al. (2018) found result of "…improvement of gait stability during normal overground walking when using active noise cancelling compared to normal hearing", which they hypothesized could be due to improved perception of the footstep sounds through the bone-conduction pathway.

In the context of office ambient noise and worker satisfaction, it has been found that generally ANC

does not improve cognitive performance, noise annoyance ratings, or physiological markers of stress like heart-rate variability in workers when used in isolation but may have modest benefits if used in conjunction with other methods such as 'sound masking' (Kari et al., 2017; Müller et al., 2019; Radun et al., 2021).

## 1.3 Acoustic amplitude modulation in human hearing

A time-domain acoustic signal can be decomposed into a fast-varying and a slow varying pair, known as temporal fine structure (TFS) and temporal envelope (ENV) respectively. The TFS is also known as carrier and the ENV is also called modulator in different sub-disciplines. For a natural signal, this decomposition is a mathematically ill-posed problem hence there are a large number of solutions to it (Turner, 2010); usually the absolute values of the Hilbert transform (Shiavi, 2010, section 10.1.4) is used as ENV and the cosine of the instantaneous phase of the Hilbert transform is used as the TFS, and this definition is adopted in the present work. About 40 to 50 years ago, researchers across multiple disciplines started focusing on these aspects of audio signals as a standard tool for understanding how biological auditory systems encode and retrieve information, and how these aspects can be utilized for designing audio acquisition, storage and transmission systems.

In neuroscience, researchers working with animals like cats or rodents have studied the representation of amplitude envelope in various locations of the brain. They usually present stimuli in the form of various modulation sinusoids to awake or anesthetized animals, and make recordings from brain centers using electrodes. The neurons' responses are then correlated using mathematical models to find the links between physical stimulus parameters and the 'neural code' for the same (Rudnicki et al., 2015). A large amount of such data have been collected from various locations and neuron types, and for various stimuli; see (Joris et al., 2004) for a review.

In auditory psychology, researchers working with human listeners design psychophysical experiments to find the mappings between stimuli parameters and limits of perceptual. For instance, the concept of 'detection threshold' is used to find the limits of the ability to detect an amplitude modulated signal among un-modulated signals by parametrically reducing the modulation depth; these thresholds when plotted against the modulation frequency form the 'modulation transfer function' (MTF), which informs us about the brain's ability to encode and process the different modulation frequencies present in natural signals (Viemeister, 1979). The goal of building an MTF of a system, much like the usual transfer function in

audio-frequency domain, is to describe the system's performance for an arbitrary modulation frequency and thus account for a broad class of modulation perception phenomena (Viemeister, 1979; Schroeder, 1981; Plomp et al., 1984; Harris, 2015). Modulation perception was also integrated with other aspects of auditory psychology such as forward and simultaneous masking (Houtgast, 1989), and some new concepts such as 'comodulation masking release' were found and began to be studied (Berg, 1996; Verhey et al., 2003). Efforts to integrate these findings and ideas have continued since the 1990s; one such model of modulation perception has become relatively popular and is known as the 'modulation filterbank' model. This model is based on the concept of an envelope extraction stage followed by a separate (succeeding) filterbank after the cochlear filterbank; this model is able to account for a broad set of human perceptual data including modulation masking and modulation detection with different carriers (Dau et al., 1997a, 1997b). Due to the cascaded architecture, this model has been extended in pieces e.g. by improving the cochlear filterbank model (gammatone to 'dual resonance non-linear') to account for more experimental data (Jepsen et al., 2008). Due to the fundamental nature of amplitude modulation in sound, the concept of 'effects of xyz on AM perception' has become a research area in its own right. For instance, Vinay and Moore (2010) investigated whether listening to music at moderately high volumes caused any change in the amplitude and frequency modulation thresholds for young, normal-hearing participants, and found subtle impairments of FM discrimination and improvement of AM detection. Noise reduction (which is different from ANC) was found to cause a "…small but consistent" improvement in AM depth detection thresholds in normal-hearing and hearing-impaired listeners (Ives et al., 2013; Ives et al., 2014).

In audio signal processing and acoustical engineering, the concept of amplitude envelope has slowly found a number of applications. Most of these come from the Dau model discussed above, which has been utilized to predict 'instrumental' speech intelligibility and quality metrics (chapter 5), or prediction of listening effort of ANC headphones (Reimes, 2019). A codec is a piece of software that allows storing or transmitting audio data at much smaller sizes by creating a compressed version of the signal, and then uncompressing it before playback. Virtually all codecs make use of psychoacoustic principles (such as masking), and some use temporal envelope properties (such as dissimilarity across envelopes across frequency bands) to optimize the compressed representation (Ghido et al., 2017). On the other hand, envelope based metrics for assessing the performance of codecs have also been proposed (van de Par et al., 2019). Design of spatial audio technologies such as ambisonics also rely on concepts related to amplitude modulation, and the relevance of the envelope in spatial sound localization by human listeners have long since recognized

(Bregman, 1994, chap. 3).

These and other developments in the 'modulation theory of hearing' have been reviewed by (Elhilali, 2019). Other studies pertaining to relevance or role of ENV in human auditory perception have been reviewed in the relevant sections of chapters 2 through 5.

## 1.4 Thesis objectives, scope and organization

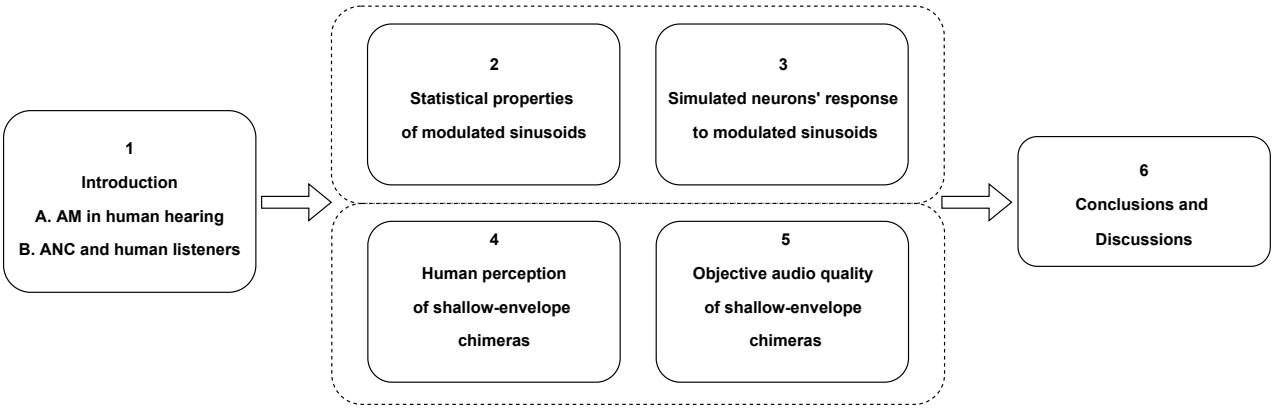



Figure 1.2: Thesis structure and chapter interconnections

The present work is an attempt to bridge the gap between psychoacoustic evaluation and ANC headphones. Given the diversity of the topics involved, and the fact that these disciplines share very little vocabulary or concepts with each other despite being all connected with human hearing, the present work does not attempt to be comprehensive in any particular direction. Rather, this attempt is to tie some threads together as a first step towards unifying our understanding of the situation when the ANC function is enabled in a headphone while simultaneously playing some sound through it. In this thesis, we study the characteristics of four commercially available ANC headphones in terms of their AM transmission properties. We are agnostic about the parameters or other internal details of the ANC algorithm used inside the headphones, and use them as a 'black-box' that can play a signal in both the 'ANC on' and 'ANC off' conditions. The thesis has four contribution chapters where two separate aspects of AM are studied. Chapters 2 and 3 are concerned with modulation frequency, and chapters 4 and 5 are concerned with modulation depth; this structure is illustrated in figure 1.2. The chapters are summarized below along with their contributions.

- Chapter 1 (this one) has laid out the common background of all chapters, which includes sections on a. sensory evaluation of audio devices, b. basic concepts of the psychoacoustics of amplitude

modulation and its relevance for audio signals, and c. review of studies that have examined active noise control in a human context. Thus the contribution in this chapter is a systematic review of the literature spanning these diverse areas.

- Chapter 2 is focused on analyzing the amplitude envelope of the playback signal in terms of distortions across the modulation frequency axis. By applying acoustics and auditory inspired time-domain metrics (statistical functions like kurtosis) to the recorded signals obtained from the headphones, it shows various patterns of differences caused by ANC. In most cases, it is seen that the ANC being enabled produces a less distorted amplitude envelope. The contribution of this chapter is in the assessment method for ANC using the envelope of modulated sinusoids. This chapter serves as a baseline against which other more advanced findings can be compared, both in the present work and future studies.

- Chapter 3 extends the previous chapter by studying the representation of amplitude envelope in the auditory brainstem and the changes produced (if any) due to ANC. The recorded signals discussed in chapter 2 are provided as input to computational models of neurons of the auditory pathway of the brain, and the responses of these simulated neurons are analyzed to find differences in the AM representation across ANC conditions. The contribution of this chapter is the conceptual connection between the literature of ANC, perceptual evaluation and computational neuroscience; these areas all deal with audio in some form or the other but ANC headphones have not been evaluated using simulated neurons of auditory pathway before.

- Chapter 4 describes a human listener based psychometric experiment with speech-like signals that have parametrically degraded amplitude envelope. In this experiment, the participants' speech perception is studied in terms of intelligibility and listening effort, and these are compared across ANC conditions. Thus, the notion of ANC delivering a less-distorted signal is put to a real world test in this chapter, in the context of a particular kind of envelope distortion. The contribution of this chapter is in the experiment design (along with chimera algorithm), and the data collected with it.

- Chapter 5 extends chapter 4 by evaluating the speech-like signals with various objective audio or speech quality models. These models aim to predict intelligibility or speech quality data that would be produced by human listeners, i.e. they act as virtual listeners. By comparing the scores from such

virtual listeners across ANC conditions, this chapter provides a detailed (and sometimes surprising) description of the effect of ANC on the speech-like signals with degraded envelope. Thus, the contribution of this chapter is the systematic evaluation of ANC headphones on a new method based on amplitude envelope degradation with well-known and widely used audio-quality and intelligibility algorithms, along with some lesser known ones.

- Chapter 6 concludes the thesis, by providing some additional perspectives from the previous chapters, and highlighting some interesting future research directions that may arise from the present work.

## 1.5 Common apparatus

In all chapters, the data from measurements/experiments are presented as a comparison between four headphones, which are mentioned in table 1.1.

| Manufacturer | Model no. | First available |
|---|---|---|
| Sony | WH-1000XM4 | 2020 |
| Sony | WH-1000XM3 | 2018 |
| Sennheiser | HD 4.50BTNC | 2017 |
| Bose | QuietComfort 25 | 2014 |

Table 1.1: ANC headphones used in the present work

All four headphones are 'circumaural' type, i.e. they completely cover the ear pinna when worn by a human listener or placed around a dummy head. The headphone models were chosen based on convenience and availability, and are only meant to serve as examples of commercially available devices. They can be replaced with any headphone that has the capability of playing a signal in both 'ANC on' and 'ANC off' conditions. Furthermore, this thesis is not concerned with developing or altering any ANC algorithm, as such, it treats each headphone as a black box without regard to its inner workings.

Chapters 2 and 4 describe various signals that were recorded through headphones and used for model based analysis. These signals were generated with custom-made scripts in an iMac computer (Apple, California, United States) running MATLAB software (Mathworks, Massachusetts, United States). The signals were digital-to-analog converted with a full-duplex USB audio interface (U-PHORIA UMC204HD, Behringer, Willich, Germany) and then played through the respective ANC headphones. The headphone output was recorded with a KU 100 dummy head microphone (The Neumann Company, Berlin, Ger-

many). Finally, the recorded signal was fed back into the same computer through the same USB audio interface for storage. This setup is illustrated in figure 1.3.

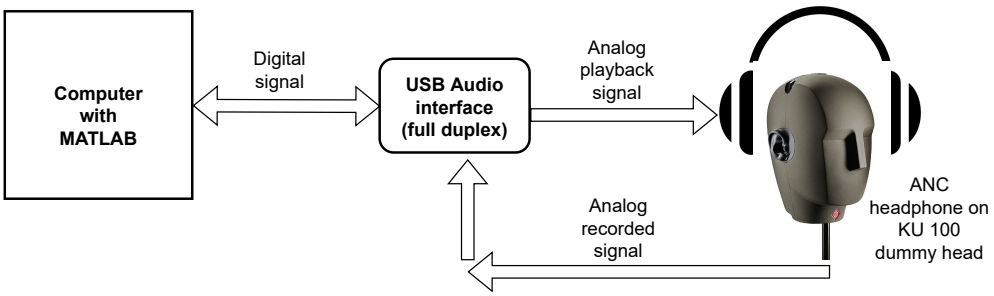



Figure 1.3: Setup for recording headphone output signals

# 2

# Electroacoustic measurement of amplitude modulated signals through ANC

## 2.1 Introduction

This chapter is concerned with characterizing ANC headphones in terms of various modulation transfer functions of its playback signal. As described in chapter 1, ANC headphones have three acoustic signals interacting in the headphone chamber when the ANC function is enabled. The goal of this chapter is to provide an initial investigation into the role of ANC in preserving, altering or enhancing the AM envelope present in the playback signals. By quantifying any such changes in the amplitude envelope of the playback signal, and visualizing the results along the modulation frequency axis, we obtain various MTFs that are informative about the modulation transmission characteristics of the device.

### 2.1.1 Related work

In loudspeaker quality research, it is common to measure the distortion of the playback signals by quantifying the 'free-field' response and plotting the amplification against the audio frequency axis (Toole, 1986b). A need for a distortion metric that is reliable in a psychoacoustic sense, such that the changes in the distortion metric faithfully correlates with the changes in evoked sensations for all inputs, has been identified for

audio devices for some time (Geddes & Lee, 2003). Measuring the reproduction accuracy of headphones is also a long-standing problem (Martin & Anderson, 1945). Headphones have the usual problems as that of a loudspeaker, and in addition a few special considerations due to the small acoustic enclosure and low-powered nature of the output. For evaluating headphones, a dummy-head microphone is usually used to record the acoustic signals generated by the headphone under test, and these signals are subjected to some kind of computational or statistical analysis. Traditionally, researchers have been interested in the question of realistic reproduction of the sound field by a headphone which would be similar to that produced by loudspeakers. Some researchers have particularly investigated the problem of coupling with human ears, i.e. the acoustic leakage and fitting of the headphone enclosure on or around the ears (Toole, 1984). Measuring headphone responses accurately with artificial ear couplers are also a challenge since they are not good replicas of human heads; in such settings, group delay, acoustic crosstalk, harmonic distortion, impulse response decay, external sound radiation are used to quantify the sound properties (Hirahara, 2004). Concepts such as acoustic load and acoustic impedance also have a role to play in quantifying the distortions produced by a headphone (Vorländer, 2000). The International Telecommunications Union has studied the problem of measuring the fidelity of acoustic output (from variety of devices like landline telephone sets, mobile telephones) and come up with a number of recommendations that lie close to the theme of this chapter. In particular, the problem of measuring acoustic fidelity is overlapping across a wide variety of devices, protocols, softwares (e.g. codecs) and so on. Recommendation P.381 (ITU-T, 2016) describes signals and test methodology to measure audio fidelity of headphones. The measured aspects include noise, frequency response, distortion and crosstalk. It also lays out the connectivity specifications such as audio socket dimensions, electrical impedance etc. Recommendation P.380 (ITU-T, 2003) describes the positioning and other physical parameters with respect to a dummy head (or head and torso simulator) when carrying out electro-acoustic tests on a headphone. Recommendation P.501 (ITU-T, 2020) describes a set of signals for sending and receiving across a speech carrying channel. Recommendation BS.1116-3 (ITU-T, 2015) lays out a quality evaluation protocol for small impairments caused by audio devices, taking into considerations the specific types of impairments caused by loudspeakers or headphones, which are all based on frequency domain concepts. In addition, when ANC is deployed as a hearing protection device, there are two standardized protocols named 'REAT' (Real Ear Attenuation at Threshold) and 'MIRE' (Microphone In Real Ear) that aims to measure the noise reduction performance in terms of the reduction of acoustic energy measured with microphones placed in strategic locations (Berger, 2002; Perala, 2006). Al-

though these are considered the 'go-to' toolkit for evaluating hearing protector devices, these standards are not concerned with the quality or fidelity of the playback signal produced by the ANC headphones.

For ANC headphones, many researchers have described the need for dedicated distortion or quality measures to be developed. Ogonowski (2003) emphasized that ANC headphones have to be treated as a class of their own due to inherent differences and conditions beyond the control of the experimenter, and suggested that for purposes of benchmarking, "…only the control structures, excitation signals and indices of attenuation quality can be standardized". They also proposed a 'mean square deviation' as one of the indices to track the signal, which is adapted in this work in the form of standard deviation gain. Wagstaff and Woxen (1997) measured the resultant noise reaching the eardrum and considered various aspects of the ANC performance, such as the frequency dependent amplification and various 'resonance phenomena' occurring due to the shape of the outer ear canal and concluded that the inter-subject variability in this regard should be studied more thoroughly. Ang et al. (2017) studied the performance of commercially available ANC headphones in terms of attenuation of the spectra of realistic noise environments such as construction site, aircraft takeoff, bus cabin and concluded that ANC headphones need to be evaluated using more than pink noise or white noise, and especially focus on transients, i.e sounds that have short period of decay or onset. In an attempt to produce a systematic assessment and comparison protocol for ANC headphones, Rudzyn and Fisher (2012) studied a number of commercially available ANC headphones in terms of attenuation of impulse and broadband noise, and also overload response and movement related stability. They found that the phenomena are quite complex, with large variations in the performances caused by such environmental factors, and that virtually no ANC device provided impulse noise protection. Liebich, Richter, et al. (2018) derived a detailed empirical relationship between the nature of the ANC system and its dependency on the direction of arrival of external noise, and proved certain bounds on the resultant anti-noise that it can produce. On a similar note, Liu et al. (2022) studied ANC headphones in a simulated realistic aircraft environment and provided detailed analysis showing that the cancellation performance depended heavily on the spatial distribution of the sound. Even though these and other similar works may be directly relevant to the perceptual outcomes, these (and most other) ANC publications do not consider the possible effects of ANC on the playback signal, i.e. the desired outcome from the users' point-of-view. Brammer et al. (2005) examined the phenomenon of alteration of the playback signal (referred to as 'communication signal') with two different ANCs, one with feedback and the other with feedforward control structure. Among other measures, they also utilized the speech transmission index which is based on the

modulation transfer concept and found that the feedforward control structure was superior in terms of not corrupting the playback signal. Denk et al. (2017) studied the problem of acoustic transparency and corresponding perceived sound quality in smart earphones (also applicable to hearables) using insertion loss and other physical metrics. They went to a great detail to understand the physical sound radiation patterns at various locations in the ear pinna using frequency domain approaches (transfer functions). Struck (2019) measured insertion loss and proposed methods for converting the subjective REAT or MIRE protocols into objective formats. In addition, they also touched upon the concept of measuring amplitude envelope fidelity and discovered that although one of their test headphones were better in terms of insertion loss, the envelope analysis revealed a "different story". Given that closed-back headphones are known to introduce spectral coloration, and that open-back ones are preferred for this reason, Schlieper et al. (2019) measured the 'openness' or acoustic impedance of ANC headphones and found that ANC being enabled reduced the acoustic impedance. The concept of acoustic impedance is similar to the 'acoustic load' studied in the context of headphones of other types and in other contexts (Vorländer, 2000). Liebich, Fabry, et al. (2018) briefly reviewed the signal processing as well as psychoacoustics aspects of ANC headphones and noted that system noise such as digital-to-analog or analog-to-digital converters may all corrupt the playback signal, but it is a "…largely disregarded subject".

Taken together, the literature suggests that there is a need to better understand the playback signal of an ANC headphone in terms of the amplitude envelope. An exact process for doing so may require attention to specifics of the application environment, and as such, may require flexible method that allows quantifying the envelope in different ways. A recently developed body of work has proposed a quality metric of professional monitoring loudspeakers based on their ability to accurately reproduce the amplitude modulations sent to them in the playback signal (Harris, 2015; Harris et al., 2018). In much the same vein, ANC headphones could be evaluated in terms of the properties of the amplitude envelope of the reproduced signal. Due to the adaptive signal-processing algorithm and its interaction with ever changing environmental sound, the ANC headphone is a more complex device than a simple headphone or a telephone system. None of the existing audio-device evaluation methods cover the measurement of acoustic outcomes given the destructive interference happening inside an ANC chamber, as outlined in figure 1.1. Thus, the core motivation behind the metrics used in this chapter is to utilize some perceptually or acoustically motivated functions to quantify the alteration of envelope under ANC. Some such ideas are presented next.

## 2.2 Comparing original and altered envelopes

The concept of 'change in the amplitude envelope' is sometimes used as part of some larger investigation, e.g. in the context of hearing aid compression algorithms. In particular, dynamic range compression usually corrupts the amplitude envelope by changing the modulation depth and also possibly clipping the upper portions of the signal, so its effects are frequently studied in terms of the change of shape of the envelope (Jenstad & Souza, 2005; Sabin et al., 2013). It has been recognized that in acoustics, especially when dealing with biological or neuroscientific considerations, considering time-domain behavior of the signal is essential; for instance, marine animals' noise exposure from drilling or other types of underwater activity needs to be measured considering the time-domain characteristics like kurtosis of the acoustic noise instead of relying on 'overall' concepts like equal-energy exposure (Martin et al., 2020). In computational models of auditory system, certain functions are used to collapse or reduce internal representations of signals into a scalar for the purposes of making a decision, such as 'detect' vs 'not detect'. This process is called a decision device, and in auditory amplitude modulation models these functions are frequently implemented as time-domain 'signal crushing' functions such as crest factor (Nelson & Carney, 2006). Taking inspiration from this concept, we asked whether there may be other functions used in the literature for similar purposes, i.e. it should be capable of producing a single scalar value from an arbitrarily long vector, and it should have been used in acoustics (preferably modulation) related studies. The aim in using these functions is to quantify the change in the envelope after being played through the ANC headphone. For all of these adapted functions or metrics (except envelope difference index), the gain is calculated using the relation:

$$\mathrm{M}_{\mathrm{gain}} = \frac{\mathrm{M}(y)}{\mathrm{M}(x)} \tag{2.1}$$

where $M$ is the function in question, $y$ is the envelope of the output signal measured at dummy head and $x$ is envelope of the source signal generated in the computer. Reporting the gain instead of raw value makes the reported values easy to interpret, because it puts all reported values on a common scale (ideal value being 1.0), and the magnitudes indicate how much the recorded signal has changed relative to the original signal. As described in section 1.3, the modulation characteristics of a system are often studied in terms of its modulation transfer function, which is essentially a plot of some computed quantity that represents the system plotted against modulation frequency in the x-axis. Thus, the metrics mentioned below are to be

considered against the modulation frequency axis, to enable a characterization of the headphone in terms of changes in the modulation transfer function caused by the ANC.

**A. Modulation Gain**

One of the frequent questions in amplitude envelope studies is the change in modulation depth of the envelope after something happens to it, such as reverberation or being played through a loudspeaker (Drullman et al., 1994). Modulation depth is usually defined as the ratio between the maximum and the minimum of the envelope signal, and the modulation gain can be quantified as a ratio between the modulation depths of output and the input envelopes. This quantity goes by a number of names in the literature including 'modulation transfer index' (Steeneken, 1992, fig. 4) and 'modulation index' (Harris, 2015). In the present work this is termed as 'modulation gain', and is defined as the gain of the two modulation depths. The modulation depth is defined as

$$\mathrm{m}(x) = 1 - \frac{\min(x)}{\max(x)} \tag{2.2}$$

where $x$ is the Hilbert envelope of the signal, min and max represents the minimum and maximum functions respectively. It is interesting to note that this metric is not at all 'sensitive' to the change in shape of the envelope signal, inasmuch as it is only defined in terms of the min and max of the signal. This is an important property which may help acoustical engineers focus on aspects of the signal that are perceptually relevant, as opposed to get distracted by changes that are quantifiable but don't matter. Modulation depth change can result in loss of speech intelligibility, poor music quality (Harris et al., 2018) and other issues in the perception of the playback signal. In the auditory psychophysics and neuroscience literature, perception of the modulation depth by the brain is widely studied, such as differences in modulation depth detection thresholds across normal and impaired-hearing individuals (Wakefield & Viemeister, 1990; Schlittenlacher & Moore, 2016).

**B. Envelope Difference Index**

Proposed by Fortune et al. (1994), the envelope difference index (EDI) is the mean absolute difference between the envelopes of the original and the output signals. That is, for two envelopes both of length $N$, the EDI is given by

$$\mathrm{EDI}(e1, e2) = \frac{\sum_{i=1}^{N} |e1(i) - e2(i)|}{2N} \tag{2.3}$$

This is obviously extremely sensitive because it quantifies all differences in the two envelopes, and can give spuriously high values if the recording delay is not eliminated properly before computing this quantity. EDI

has been used to study the effect of hearing aid compressor parameters such as attack time and release time, and its corresponding sensation and speech intelligibility differences (Souza et al., 2012), where it was found that different parameters of the compression interact with audibility and these effects can be predicted to a certain extent by this quantity. EDI has been used to fine-tune hearing aid compressor parameters (attack and decay times which affect the envelope shape) (Jenstad & Souza, 2005), and is said to have an acceptable upper limit of 0.2 beyond which consonant recognition degrades significantly. An extension of EDI has been proposed, which discards the portions of the signal which are not relevant to the speech perception, and operates on the first derivative of the temporal envelope instead of the raw envelope; this is said to improve EDI because it more accurately reflects human perceptual needs (Hoover et al., 2012).

**C. Peak of Cross-Correlation**

Cross-correlation is a function that measures the similarity between two signals, by shifting one of them and measuring the similarity as a function of this shift. Cross-correlation is thus a measure of similarity between a signal ($x$), and delayed copies of another signal($y$), as a function of the delay. It is defined as

$$R_{xy}(m) = \mathbb{E}\{x_{n+m}y_n^*\} \tag{2.4}$$

Where $\mathbb{E}$ is the expectation operator, $*$ denotes complex conjugation, and $-\infty < n < \infty$. Backman (1997) used the normalized cross-correlation between original and hearing-aid processed envelopes as a metric to quantify the degradation of speech signals. In models of binaural auditory information processing, cross-correlation is commonly used for coincidence detection (Trahiotis & Bernstein, 2014). Computational psychoacoustic models of AM perception often use cross-correlation as a 'decision device', i.e. to compare intermediate representations based on a single number and arrive at a decision of detection or not (Dau et al., 1996; Jepsen et al., 2008). Speech intelligibility or quality models also adopt cross-correlation to compute similarity between internal representation of the input signals, such as PEMO-Q (Huber & Kollmeier, 2006) which is used in chapter 5. Taking the maximum value, i.e. the peak of the cross-correlation vector thus allows us to represent the highest similarity between two signals, which in the context of envelopes provides an indication similar to the EDI.

**D. Standard Deviation Gain**

This metric is inspired by the concept of 'transient of mean square deviation' discussed by (Ogonowski, 2003). The standard deviation is a well-known and widely used statistical concept, and for a signal $x(n)$ is

defined as

$$\sigma = \sqrt{\frac{\Sigma[x(n) - \mu]^2}{N - 1}} \tag{2.5}$$

where $N$ is the length of $x(n)$, and $\mu$ is its mean. Standard deviation is frequently used in descriptive and inferential statistics to describe the dispersion, i.e. degree to which the numbers are 'spread out' relative to the mean value. When applied to a signal, the standard deviation represents the mean fluctuation of power of the signal (Smith, 1999, p. 14). Standard deviation is sensitive to small fluctuations in the input signal and therefore is ideal for quantifying the relatively small changes in the envelope due to ANC. Standard deviation has been used for designing optimal signals for novel radar and sonar applications for detecting plastic and other non-elastic objects (Hamschin & Loughlin, 2013). Standard deviation has also been utilized for quantifying the change in amplitude envelope of a signal due to dynamic range compression in hearing aids, and was found to be highly correlated with behavioral measure of sensitivity (Sabin et al., 2013).

**E. Kurtosis Gain**

The Kurtosis of a signal $x(n)$ is defined as

$$\mathrm{K} = \frac{\mathbb{E}[x(n) - \mu]^4}{\sigma^4} \tag{2.6}$$

where $\mathbb{E}$ is the expectation operator, $\mu$ is the mean of $x(n)$ and $\sigma$ is the standard deviation of $x(n)$. Kurtosis is a measure of signal distortion, and widely used in contexts like mechanical fault diagnosis (Marchesiello et al., 2015). In auditory psychophysics, kurtosis has been used as a decision device in models of amplitude modulation perception (Lorenzi et al., 1999). Kurtosis has been used as a measure of noise-exposure and consequently to inform protocols for the protection of hearing (Lei et al., 1994; Hamernik et al., 2003; Davis et al., 2009). The Kurtosis function has also been used to characterize crackling sounds; 'crackle' is a perceptual aspect of many sounds such as fire sounds, and it had been hitherto hard to define solely in terms of the long-term spectral characteristics (Swift et al., 2017).

**F. Crest Factor Gain**

The crest factor of a signal $x(n)$ of size $N$ is defined as

$$\mathrm{C} = \frac{\max[x(n)]}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}|x(n)|^2}} \tag{2.7}$$

Thus, crest factor is also known as the 'peak-to-rms' ratio of the signal. In music production, artists and

composers use a variety of signal monitoring and shaping tools using digital audio workstations to arrive at the desired signal combinations. One of the time-domain metrics such producers and artists observe is crest factor, which is related to the 'mellow' sounds and quantify the concept of percussiveness (Ziemer et al., 2020). In auditory psychophysics, Rybicka and Sęk (2003) studied the detection and discrimination of various amplitude modulated signals with different crest factors and concluded that high crest factor enables listeners to detect AM more easily than the traditional models suggested. Crest factor is also considered an important metric for quantifying the noise exposure of wild animals, e.g. marine mammals exposed to drilling noises (Martin et al., 2020). Crest factor has also been used to characterize loudspeaker distortions and to quantify relationships between signal power and acoustic power (Zuccatti & Bandiera, 2009). Considering these diverse applications of the function, in the present work, the change in crest factor of the envelope is taken to be relevant for the psychological aspects of the playback signal.

**G. Loudness Gain**

Loudness is the perceptual dimension of sound which closely depends on the intensity, and also on fluctuations over time (amplitude modulation). Many models of loudness have been proposed, and they are able to predict human perception of loudness to varying degrees (Plack & Carlyon, 1995; Marozeau, 2011). Here, the loudness model implemented by the function acousticLoudness in MATLAB audio toolbox [1], which implements the loudness model standardized by ISO 532-1 (ISO, 2017). This function was utilized to calculate the loudness gain (as per equation 2.1) and the resulting values are reported. This and the next metric is not defined purely in terms of envelope, i.e. it is not purely a quantification of the change of envelope of the playback signal. Nevertheless, it is an important quantification of the effect that ANC has on the playback signal.

**H. Sharpness Gain**

Sharpness is another perceptual dimension of sound which is more complex than loudness or pitch. It is not a time-domain metric; rather, it is based on a time-frequency representation and the relative distribution of higher and lower frequencies (Swift et al., 2017). Thus, a relative reduction of sharpness indicates a relative increase of lower frequencies and vice versa. Sharpness is known to be correlated with listener preferences and annoyance in noises like engines and domestic appliances (Zwicker & Fastl, 2013, chapter 9). As discussed in section 1.2, number of studies have attempted to create 'psychoacoustic ANC' systems that evaluate themselves using perceptual models of sharpness (Gonzalez et al., 2003; Bao & Panahi, 2010a).

[1] https://in.mathworks.com/help/audio/ref/acousticloudness.html

Here, the MATLAB audio toolbox function acousticSharpness [2] was utilized, which is also standardized in ISO 532-1 (ISO, 2017) to compute a value of perceived sharpness. Similar to loudness gain, this function was utilized to calculated the gain (equation 2.1) and the resulting values are reported. Although primarily defined as a frequency-domain quantity, sharpness has also been studied in the context of amplitude modulation domain, e.g. for triangular and rectangular shapes of the envelope (Ueda & Akagi, 1990). Thus a change in sharpness can indicate a corresponding underlying change in the envelope. Since sharpness models are validated against human perception, it would be interesting to correlate the results obtained from this metric with the more basic changes reported by the simple 'physical' metrics like kurtosis etc.

## 2.3 Materials and Methods

### 2.3.1 Test signals and computations

A set of sinusoidally amplitude modulated signals were used as inputs to the headphones. The following subsections lay out the parameters and other details of these signals. For carrying out both the dummy-head microphone recording and computations, MATLAB software (version R2021a) was used.

The signals were generated in time-domain, which took the form

$$y = [1 + m.\sin(2\pi f_m t)].a.\sin(2\pi f_c t) \tag{2.8}$$

where $f_c$ is the carrier frequency and $f_m$ is the modulator frequency, and $m$ is modulation depth which will be reported in dB scale (= $20\log_{10}(m)$) in outputs. To avoid sudden onset and offset effects, the signals were gated with a tapered cosine window for both the onset and offset, for 0.1 seconds. The total duration of each signal (including gating window) was 2.0 seconds. The modulation depth was kept at 0 dB (i.e. $m$=1). Scaling factor $a$ was kept at 0.66 to ensure adequate signal power without clipping distortions.

10 carrier frequencies were used, starting from 200 Hz and ending in 2500 Hz in a roughly logarithmic space, using the audspace function of AMtoolbox [3]. The values of the carrier frequencies are provided in table 2.1. Modulator frequencies were in range of 1 Hz to 50 Hz, with a step size of 1 Hz. Thus, a total of 500 (50*10) base signals were generated. Each of these signals were recorded with the 'ANC on' and 'ANC off'

[2] https://in.mathworks.com/help/audio/ref/acousticsharpness.html
[3] https://www.amtoolbox.org/

Table 2.1: Carrier frequency values

| No. | Frequency (Hz) |
|---|---|
| 1. | 200 |
| 2. | 298 |
| 3. | 418 |
| 4. | 566 |
| 5. | 747 |
| 6. | 970 |
| 7. | 1244 |
| 8. | 1580 |
| 9. | 1993 |
| 10. | 2500 |

conditions of each headphone under test. After recording, the signals were stored (without compression, in wav format) for offline analysis. During the analysis, the recorded signals were time-aligned with the original signals using ALIGNSIGNALS [4] function which uses a correlation based method to find and eliminate the time delay. This time alignment is important to avoid errors and ensure accurate comparisons of original and recorded envelopes.

Envelopes of the recorded signals were computed using the Hilbert transform. The Hilbert transform implemented by the HILBERT [5] function from MATLAB signal processing toolbox which uses an FFT-based method (Marple, 1999), which is implemented as the algorithm given in table 2.2. The envelope is calculated by taking the absolute value of the result of the Hilbert transform. The envelopes were then analyzed according to the metrics defined above. The scores or values obtained from each envelope metric were analyzed to see if there were any statistically significant differences across headphones and across 'ANC on' and 'ANC off' conditions, with the help of a linear model (ANOVA).

Table 2.2: Steps for Hilbert transform

| Input: s |
|---|
| Step 1. Calculate the FFT of s, storing the result in a vector x.<br>Step 2. Create a vector h whose elements h(i) have the values:<br>1 for i = 1, (n/2)+1<br>2 for i = 2, 3, … , (n/2)<br>0 for i = (n/2)+2, … , n<br>Step 3: Calculate the element-wise product of x and h.<br>Step 4: Calculate the inverse FFT of the sequence obtained in step 3 and returns the first n elements of the result. |

[4] https://in.mathworks.com/help/signal/ref/alignsignals.html
[5] https://in.mathworks.com/help/signal/ref/hilbert.html

### 2.3.2 Recording environment

The signals described above were recorded with the hardware setup described in section 1.5. The input gains of the microphones in the USB interface were adjusted to ensure no clipping occurred in the recorded signals. The headphones were placed on the dummy head to ensure a good fit. The recording was carried out in a quiet office environment which had an average sound level of 50 dbA, which was mainly from the air-conditioning system. The audioPlayerRecorder and associated functions were used for conducting simultaneous playback and recording with the USB audio interface. The recording script was written to check for overflow and underflow errors (i.e. samples being rejected due to underlying issues with the synchronization between computer clock and audio interface clock), and any signal that suffered from these types of error was discarded and the recording was repeated. The recorded signals were stored in uncompressed wav format for later analysis.

## 2.4 Results and discussions

The plots of different envelope metrics are shown in figures 2.2 to 2.8. In each case, the error-bars represent standard error of the mean, where the metric value is averaged across the 10 carrier frequencies (as defined in section 2.3). The mean and standard deviation of the metrics across different headphones and ANC conditions are reported in table 2.3. These values are then compared using analysis of variance (ANOVA) to test for overall differences across the four headphones and across 'ANC on' and 'ANC off' conditions, the corresponding F and p-values are reported in table 2.4. A discussion on F and p-values and their interpretation can be found in standard statistics textbooks, such as (McDonald, 2009). Effect sizes and their usefulness in statistical inference are discussed in detail in (Ellis, 2010).

Overall, according to the plots of most metrics the ANC is seen to be providing a less distorted envelope. It is also obvious that not all headphones distort the signals to the same extent. The different metrics help illustrate different aspects of the signal, and hence do not produce the same magnitude of the 'ANC effect' even though the input signals are the same for all of them. Additionally, the size of the errorbars in each plot helps illustrate the amount of dependence that the headphone and that metric has on the carrier frequency, ideally all errorbars should be as close to zero as possible because modulation frequencies do not depend on the carrier frequencies.

In table 2.4, the two factors (headphone type and ANC condition) are answering the questions whether the metric has any significant differences across the 'ANC on' and 'ANC off' conditions, and whether any of the headphones are different from the rest. The interaction between these two factors answer the question whether the 'ANC effect' (if any) are significantly different across different headphones. The effect size of each factor and interaction quantifies the strength of such differences, with higher values indicating more differences. Since all the metrics are statistically highly significant across both the factors, the effect size helps understand the relative change as detected by each metric (Ellis, 2010). As expected, modulation gain shows one of the smallest effect sizes. Similarly, loudness shows the largest effect sizes.

These results establish that ANC headphones do help deliver a less distorted envelope, and the nature or degree of this distortion can be quantified and visualized using a variety of metrics related to envelope research. These results are not meant to serve any definitive purpose such as making recommendations about headphone goodness; rather, the goal here is to inspire the beginning of a dialogue across disciplines to help fill the significant gap of evaluation and characterization of smart audio devices as discussed in chapter 1.

This work presents an important initial exploration into a topic that is ripe for investigation, yet quite unexplored, i.e. the modulation transmission properties of ANC headphones. These methods can serve as a complement to the existing assessment protocols for defects or perceptual eccentricities when a new ANC device or algorithm is proposed. The different time-domain and psychoacoustically motivated metrics serve to illustrate different aspects of change of the signal, such as the change in 'peakiness' with kurtosis or simply the change in depth by modulation index. By plotting the gain of various physical time-domain metrics, and comparing the scores across 'ANC on' with 'ANC off', we gleaned an insight into the potential benefits conferred by the ANC on the playback signal. Other metrics from statistics or signal processing could be used as well for gaining deeper understanding of these changes.

One important caveat regarding the work reported in this chapter is as follows: when it comes to audio device evaluations, the ears do not always correspond to what the eyes see in the form of various signal representations. Toole (1986a) emphasized that "the auditory perceptions of musical sounds are not the same as visual analyses of data from clinical sine waves, impulses, and pink noise". This chapter offers a robust alternative to these problems by characterizing the headphones in the modulation frequency domain; furthermore, any transient and visually over-represented distortion would be minimized when the entire 2 seconds of the signal is 'crushed' to produce a single number and then that number is compared against the

source signal. Indeed, as most of the results show in figures 2.1 to 2.8, in most metrics the response is fairly stable across the modulation frequency axis.

Table 2.3: Mean and standard deviation of envelope metrics across headphones

| | ANC | QC25 | | HD 4.50 | | WH-1000XM3 | | WH-1000XM4 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Stdev | Mean | Stdev | Mean | Stdev | Mean | Stdev |
| Modulation gain | On | 0.999501 | 0.0008365 | 0.999516 | 0.000321 | 0.999239 | 0.0009142 | 0.999416 | 0.0005841 |
| | Off | 0.997198 | 0.008501 | 0.999143 | 0.0008397 | 0.998776 | 0.00128 | 0.998871 | 0.0008694 |
| Envelope Difference Index | On | 0.000809366 | 0.0006252 | 0.00126339 | 0.0005809 | 0.00130076 | 0.0003812 | 0.00193795 | 0.002157 |
| | Off | 0.00264621 | 0.001381 | 0.00219959 | 0.0009247 | 0.00293748 | 0.001543 | 0.00316568 | 0.0011189 |
| Peak cross-correlation | On | 111.362 | 1.958 | 111.259 | 1.93 | 111.288 | 1.887 | 110.246 | 3.803 |
| | Off | 110.187 | 2.582 | 110.691 | 2.079 | 109.855 | 2.654 | 109.777 | 2.306 |
| Loudness gain | On | 1.00975 | 0.006422 | 1.15188 | 0.05089 | 1.03482 | 0.01813 | 1.02153 | 0.02789 |
| | Off | 1.10089 | 0.06849 | 1.22361 | 0.06934 | 1.13372 | 0.08897 | 1.21056 | 0.1095 |
| Sharpness gain | On | 1.00903 | 0.01609 | 1.12769 | 0.1524 | 0.987805 | 0.02409 | 0.991142 | 0.02726 |
| | Off | 0.937082 | 0.055 | 1.0907 | 0.2906 | 0.909754 | 0.06797 | 0.899016 | 0.09507 |
| Kurtosis gain | On | 1.00168 | 0.01414 | 0.997657 | 0.02106 | 0.998156 | 0.004219 | 1.0642 | 0.231 |
| | Off | 1.04959 | 0.1415 | 1.02541 | 0.04972 | 1.06022 | 0.0846 | 1.06662 | 0.07857 |
| Crest factor gain | On | 1.02907 | 0.04912 | 1.04643 | 0.05194 | 1.03771 | 0.04146 | 1.09296 | 0.1783 |
| | Off | 1.11627 | 0.1275 | 1.09757 | 0.09359 | 1.12772 | 0.1167 | 1.16162 | 0.1162 |
| Standard deviation gain | On | 0.990895 | 0.01044 | 0.983051 | 0.009993 | 0.984005 | 0.008243 | 0.974866 | 0.03927 |
| | Off | 0.952073 | 0.03296 | 0.961148 | 0.0226 | 0.941614 | 0.04044 | 0.939299 | 0.02792 |

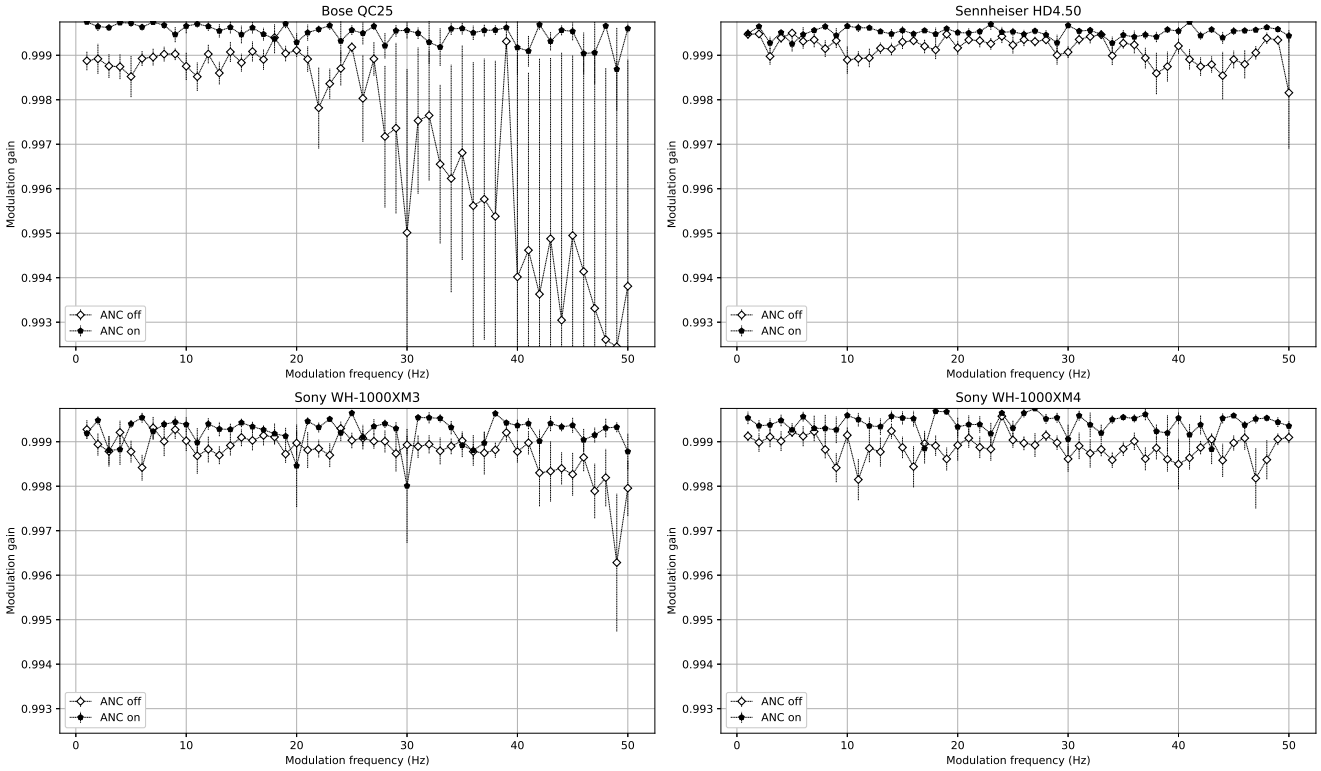



Figure 2.1: Modulation gain against modulation frequency

Table 2.4: Statistical comparison (two-way ANOVAs) of envelope metrics

| Dimension | factor | F-value | p-value | Effect size ($\eta_p^2$) |
|---|---|---|---|---|
| Modulation gain | Headphone Type | F(3,3992)=18.79 | **<0.001** | **0.01** |
| | ANC condition | F(1,3992)=87.72 | **<0.001** | **0.02** |
| | Interaction | F(3,3992)=22.10 | **<0.001** | **0.02** |
| Envelope Difference Index | Headphone Type | F(3,3992)=101.7 | **<0.001** | **0.07** |
| | ANC condition | F(1,3992)=1315 | **<0.001** | **0.25** |
| | Interaction | F(3,3992)=27.11 | **<0.001** | **0.02** |
| Peak cross-correlation | Headphone Type | F(3,3992)=28.18 | **<0.001** | **0.02** |
| | ANC condition | F(1,3992)=135.8 | **<0.001** | **0.03** |
| | Interaction | F(3,3992)=8.915 | **<0.001** | **0.007** |
| Loudness gain | Headphone Type | F(3,3992)=782.8 | **<0.001** | **0.37** |
| | ANC condition | F(1,3992)=3065 | **<0.001** | **0.43** |
| | Interaction | F(3,3992)=164.1 | **<0.001** | **0.11** |
| Sharpness gain | Headphone Type | F(3,3992)=384.3 | **<0.001** | **0.22** |
| | ANC condition | F(1,3992)=309.4 | **<0.001** | **0.07** |
| | Interaction | F(3,3992)=8.724 | **<0.001** | **0.007** |
| Kurtosis gain | Headphone Type | F(3,3992)=46.90 | **<0.001** | **0.03** |
| | ANC condition | F(1,3992)=109.3 | **<0.001** | **0.03** |
| | Interaction | F(3,3992)=14.93 | **<0.001** | **0.01** |
| Crest factor gain | Headphone Type | F(3,3992)=60.54 | **<0.001** | **0.04** |
| | ANC condition | F(1,3992)=486.0 | **<0.001** | **0.11** |
| | Interaction | F(3,3992)=7.210 | **<0.001** | **0.005** |
| Standard deviation gain | Headphone Type | F(3,3992)=71.76 | **<0.001** | **0.05** |
| | ANC condition | F(1,3992)=1649 | **<0.001** | **0.29** |
| | Interaction | F(3,3992)=27.51 | **<0.001** | **0.02** |

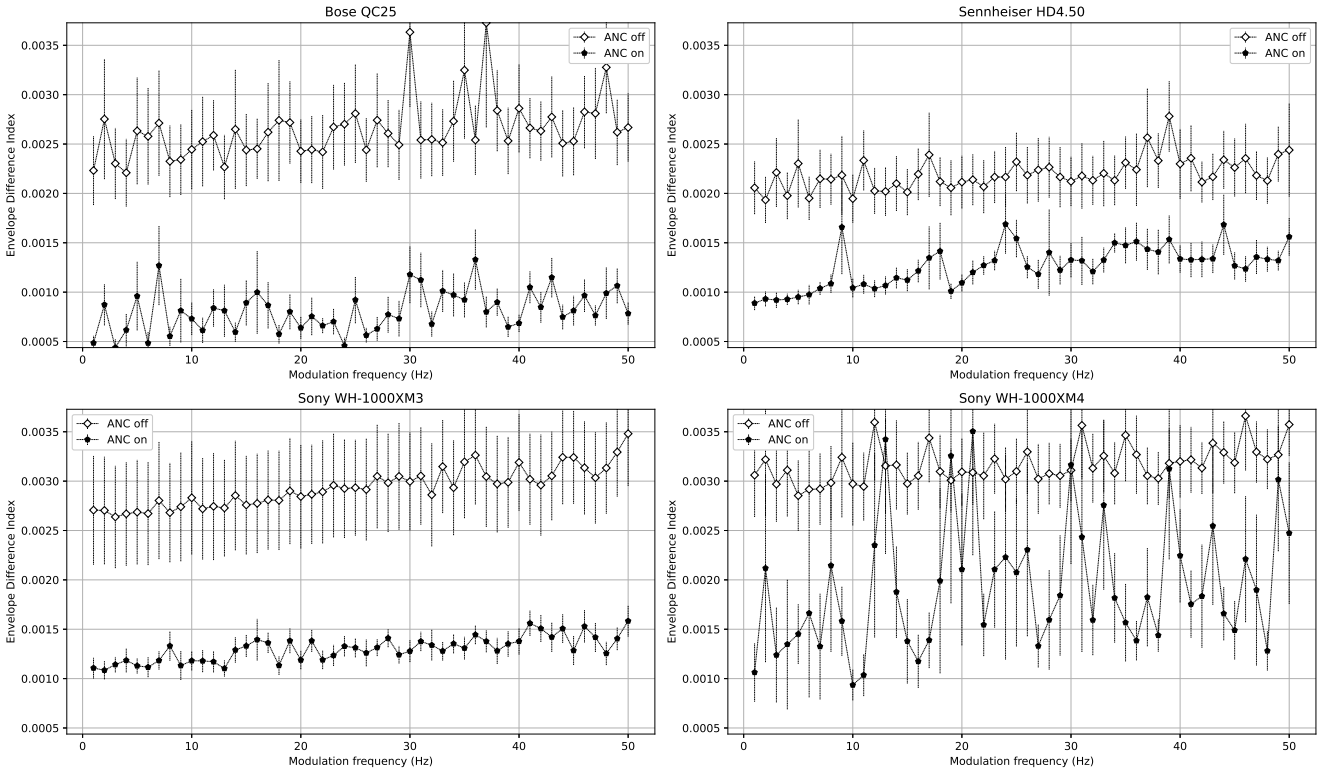



Figure 2.2: Envelope Difference Index against modulation frequency

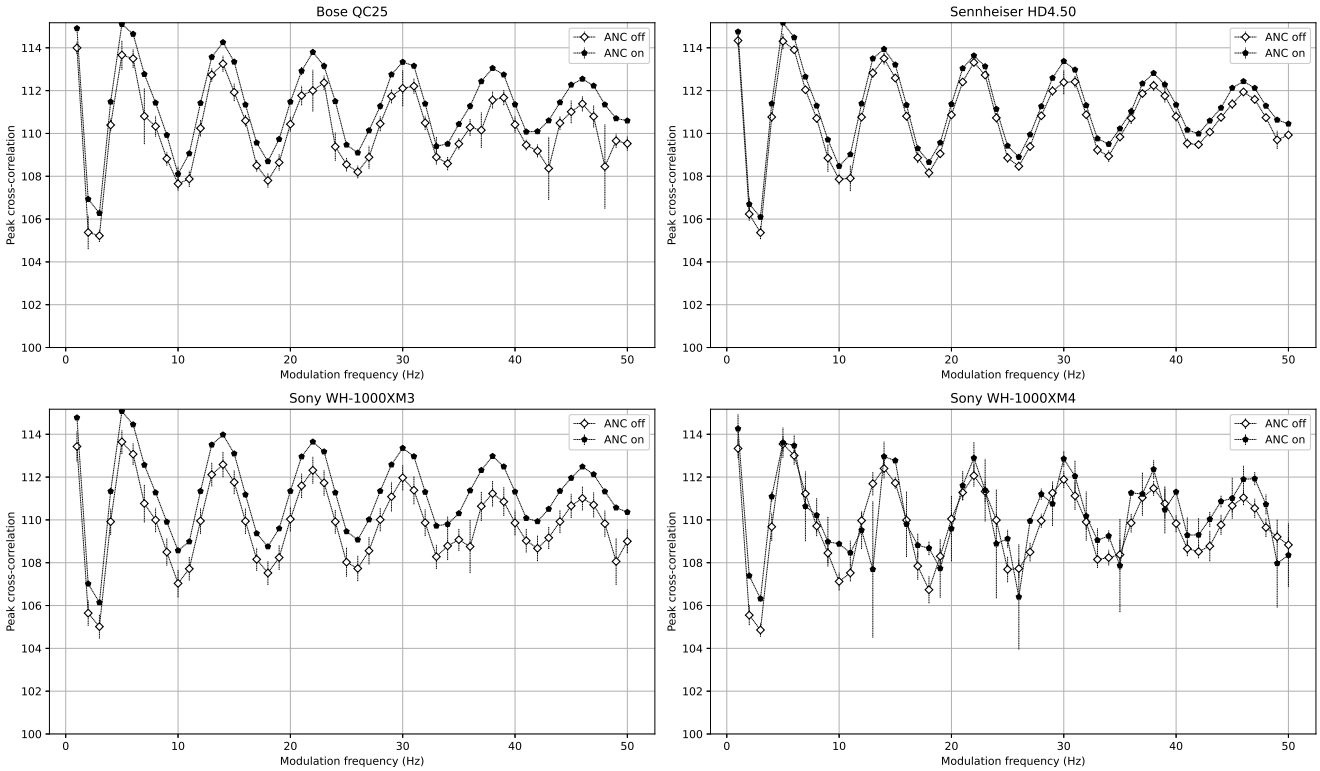



Figure 2.3: Peak cross-correlation against modulation frequency

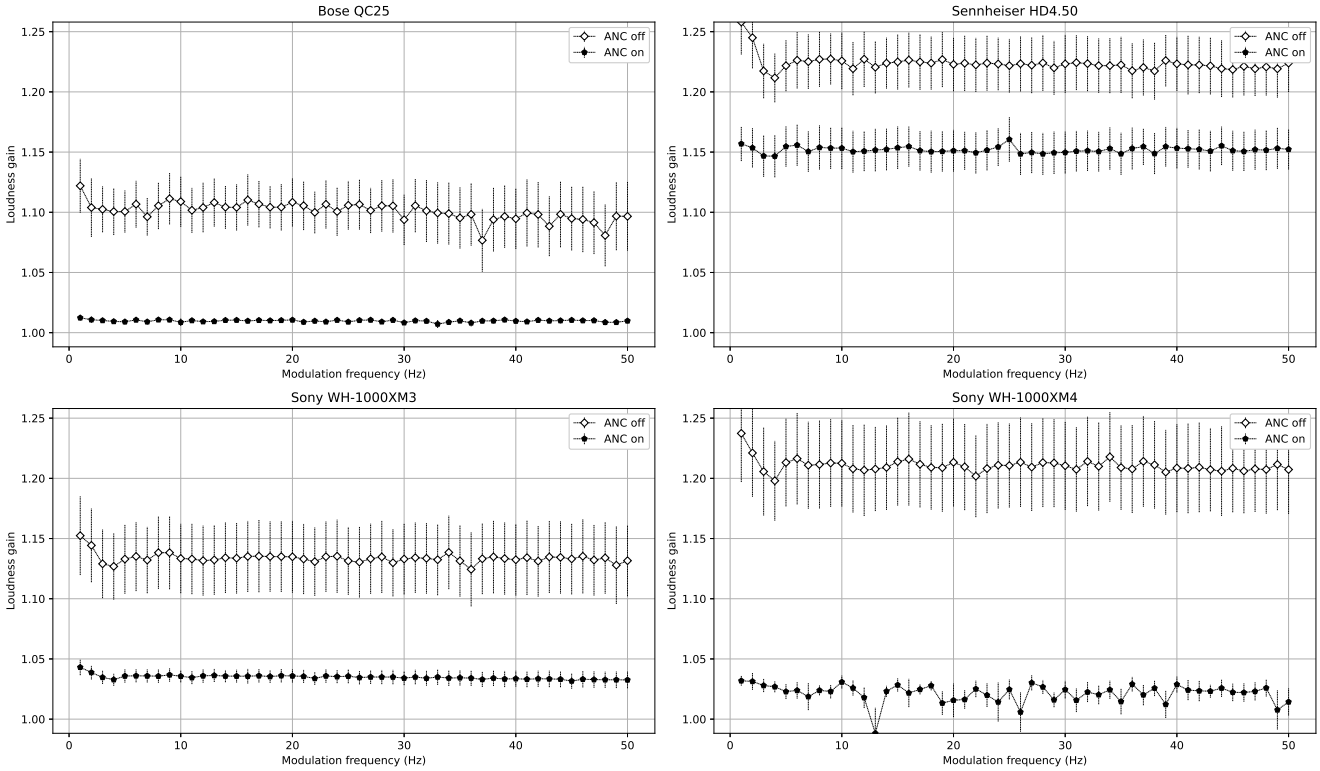



Figure 2.4: Loudness gain against modulation frequency

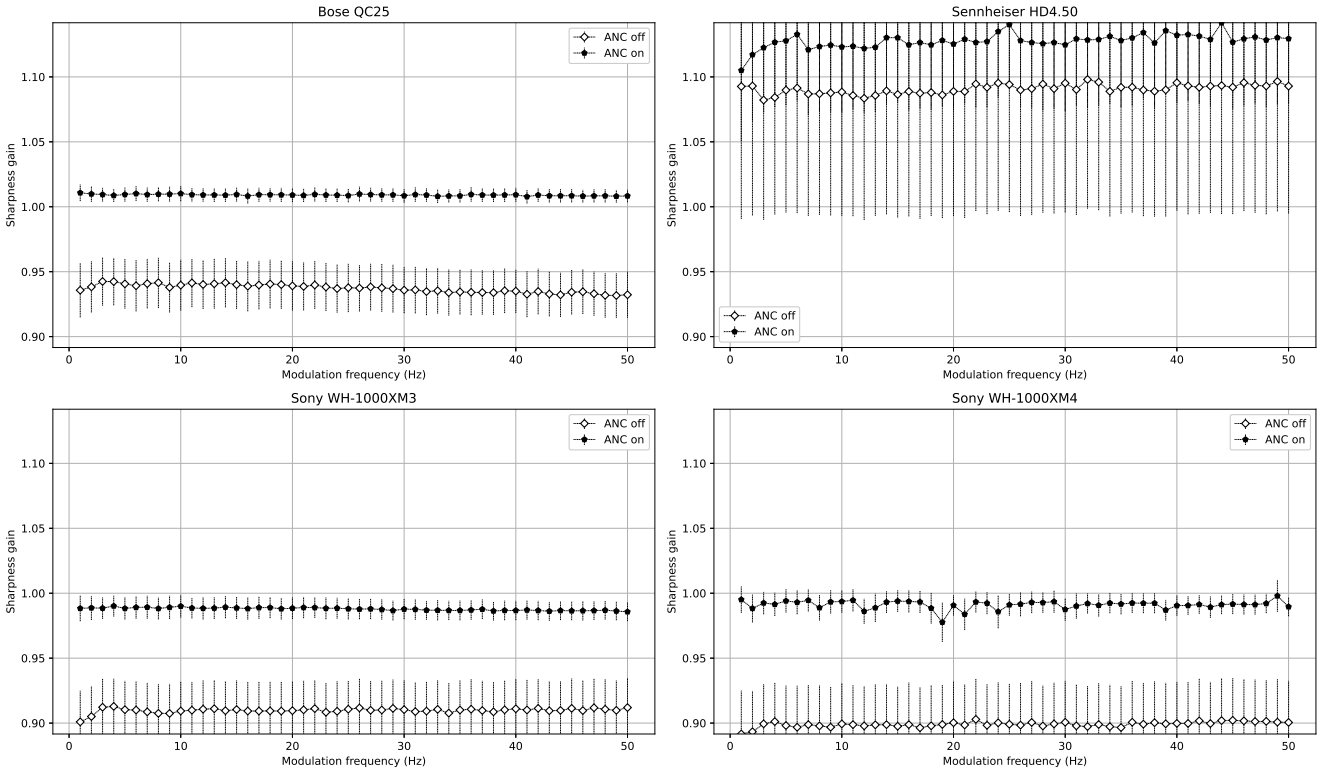



Figure 2.5: Sharpness gain against modulation frequency

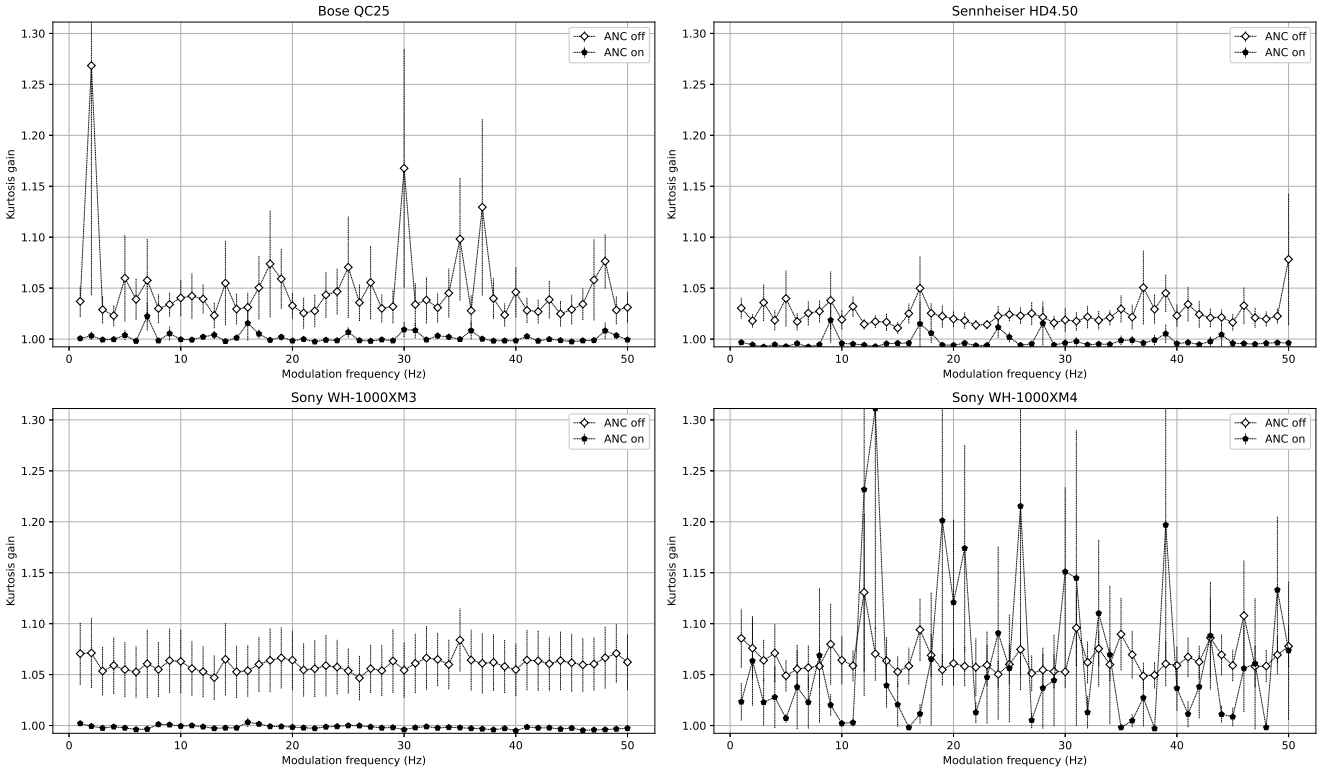



Figure 2.6: Kurtosis gain against modulation frequency

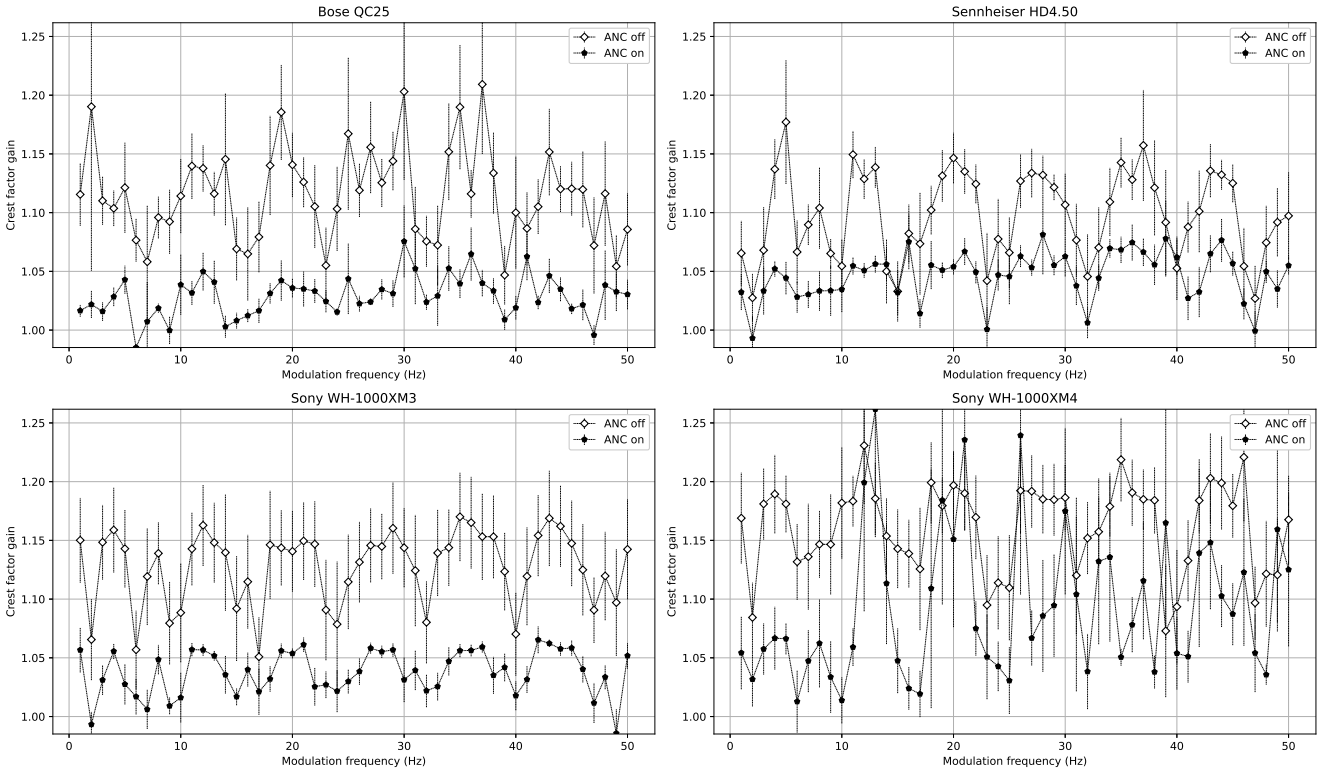



Figure 2.7: Crest factor gain against modulation frequency

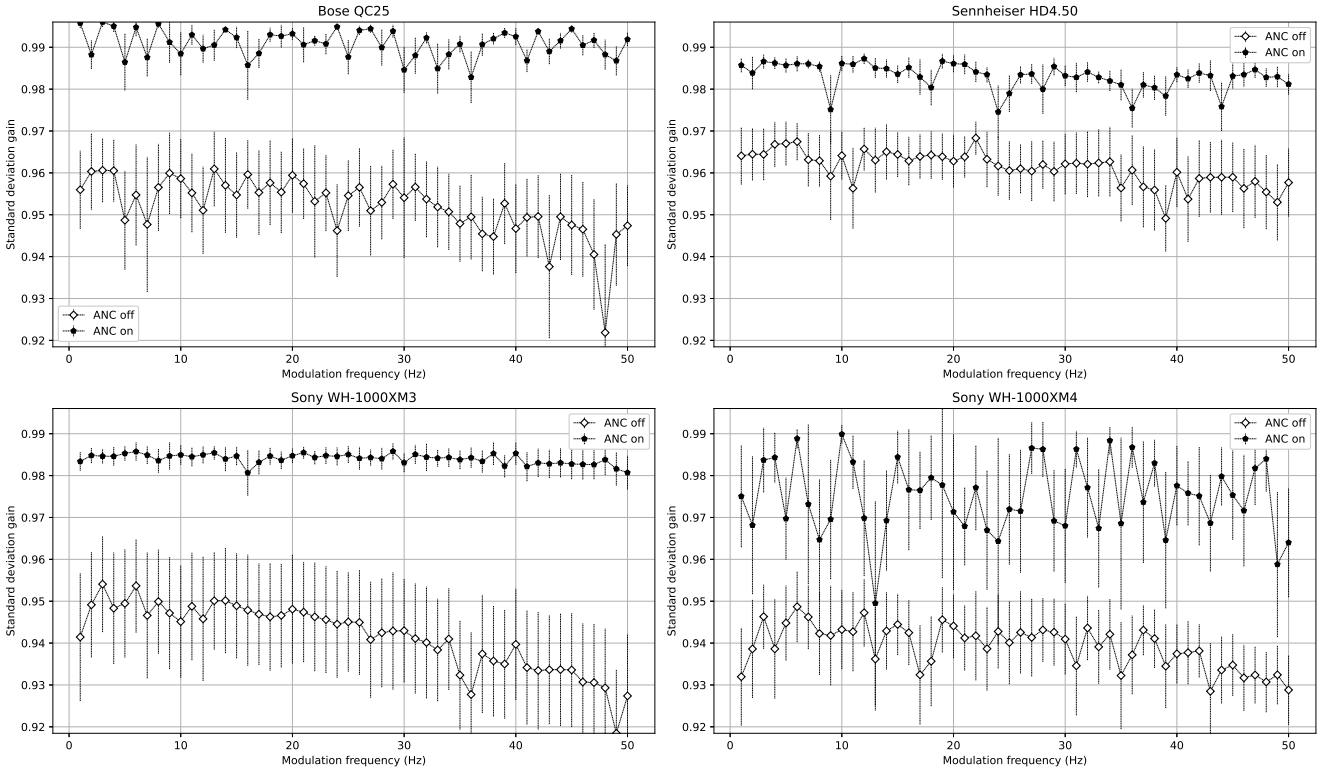



Figure 2.8: Standard deviation gain against modulation frequency

## 2.5 Chapter summary

This chapter characterized four ANC headphones using the amplitude envelope of the playback signal. The headphones were used to play a range of modulated sinusoids and the resulting signals were recorded using a dummy-head microphone. The amplitude envelope of the resulting recorded signals were compared with the original envelopes to quantify the distortion across the ANC conditions, and across four headphones under study. Eight metrics were chosen for computing the distortion of envelope, most of which were time-domain statistical functions such as kurtosis. The values of these metrics were averaged across the carrier frequencies and plotted against the modulation frequencies, to obtain the modulation transfer functions for each headphone and ANC condition. Such functions serve as important visual tools to highlight and characterize a system in terms of its modulation transmission capabilities, much like the 'transfer function' in the audio-frequency domain. Statistical tests (ANOVA) revealed significant differences between the 'ANC on' and 'ANC off' conditions in all the metrics, and also between the headphones. Furthermore, the interaction between the ANC factor and the headphone factor was significant, implying that the 'ANC effect' on the envelope of the playback signal is not the same across all the headphones.

### 2.5.1 Future work

The work reported in this chapter is a preliminary investigation into the modulation characteristics of ANC headphones. These methods and analyses can be extended in the following ways:

- Other statistical functions could be developed to track the envelope fidelity. This could be done by taking inspiration from computational neuroscience to better understand what kinds of distortions are actually relevant for the auditory system of the brain. Such metrics would pave the way for future studies in control system engineering as well as signal processing for understanding the effect of these devices on human perception.

- Continuing on the above, additional envelope fidelity metrics could be constructed based on the modulation spectrum. That is, in addition to the time-domain representation of the envelope signal, the spectrum of the envelope can be computed and analyzed using some kind of weighted-sum across modulation frequencies, such as the analysis of music playback in vehicular ANC done by Basu et al. (2021).

- Schroeder (1981) had developed a method for estimating the MTF of a linear passive system using the square of the Fourier transform of the impulse response of the system. ANC violates the assumptions of passive and linear systems, but this idea could potentially be extended by applying some kind of probabilistic signal processing theory, and by taking into account the full set of properties such as phase of the envelope signal, e.g. (Backman, 2010). This would eliminate the need for carrying out time consuming measurements (recordings) such as the ones reported in this chapter, and allow researchers to automate the assessment of future ANC systems.

- The performance of ANC system may be degraded in a variety of ways, such as mechanically induced vibration of the cables, variation of the acoustic load, and many others (reviewed by (Jiang & Li, 2018; Liebich, Fabry, et al., 2018)). The interaction between the ANC system and the external sound field can be complex, and proper models of such interaction could be developed to study the benefits on the playback signal in more realistic conditions.

- The findings of this chapter could be integrated back into the ANC design process itself in consideration of the controller architecture, such as hinted by comparative advantage of feedforward over feedback architectures in terms of speech intelligibility (Brammer et al., 2005).

# 3

# Simulated neurons' response to amplitude modulated signals through ANC

## 3.1 Introduction

This chapter characterizes ANC headphones in terms of responses obtained from simulated computational models of neurons. As with the goals of perceptual evaluation, we are interested in using models of hearing (or specific parts of it) to characterize the behavior of audio devices (Bech & Zacharov, 2007). In chapter 2 we observed some physical differences caused by ANC in the headphones' playback signal in terms of the statistical properties of the signals' envelope, such as kurtosis gain. To extend that analysis, we now turn to more advanced and neuroscientifically realistic models of perception. Hence in this chapter, computer simulated models of the auditory pathway of the brain are being utilized as virtual observers. The modulated sinusoids described in section 2.3 were given as inputs to computer simulated neuron models of the auditory pathway. The modulation representation is studied in these neurons' responses, in order to find whether the ANC caused any difference in such representation.

### 3.1.1 Related work

Computational neuroscience of sensory systems attempt to build bridges between stimuli, sensation and behavior. Researchers working in disciplines related to audio and speech processing have frequently found applications of models of the auditory system, i.e. the models from neuroscience research have been able to improve specific tasks in audio and speech processing technologies. For instance, Hines (2012) developed a model of speech intelligibility based on neural responses, which was later refined into an instrumental audio evaluation metric named VISQOL (Hines et al., 2015), which has been used in chapter 5. Much of computational neuroscience of the auditory system is dedicated to the problem of signal representation, i.e. the neural firing patterns that correspond to various patterns in the externally delivered stimuli (Eggermont, 2001; Rudnicki et al., 2015). The responses of inner hair cell and auditory nerve synapse models have been used to predict the behavioral outcomes such as speech intelligibility (Hauenstein, 1998; Wirtzfeld et al., 2017). Baby and Verhulst (2018) found that deep neural networks being trained for noise suppression tasks performed better when a hybrid biophysical based cochlear model was used as a pre-processing or feature extraction mechanism. Lyon (2017) summarized in detail various applications of the models of the human auditory system including cochlea, auditory nerve fibers and more in applications such as predicting pitch perception, speech transcription, source separation, noise removal, object detection and more by artificial listeners a.k.a 'machine hearing'. Reimes (2019) formulated a signal processing model (not neuroscience based) for predicting the 'listening effort' (a psychological measure) of human participants when using ANC headphones. To the best of the authors' knowledge, no model of auditory nerve or brainstem neurons has been used to characterize or evaluate ANC headphones.

### 3.1.2 Models of auditory neurons

In this section, the neuroscientific findings and models of the auditory pathway of the brain are reviewed briefly, and the neural centers used in the present work are described in terms of their amplitude modulation properties.

The mammalian auditory system processes sounds by first filtering it through the cochlea, in which a structure known as basilar membrane splits the signal into frequency components due to the mechanical/physical properties. It also acts as an adaptive filterbank, allowing the brain to control the frequency tuning of the filters as needed. The basilar membrane's motion causes small mechano-sensitive structures

called inner hair cells to get excited and release certain chemical messengers (called neurotransmitters) into gaps (called synapses) which then excite auditory nerve fibers (neurons). The auditory nerve fibers 'meet' other neurons at the cochlear nucleus, which is the first synapse for auditory information. Neurons from the cochlear nucleus then form synapses with higher nerve centers such as superior olive, inferior colliculus (Møller, 2012; Rudnicki et al., 2015; Plack, 2018). This entire process has been studied with the help of biological as well as simulation approaches. The standard tools and methods for doing this includes giving systematically controlled naturalistic or artificial stimuli to an organism, and collecting its brain response in the form of neuronal membrane potentials or EEG recordings or other forms of measurement. Simultaneously, similar stimuli are used to elicit behavioral responses from animals and more importantly from human participants. These two kinds of data are then subjected to correlation based analyses to find links between the stimuli and sensation. One common method to investigate the representation of stimuli in the brain's sensory pathway includes statistically analyzing the spike-train properties of a neuron (Kass et al., 2018). The biological data are collected from animals (such as rodents or cats) and analyzed in specific stimulus paradigms, e.g. tone-on-tone masking. Large volumes of such data have been then assimilated into computational models of the auditory pathway, some of which are used in this chapter.

Since the 1940s, researchers have attempted to extract the essence of the processes of hearing into mathematical abstractions to glean deeper insights into the apparent diversity of phenomena (Galambos & Davis, 1943; Jeffress, 1948); reviewed and discussed by Meddis et al. (2010, chap. 1). Computational models of the auditory system are frequently developed as a cascade of stages that are anchored along the route of the information, from the cochlear nonlinearities, basilar membrane, electrical properties of outer and inner hair cells, synapse chemicals and finally the nerve potentials of the auditory nerve fibers. These 'way points' serve as benchmarks to validate the computational models; by comparing the model outputs of such cascades with physiological data the researchers establish the validity of their model (Meddis & Lopez-Poveda, 2010). Some researchers classify models of the auditory system into a few groups known as 'biophysical', 'phenomenological', and 'functional effective' (Vecchi et al., 2021, fig. 1). In the present work, only phenomenological models are used. Although derived from some of the knowledge of internal biological structures, the phenomenological models attempt to replicate only the input-output characteristics of the system being studied using mathematical structures (e.g. hierarchical activation networks) without taking into account the precise inner biological or chemical processes that give rise to those characteristics. Phenomenological models have certain advantages, including simplicity and speed of execution; however, this

advantage may come at the expense of interpretability (Trappenberg, 2009, sec. 1.2.1). Like any model, these also provide researchers the ability to study the response to arbitrary stimuli without having to carry out messy and ethically painful experiments on live animals.

Neurons in the mammalian auditory brainstem represents AM parameters like modulation frequency and depth by adjusting the mean firing rate over time, and/or by firing the spikes at specific phases of the AM envelope (Krebs et al., 2008; Rudnicki et al., 2015). The former is usually known as rate-coding and the latter is known as synchronization. These strategies are then distributed across populations of neurons, so that the higher processing centers in the brain (such as cortex) can analyze the incoming spike trains from hundreds or thousands of neurons to find temporal regularities and decide on the nature of the input signals. A significant amount of computational neuroscience literature is devoted to uncovering the specifics of these adaptations and representation tactics (Nelson & Carney, 2006; Zilany et al., 2009).

**A. Inner hair cell and auditory nerve**

The computational models of the inner hair cell (IHC) and auditory nerve (AN) synapse developed over two decades by Carney and colleagues (Carney, 1993; Zilany et al., 2009; Zilany et al., 2014) were used. This model includes simulations of auditory periphery including basilar membrane and cochlear filtering processes with nonlinearities. Internally, the model has two stages: the inner hair cell model and the auditory nerve synapse model; both these outputs are made available to the calling function. The model implements multiple scales of power-law adaptation and is tuned to replicate physiological data including simultaneous and forward masking, compression, suppression, shift in characteristic frequency as a function of level, and adaptation properties and other phenomena observed at the auditory nerve output (Zilany et al., 2009). In terms of amplitude modulation, the model is tuned to reproduce the salient trends in the encoding observed in auditory nerve fibers in cats and other animals across a wide range of carrier frequencies. This model has been used in terms of amplitude envelope representation and other ways to study speech perception (Wirtzfeld et al., 2017). It has also been used to draw parallels between the individual neurons responses and collective electric field fluctuations (known as auditory brainstem response) to understand speech perception (Saiz-Alía & Reichenbach, 2020). The 2014 version of the model has also been validated further and is "…now a better candidate to examine realistic neural-encoding hypotheses" (Zilany et al., 2014).

**B. Cochlear nucleus and inferior colliculus**

The cochlear nucleus is the first nucleus (where neurons 'meet') along the auditory pathway, and encoding

of amplitude modulation parameters such as modulation frequency is known to be enhanced compared to the auditory nerve (Rhode & Greenberg, 1994). As such, it has been heavily studied in neurophysiological experiments and detailed data about representation of AM has been collected, such as the modulation frequency encoding in the firing rate distribution of populations of neurons (Joris et al., 2004). Inferior colliculus is known to be an important 'relay station' or even 'obligatory synapse' for auditory information which encodes temporal structure of the incoming audio waveform, and as such has received significant modeling attention from researchers (Nelson & Carney, 2004, 2007). Computational models of these neural centers have specific assumptions regarding the relation between the model architecture and their corresponding firing properties, and hence some researchers have attempted to simplify and unify existing models to arrive at common shared structural and functional properties of these models (Goodman et al., 2018). The models used in the present work are phenomenological in nature, and were validated against physiological data in terms of both the broad trends and absolute values in terms of average firing rate and synchronization metrics for a variety of inputs with different modulation frequencies, intensities and carrier types and frequencies (Nelson & Carney, 2004). It includes a model structure called 'same frequency inhibition and excitation' which is basically a way of specifying the activation and connectivity between the neurons being simulated; this particular model is apparently able to achieve a vast majority of the observed data with a relatively simple network (Nelson & Carney, 2004, fig. 2). These models expect the output of an auditory nerve model as input, and produce the mean firing rate of the corresponding nuclei as output.

## 3.2 Materials and methods

The signals described in chapter 2 were used as inputs for the computer simulations of neurons. Briefly, these were sinusoidally amplitude-modulated signals of duration 2.0 seconds each, for a set of carrier frequencies ranging from 200 Hz to 2500 Hz, and modulation frequencies ranging from 1 Hz to 50 Hz. These signals were recorded with the four headphones under study, with 'ANC on' and 'ANC off' conditions. The neural centers used in the present work, along with the name of the toolbox/function used for the simulation, are shown in figure 3.1.

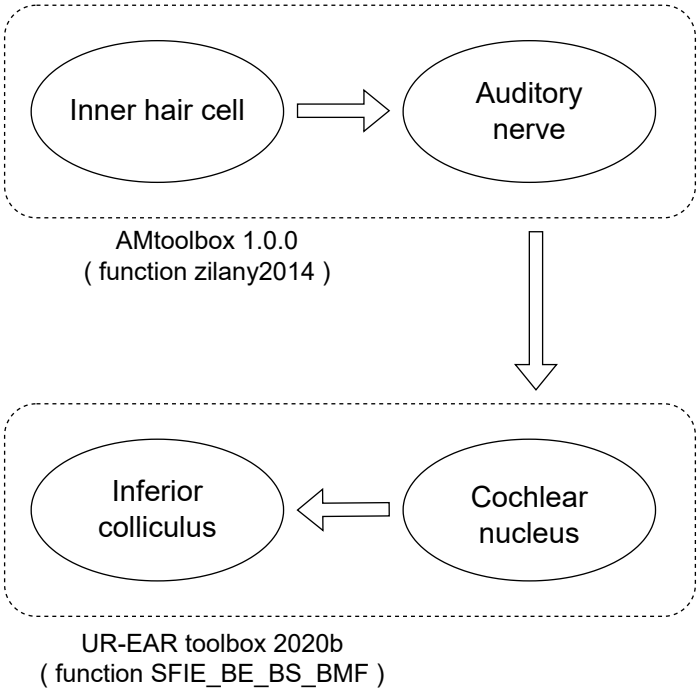



Figure 3.1: Simulated neural centers and inter-connections

### 3.2.1 Simulation procedure and parameters

For simulating neurons of auditory nerve, the AMtoolbox [1] (version 1.0.0 May 2021) was used. For simulating neurons of cochlear nucleus and inferior colliculus, the UR-EAR [2] package (version 2020b) was used. The schematic of the toolboxes and their relation in terms of input-output connections are shown in figure 3.1. Custom wrapper scripts were developed to read the headphone-recorded signals from hard disk, provide these signals to the models, and save their responses as csv files to disk.

The functions provided by the original authors included appropriate simulations of the outer auditory periphery such as the ear canal, middle ear, basilar membrane and so on. The models were essentially given as 'sound in, spikes out' i.e. they expect a time-domain waveform and produce a spike timing vector and/or the instantaneous firing pattern of the neuron for the entire stimulus period. Two examples of neuron spike trains from one of our headphone's recorded signals are shown in figure 3.2. Some models actually produce a longer output in order to enable studying adaptation and memory effects, which were ignored for the purpose of the present study. The models were used 'as-is', i.e. there was no attempt to improve or extend the models. Moreover, unlike computational neuroscience, the goal of this chapter is to use auditory models as a tool to understand the headphones, rather than to understand the brain itself.

[1] available from https://www.amtoolbox.org/

[2] available from https://www.urmc.rochester.edu/labs/carney/publications-code/auditory-models.aspx

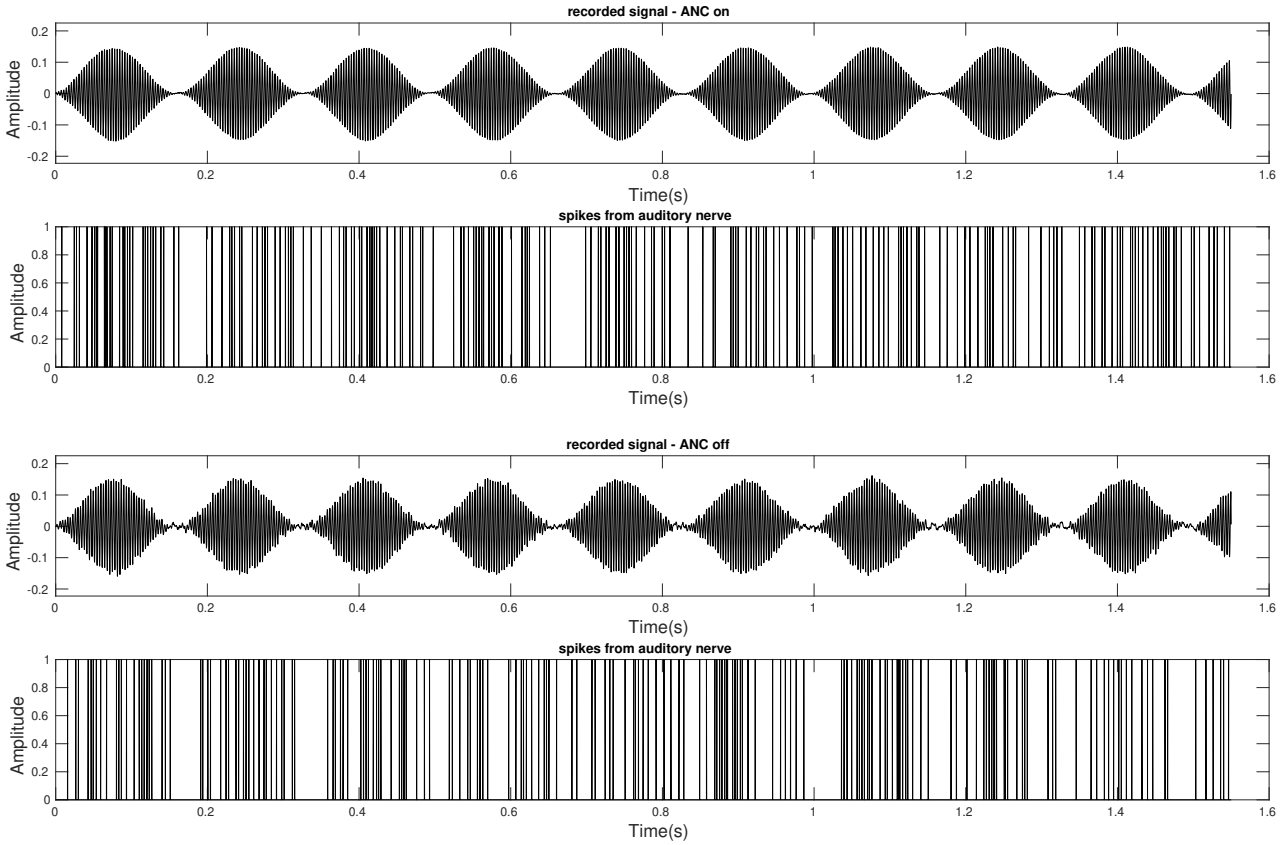



Figure 3.2: Examples of simulated neurons' spike timing vector in response to recorded AM signals

For auditory nerve fibers, the CF (characteristic frequency) was set to the carrier frequency of the input signal. That is, for each of the 10 carrier frequencies, a different neuron was simulated. Similarly, for inferior colliculus neurons, the BMF (best modulation frequency) parameter was set to the modulation frequency used in the input signal. Other neurons (with different CFs) would also show some response to the same signal; in fact this is the approach used to construct the so-called 'neurogram'. However, the focus of this work is characterizing the ANC headphones in terms of the best possible neuron to carry the corresponding signal, hence, neurogram or similar methods are outside the scope of the current work.

The simulated neurons are stochastic models with randomness included inside. To arrive at statistically reliable results, for all measures, 10 repetitions were computed for each available signal and the values were averaged. There were a total of 40000 values for each metric: 10 carrier frequencies, 50 modulation frequencies, 10 repetitions, 2 ANC conditions, 4 headphones; these were analyzed using the statistical procedure described in section 3.2.3.

### 3.2.2 Measures of AM representation

The outputs obtained from the simulated neurons were analyzed in terms of three different metrics, the mean firing rate, the coefficient of variation of inter-spike interval, and the vector strength. These values were computed from the neuron response for each headphone recorded signal. The resulting averaged scores were plotted across the modulation frequency axis to obtain the MTFs. As mentioned earlier, MTFs are used widely in computational neuroscience as a qualitative tool to uncover trends or patterns across conditions, e.g. to find if there were any changes due to some external condition on the neural representation.

**Mean rate** is defined as the mean firing rate of the neuron over the entire duration of the stimulus presentation. It represents the neuron's overall firing intensity to a given stimulus, and thus any change in this quantity represents a large change in the coding strategy used by a neuron while responding to the given stimuli. Mean rate is the metric to assess the representation of the ANC generated signal in terms of the 'rate coding' principle. It can be thought as an analogue to the loudness function used in the previous chapter, except here it is being studied as the response of a single simulated neuron whereas a sensation like loudness is the result of millions of neurons. Mean firing rate is frequently used in studies involving actual neurons or simulated ones to quantify the encoding of amplitude modulation (Guérin et al., 2006). When plotted against modulation frequencies, mean rate produces the so called 'rate modulation transfer function' (rMTF) and is one of the methods usually employed for studying a neurological system's (e.g. single neuron or a population) response to amplitude modulated stimuli (Joris et al., 2004). Studies typically use these rMTF plots as an aggregate visualization tool to qualitatively understand processes of neuron behavior, e.g. a rMTF may be used to assess if there was any change of representation across two external conditions, which may be seen as a shift from bandpass to lowpass shape of the rMTF (Nelson & Carney, 2004). An increase in the mean firing rate of a neuron is interpreted as a change in the overall activity of the neuron; if other factors such as learning etc. are accounted for, this implies a change in the stimulus properties being encoded by the neuron. Mean rate when computed against a long duration is a less-sensitive metric because small fluctuations in the instantaneous firing rate gets swamped out due to the averaging process.

**Coefficient of variation (Cv)** of the inter-spike interval is a measure of spike variability that is defined

as the ratio of the standard deviation and the mean of the inter-spike interval distribution, i.e.,

$$Cv = \frac{\sqrt{\frac{1}{N-2}\sum_{i=1}^{N-1}(\Delta t_i - \overline{\Delta t})^2}}{\overline{\Delta t}} \tag{3.1}$$

where $\overline{\Delta t}$ is the mean inter-spike interval, N is the total number of spikes in the response and $\Delta t_i$ represents individual inter-spike intervals. This is a measure of irregularity, and is often used to characterize units or populations of neurons, real or simulated (Trappenberg, 2009, p. 108). Higher values indicate more disorderly firing behavior by the neurons, and as such has been utilized to understand the 'regularity' of firing patterns by neurons (Rees et al., 1997). This measure has been often used in studies of amplitude modulation to classify neurons into groups such as 'transient choppers' or 'sustained choppers' in the cochlear nucleus (Rhode et al., 2010), and it has also been used in studies involving the auditory cortex (Wang, 2007; Imaizumi et al., 2010). This measure has also been utilized to study things like learning and plasticity in neural models (Song & Abbott, 2000), highlighting the versatile nature of the method. Coefficient of variation is not defined directly in terms of aspects of the input signal (unlike vector strength), hence it can be computed for any spike train without regard to the nature of the input stimulus that elicited that spike train. An increase in coefficient of variation indicates that the neuron has become more erratic in terms of the timing of the spikes, i.e. the variability of the gaps between spikes have increased.

**Vector strength (VS)** is a measure to represent the degree of phase-locking a.k.a synchronization to the stimulus envelope by auditory neurons. It is defined as the length of the average phase of the spikes when projected onto a unit circle, produced by the neuron in response to a given envelope. This average spike has a length of unity if all the neurons fire exactly on the same phase, and zero if the spikes are totally random with respect to the envelope. Thus, it is defined as

$$\text{VS} = \frac{\sqrt{(\sum_{i=1}^{n}\cos\theta_i)^2 + (\sum_{i=1}^{n}\sin\theta_i)^2}}{n} \tag{3.2}$$

where $n$ is the total number of spikes elicited in the neuron under study and $\theta_i$ is the phase angle of the $i$th spike. Vector strength has been utilized by researchers studying actual animal data as well as computational models, and is widely used to study amplitude modulation representation in neurons from the auditory nerve (Sayles & Heinz, 2017), cochlear nucleus (Sayles et al., 2013), inferior colliculus (Krishna & Semple,

2000) and the cortex (Yin et al., 2011). As vector strength is defined in terms of the stimulus envelope, any change in this measure is interpreted due to actual differences in the represented envelope and not computational or biological irregularities or internal noise issues. An increase in vector strength indicates that the neuron is able to phase-lock more accurately to the envelope in that condition.

### 3.2.3 Statistical analysis

The measured values of mean rate, coefficient of variation and vector strength were also analyzed using a linear model (similar to chapter 2), to find if there were any statistically significant differences between headphones and ANC conditions.

## 3.3 Results and discussions

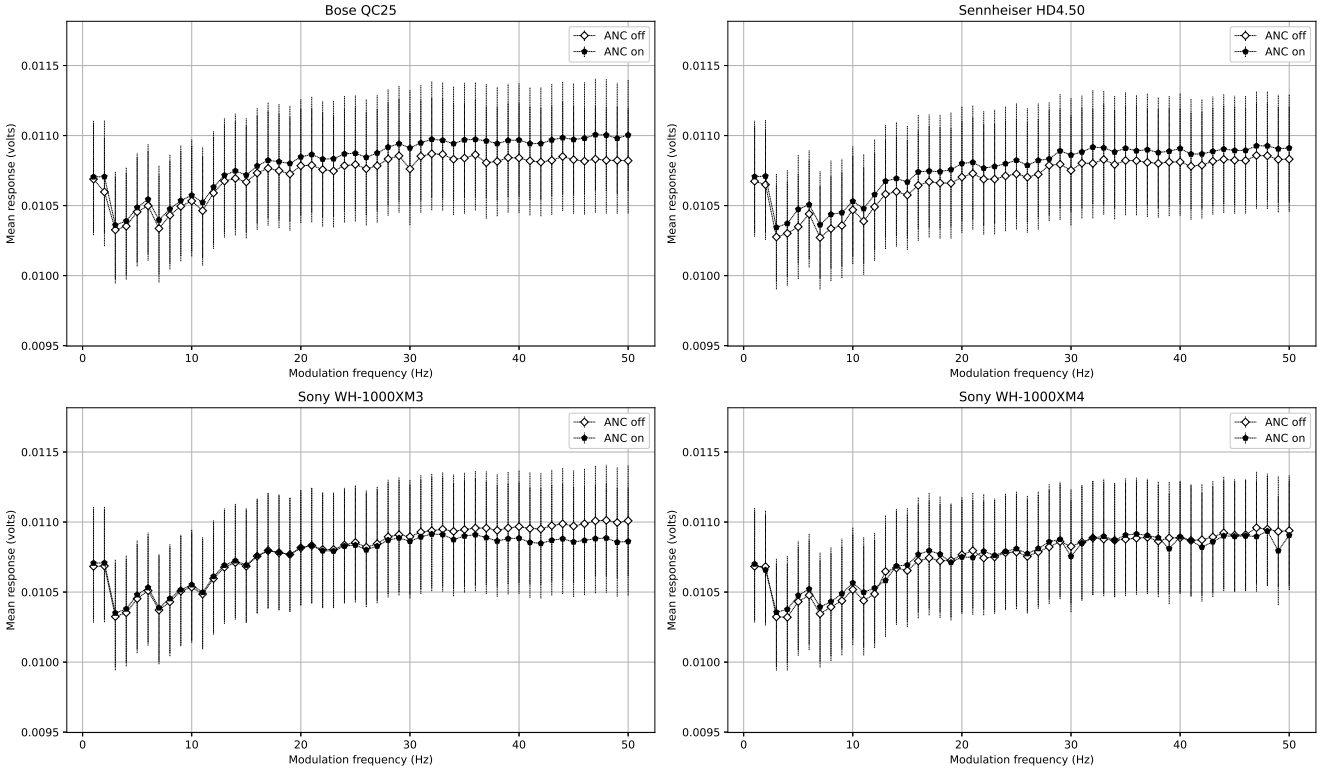



Figure 3.3: rMTF of simulated inner hair cell

The plots of the various MTFs (for all headphones) are shown in figures 3.3 to 3.12. The results of the statistical comparison with analysis of variance (ANOVA) are given in table 3.2. Since all the p-values are below threshold of 0.001, the most interesting information in this table is the effect size, which is a quantitative summary of the amount of difference due to a particular factor (Ellis, 2010). For instance, we can see

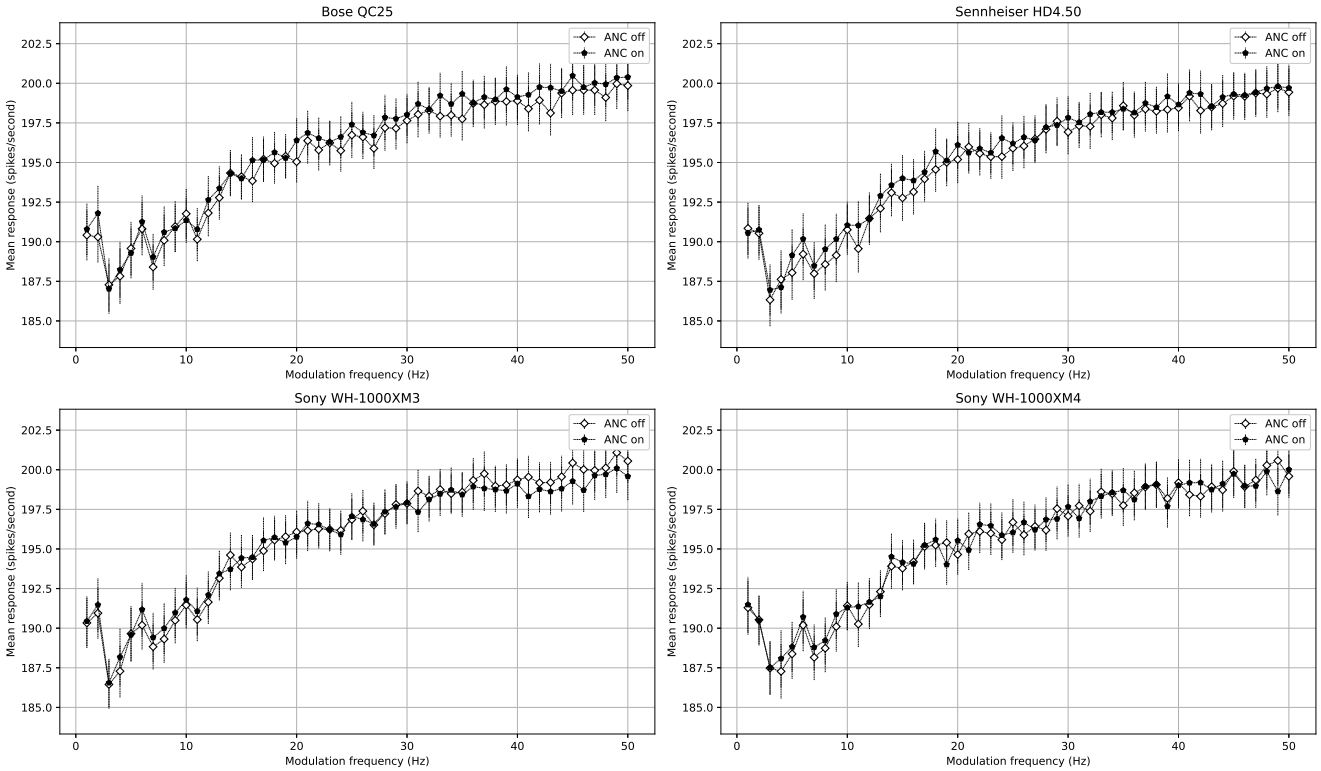



Figure 3.4: rMTF of simulated auditory nerve

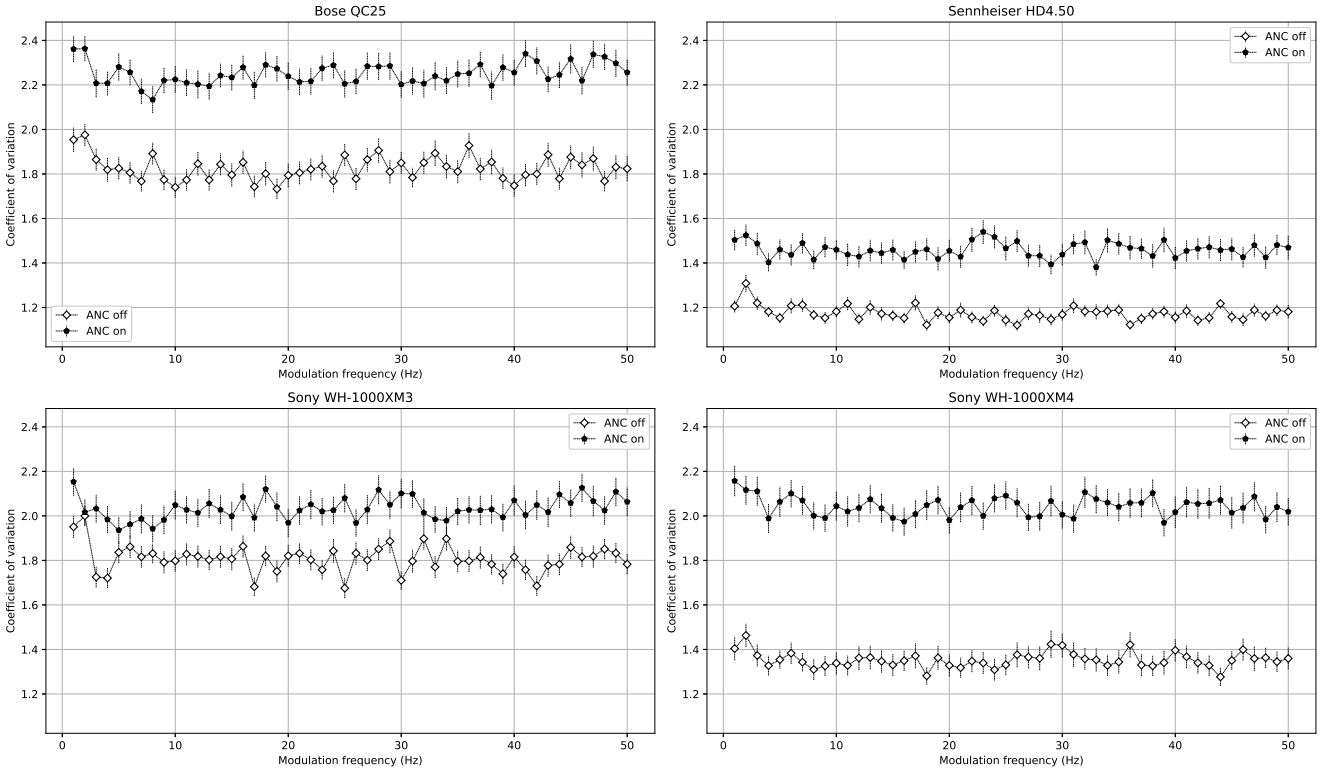



Figure 3.5: MTF of simulated auditory nerve with coefficient of variation

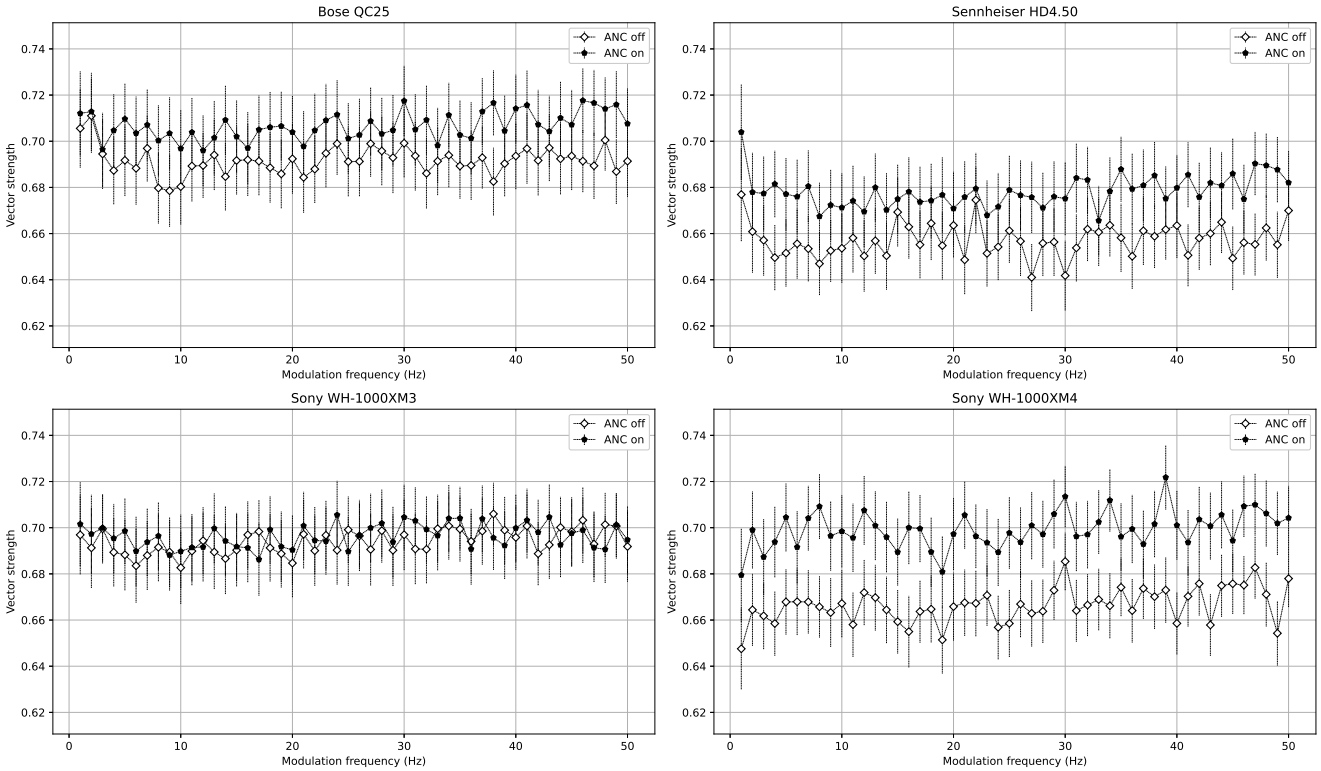



Figure 3.6: Sync-MTF of simulated auditory nerve with vector strength

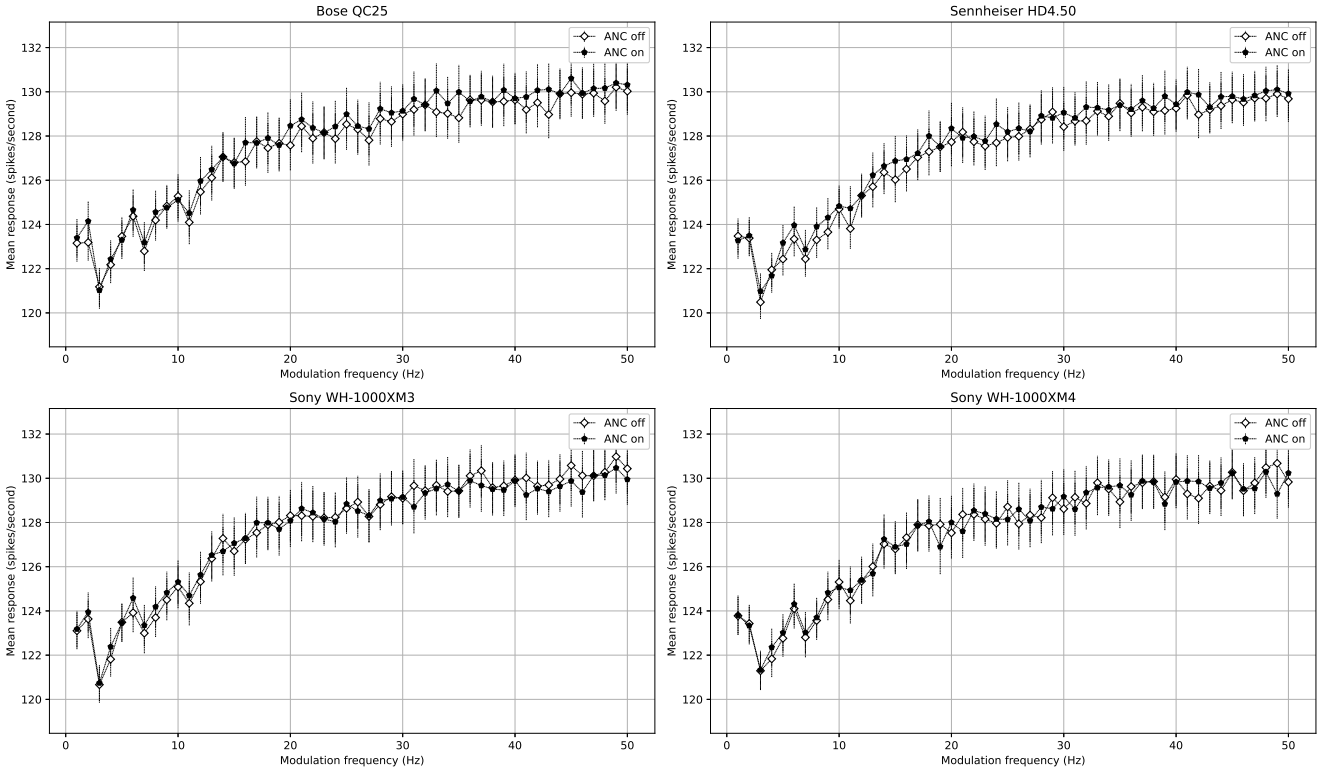



Figure 3.7: rMTF of simulated cochlear nucleus

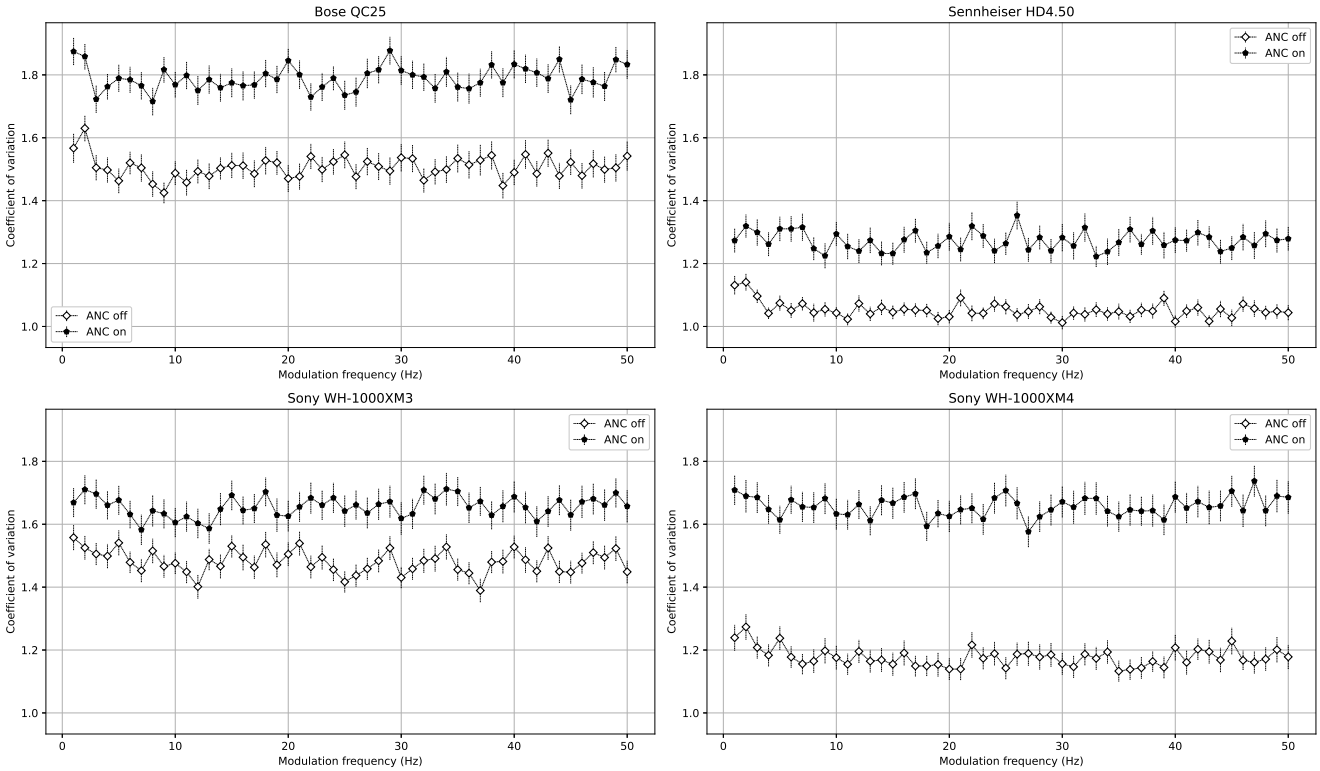



Figure 3.8: MTF of simulated cochlear nucleus with coefficient of variation

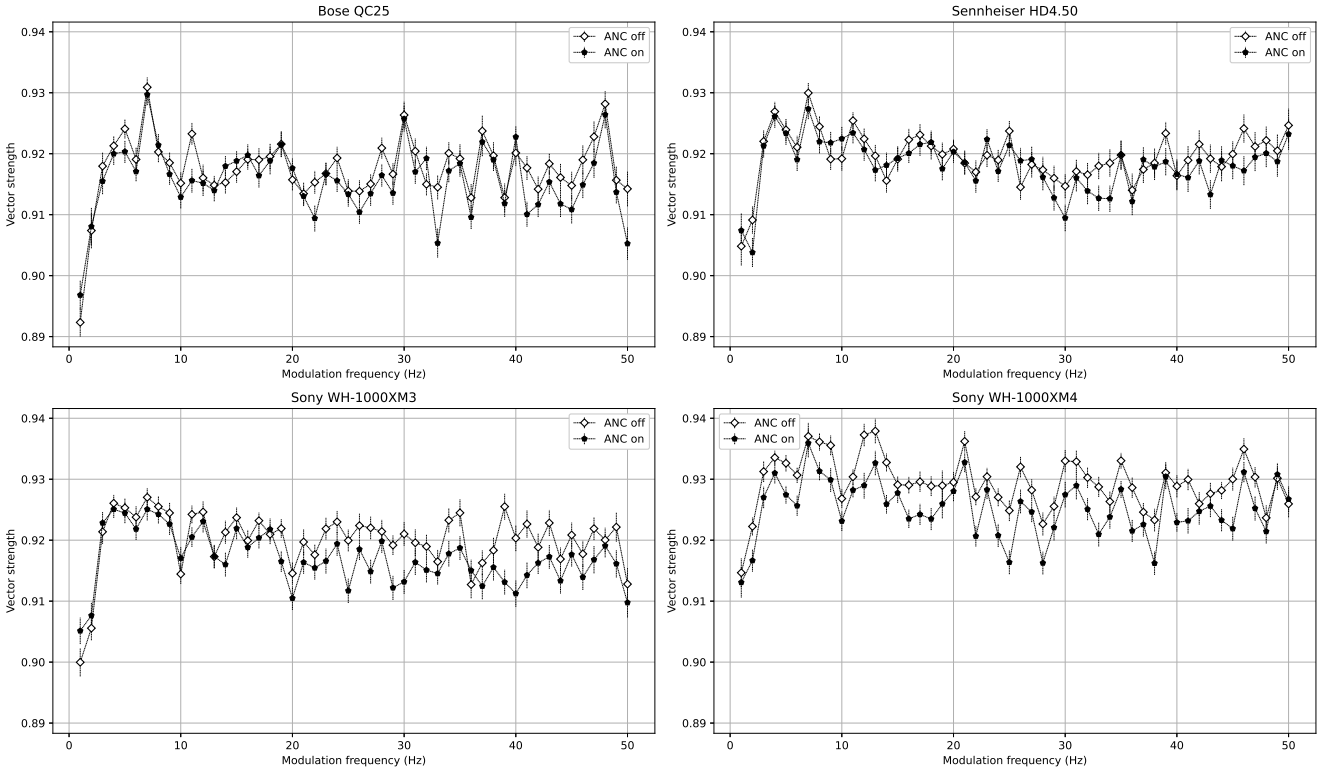



Figure 3.9: Sync-MTF of simulated cochlear nucleus with vector strength

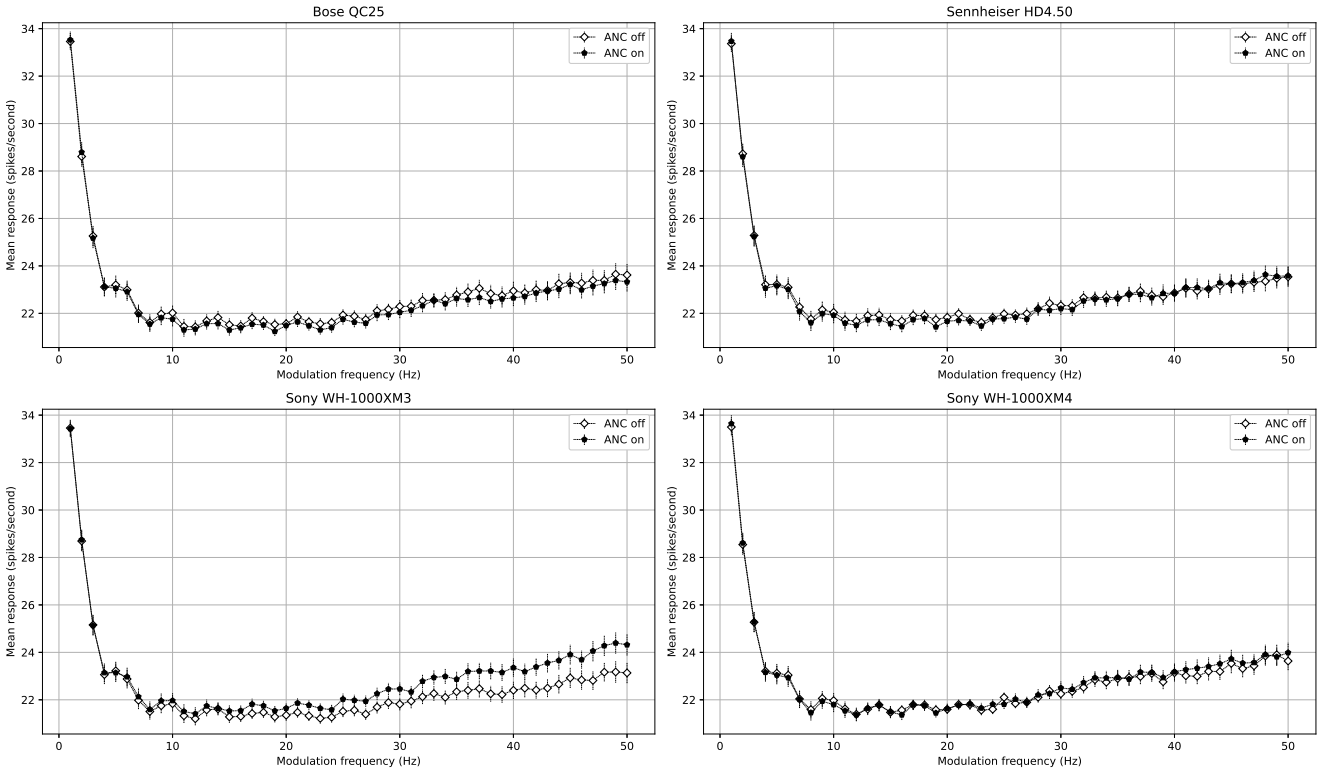



Figure 3.10: rMTF of simulated inferior colliculus

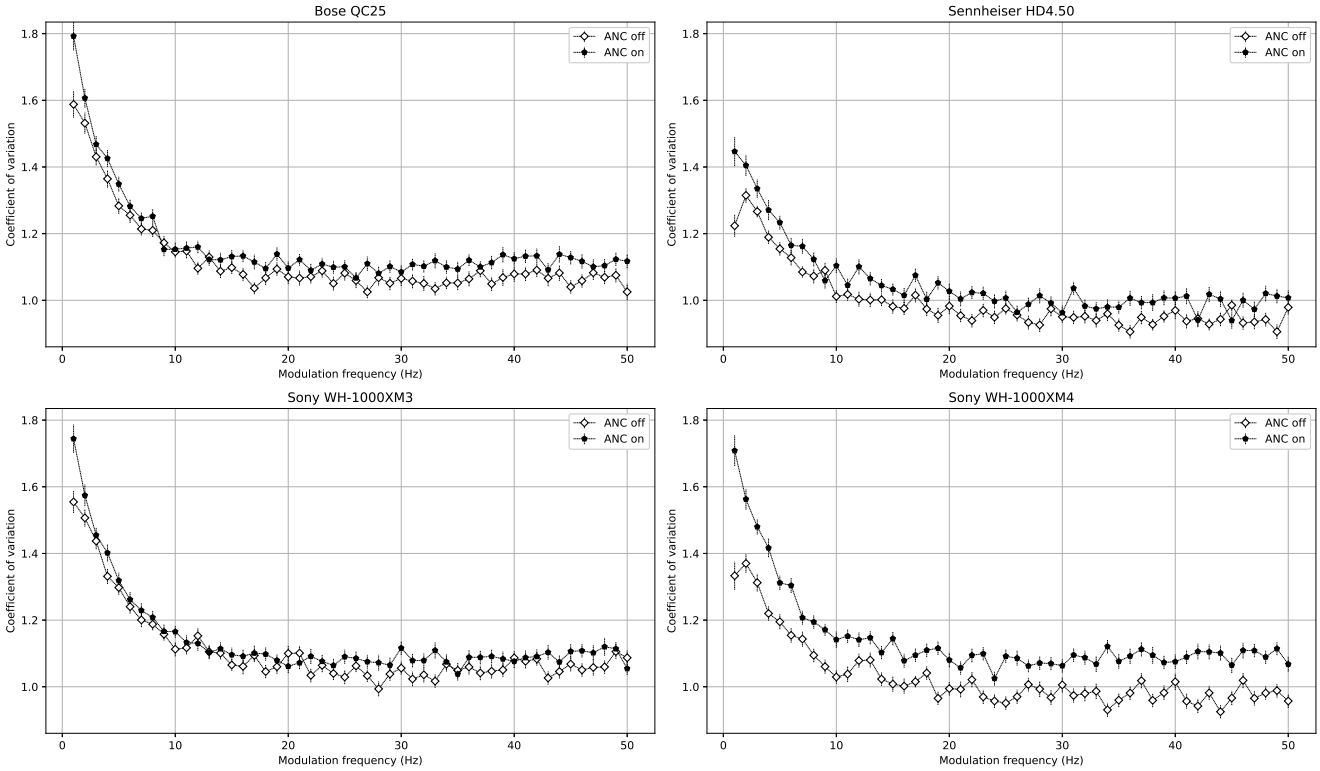



Figure 3.11: MTF of simulated inferior colliculus with coefficient of variation

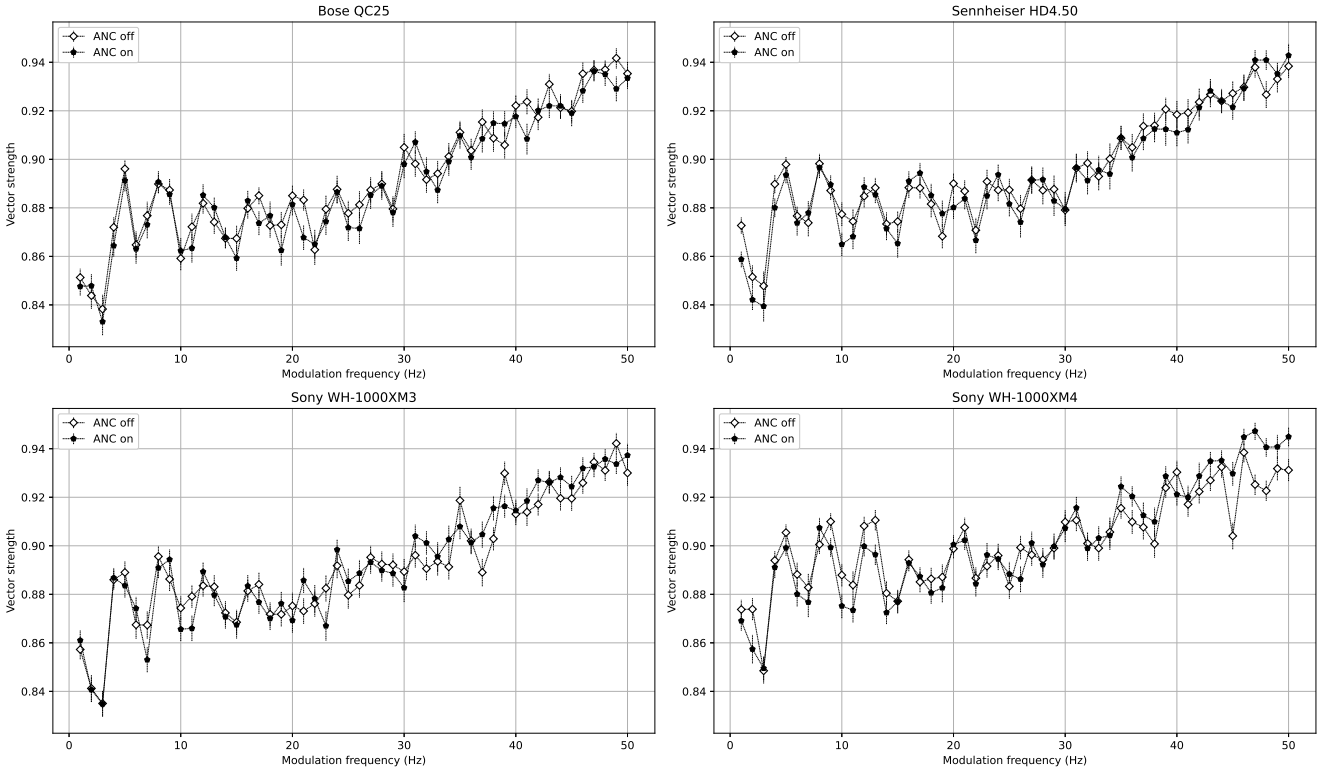



Figure 3.12: Sync-MTF of simulated inferior colliculus with vector strength

that among the significant differences, the coefficient of variation metric has the highest difference (0.2100) across headphone type factor; this means that the different headphone models caused the maximum difference if we choose to focus on the variability in the inter-spike intervals elicited from the simulated neurons. The mean and standard deviation values for each metric across all headphones and ANC conditions are shown in table 3.1.

The inner hair cell is not showing any significant difference across either the headphone conditions or the ANC conditions (figure 3.3). This could be due to a functional incompleteness of the model in its present form; the authors explicitly state that "…it was assumed that there is no adaptation in the voltage responses of the IHC, but recent studies suggest that there is indeed some adaptation at this level" (Zilany et al., 2009). Nevertheless, this result could be interesting when taken together with the rest of the results. Given that the IHC model is the first 'process' involved in the auditory transduction in the cochlear, and yet the ANC effect does not show up there, but shows up in later stages of the cascade.

The auditory nerve is not showing any difference in terms of mean rate across the ANC conditions (figure 3.4). However there is a weak effect across the different headphones; this can be explained by the fact that the different headphones do not work the same as we saw in the various enveolpe distortion metrics shown in chapter 2. However, the auditory nerve vector strength (figure 3.6) is showing a clear difference due to

ANC (higher values for 'ANC on'), which is significant across the headphones and also the interaction between them. This means that the ANC is enabling the simulated auditory nerve to fire more accurately at regular phases of the amplitude envelope of the signal, thus showing that effect of ANC in the physical signal also translates to a better ability of neurons in representing the amplitude envelope. Furthermore, the coefficient of variation shows an even larger effect across the ANC conditions and the headphones (figure 3.5). This can be interpreted in conjunction with the vector strength result; in the 'ANC on' conditions, the neurons' become more synchronized to the envelope. As a consequence, their inter-spike intervals become more deviated from the mean, hence causing a higher variability i.e. a higher value of Cv. Last but not the least, the Cv metric turns out to be more sensitive i.e. the effect sizes are higher. Overall, this presents clear evidence that ANC headphones (and perhaps other smart audio devices like hearables) can be evaluated in terms of simulated auditory nerve activity.

The cochlear nucleus, similar to the auditory nerve, does not show a difference in the rMTFs (figure 3.7), probably due to the swamping out of the effect over the relatively large stimulus duration. And again, the Cv shows large effect sizes and vector strength shows smaller effect sizes (across both headphone factor and ANC conditions factor, figures 3.9 and 3.8), indicating that 'ANC effect' experienced by the auditory nerve is passed on to the cochlear nucleus, and it is doing its own neural computations and arriving at a similar but slightly different pattern of results. Interestingly, the absolute values of the vector strength have increased greatly from the auditory nerve – from about 0.7 to 0.9; this represents a change in the auditory system's coding strategy priorities and information sorting principles and has been explored in the neuroscience literature extensively (e.g. (Rhode & Greenberg, 1994; Rhode et al., 2010)). Taken together, the CN results suggest that the second stage of the auditory brainstem can serve as a confirmation of the ANC effect from headphones, and potentially the differences in absolute or relative changes can be correlated with perceptual outcomes to arrive at a more robust understanding of the effect of ANC on listeners.

Finally, the inferior colliculus as known as an obligatory synapse, is also showing some differences due to ANC but the effect sizes are very low. For all the metrics, the unusual non-linear shape of the MTF at the lower modulation frequencies can most likely be explained by the neuroscientific findings in terms of the properties of stellate or bushy cells (Guérin et al., 2006), and the model is tuned to replicate such phenomena (Nelson & Carney, 2004). In terms of mean rate (figure 3.10), surprisingly there are small but significant differences indicating that it is in principle possible for even long stimuli to elicit such 'ANC effect' or 'headphone effect' differences. The coefficient of variation absolute values are similar to the au-

ditory nerve, but the effect due to ANC is much smaller (figure 3.11). The vector strength metric (figure 3.12) shows almost no differences across ANC or headphone (compared to the previous locations), but the difference across headphone types are nevertheless significant. Overall, inferior colliculus seems to have partially or fully discarded the difference caused by the ANC in terms of all the metrics. It is possible that this is a reflection of the priorities of the AM coding strategies for the brain, or this is an artifact arising of the specifics of how the model is being utilized in the present work. At any rate, in terms of the perceptual evaluation of smart audio devices, models of the inferior colliculus can serve as a 'limit' against which more optimistic findings can be pitted and nullified to see whether such effects would survive the journey to auditory cortex.

Overall, ANC seems to be somewhat effective in changing the representation of AM at the various neural locations. The two different synchronization metrics (Cv and VS) suggest that this difference can be quantified as a more 'disciplined' firing of the neurons and better tracking of the amplitude envelope in the input signal. Some work suggests that the brain has noise invariant mechanisms to represent acoustic modulations faithfully regardless of noise levels; that is, they dynamically adjust their firing behavior to account for varying noise levels (Rabinowitz et al., 2013). Such processes may or may not be already emergent in the phenomenological models being deployed here, it is not clear whether such adaptation effects would enhance, alter or destroy the 'ANC effect'.

Table 3.1: Mean and standard deviation of simulated neurons' response across headphones

| | ANC | QC25 | | HD 4.50 | | WH-1000XM3 | | WH-1000XM4 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Stdev | Mean | Stdev | Mean | Stdev | Mean | Stdev |
| IHC potential (volts) * 100 | On | 1.0815 | 0.4031 | 1.07582 | 0.3958 | 1.07626 | 0.3986 | 1.07468 | 0.3992 |
| | Off | 1.0722 | 0.3944 | 1.06749 | 0.3861 | 1.07949 | 0.4003 | 1.07406 | 0.3992 |
| AN mean rate (spikes/s) | On | 196.061 | 14.76 | 195.489 | 14.7 | 195.737 | 15.01 | 195.509 | 14.78 |
| | Off | 195.537 | 15.15 | 195.07 | 15.5 | 195.857 | 14.24 | 195.404 | 14.77 |
| AN coefficient of variation | On | 2.25078 | 0.5966 | 1.45914 | 0.4883 | 2.03405 | 0.6335 | 2.04381 | 0.65 |
| | Off | 1.8254 | 0.5133 | 1.17501 | 0.2789 | 1.80803 | 0.4849 | 1.35384 | 0.4954 |
| AN vector strength | On | 0.706618 | 0.1526 | 0.678064 | 0.1507 | 0.696429 | 0.1473 | 0.699361 | 0.1468 |
| | Off | 0.691644 | 0.1481 | 0.657123 | 0.1438 | 0.694039 | 0.1474 | 0.666552 | 0.1386 |
| CN mean rate (spikes/s) | On | 127.846 | 11.49 | 127.513 | 11.05 | 127.694 | 11.45 | 127.588 | 11.54 |
| | Off | 127.514 | 10.98 | 127.207 | 10.44 | 127.738 | 11.21 | 127.559 | 11.39 |
| CN coefficient of variation | On | 1.78915 | 0.4428 | 1.27289 | 0.3962 | 1.65618 | 0.4824 | 1.65755 | 0.49 |
| | Off | 1.50636 | 0.4163 | 1.05282 | 0.2322 | 1.48156 | 0.3722 | 1.17706 | 0.3665 |
| CN vector strength | On | 0.915666 | 0.02089 | 0.918243 | 0.02118 | 0.916933 | 0.01974 | 0.925256 | 0.01905 |
| | Off | 0.917604 | 0.02094 | 0.919651 | 0.02108 | 0.920166 | 0.01976 | 0.929571 | 0.01752 |
| IC mean rate (spikes/s) | On | 22.6099 | 3.687 | 22.768 | 3.771 | 22.9759 | 3.791 | 22.8965 | 3.68 |
| | Off | 22.7776 | 3.888 | 22.8507 | 3.811 | 22.4627 | 3.821 | 22.8447 | 3.661 |
| IC coefficient of variation | On | 1.16534 | 0.2534 | 1.05256 | 0.2534 | 1.14214 | 0.2598 | 1.14468 | 0.2603 |
| | Off | 1.11987 | 0.2522 | 0.996728 | 0.2286 | 1.1104 | 0.2452 | 1.02929 | 0.2457 |
| IC vector strength | On | 0.889787 | 0.05752 | 0.893835 | 0.05596 | 0.893085 | 0.05475 | 0.902665 | 0.05247 |
| | Off | 0.892491 | 0.05539 | 0.895883 | 0.05436 | 0.892348 | 0.05479 | 0.902053 | 0.04962 |

Table 3.2: Statistical comparison (two-way ANOVAs) of simulated neurons' response metrics

| Location and measure | factor | F value | p-value | Effect size ($\eta_p^2$) |
|---|---|---|---|---|
| IHC potential | Headphone Type | F(3,39992) = 0.489 | 0.689 | 0.0000 |
| | ANC condition | F(1,39992) = 0.889 | 0.346 | 0.0000 |
| | Interaction | F(3,39992) = 0.582 | 0.627 | 0.0000 |
| AN mean rate | Headphone Type | F(3,39992) = 3.029 | **0.028** | **0.0002** |
| | ANC condition | F(1,39992) = 2.436 | 0.119 | 0.0001 |
| | Interaction | F(3,39992) = 0.981 | 0.401 | 0.0001 |
| AN coefficient of variation | Headphone Type | F(3,39992) = 3597 | **<0.001** | **0.2100** |
| | ANC condition | F(1,39992) = 5898 | **<0.001** | **0.1300** |
| | Interaction | F(3,39992) = 381.7 | **<0.001** | **0.0300** |
| AN vector strength | Headphone Type | F(3,39992) = 93.08 | **<0.001** | **0.0069** |
| | ANC condition | F(1,39992) = 145.67 | **<0.001** | **0.0036** |
| | Interaction | F(3,39992) = 18.48 | **<0.001** | **0.0014** |
| CN mean rate | Headphone Type | F(3,39992) = 2.044 | 0.105 | 0.0001 |
| | ANC condition | F(1,39992) = 1.931 | 0.165 | 0.0000 |
| | Interaction | F(3,39992) = 0.728 | 0.535 | 0.0001 |
| CN coefficient of variation | Headphone Type | F(3,39992) = 2750 | **<0.001** | **0.1700** |
| | ANC condition | F(1,39992) = 5056 | **<0.001** | **0.1100** |
| | Interaction | F(3,39992) = 136.3 | **<0.001** | **0.0200** |
| CN vector strength | Headphone Type | F(3,39992) = 568.9 | **<0.001** | **0.0400** |
| | ANC condition | F(1,39992) = 183.8 | **<0.001** | **0.0046** |
| | Interaction | F(3,39992) = 10.61 | **<0.001** | **0.0008** |
| IC mean rate | Headphone Type | F(3,39992) = 4.706 | **0.003** | **0.0003** |
| | ANC condition | F(1,39992) = 4.365 | **0.037** | **0.0001** |
| | Interaction | F(3,39992) = 16.242 | **<0.001** | **0.0012** |
| IC coefficient of variation | Headphone Type | F(3,39992) = 440.4 | **<0.001** | **0.0300** |
| | ANC condition | F(1,39992) = 617.2 | **<0.001** | **0.0200** |
| | Interaction | F(3,39992) = 54.36 | **<0.001** | **0.0040** |
| IC vector strength | Headphone Type | F(3,39992) = 83.08 | **<0.001** | **0.0062** |
| | ANC condition | F(1,39992) = 2.439 | 0.118 | 0.0001 |
| | Interaction | F(3,39992) = 2.676 | **0.045** | **0.0002** |

## 3.4 Chapter summary

In this chapter, we characterized ANC headphones in terms of the responses of auditory centers of the brain. We provided modulated sinusoids played with these headphones as inputs to these models, and computed the simulated neurons' responses in terms of three standard metrics used in AM studies. Computational models of three neural centers of the brainstem, the auditory nerve fibers, cochlear nucleus and inferior colliculus were used for this purpose. We obtained the mean rate, coefficient of variation and vector strength of these neural centers across the 'ANC on' and 'ANC off' conditions for the various headphones. We found that the simulated neurons exhibited differences in amplitude modulation representation when

the ANC was enabled, and that this effect was not of the same magnitude across the different headphones under study. Moreover, we found that this effect was not seen to the equal extent from all three metrics, highlighting the need for choosing the proper metric. Even if an effect is present, studying it with improper metrics would result in false negative or false positive conclusions depending on the nature of the metric. We also found that the different neural centers exhibited some surprising patterns of the 'ANC effect', which suggests that the ANC phenomena deserves further study in terms of in-depth modulation representation investigations.

A word of caution is warranted regarding the use of computational models of neurons for practical purposes such as studying headphones. Despite the large amount of literature studying representation of modulation frequencies in auditory brainstem neurons, there are still debates around the basic idea that neurons may encode modulation parameters at all. For instance, Viemeister (2014) stated "…physiologically measured responses to AM have provided little insight into the processes underlying the psychophysics, specifically whether there is tuning for AM rate as postulated by the modulation filterbank models". It is also unclear to what extent the representation of amplitude modulation is actually relevant for speech comprehension, because electroencephalography (brain scanning using electrodes on the scalp) studies have found contradictory evidence in terms of the importance of the amplitude envelope and speech perception. Some researchers conclude that '…envelope information alone can predict neither the intelligibility of speech nor the goodness of phase-locking to the speech signal' (Obleser et al., 2012)). Thus, using the concept of envelope representation can only account for a limited amount of the 'ANC effect' and future research must look for more expanded perspectives in terms of auditory model based assessment of ANC headphones.

### 3.4.1 Future work

The characterization of ANC headphones or other 'hearables' devices can be carried out by extending the methods outlined here in the following ways:

- Modulated sinusoids can be constructed in shallow modulation depths too, unlike the 100% modulation (i.e. 0 dB depth) signals used in the present work. These shallow modulation depth signals have been studied by neuroscience researchers in search for the exact representation code used by neurons (e.g. (Nelson & Carney, 2006)). By studying the differences (if any) of neurons' responses to shallow modulation AM signals across ANC conditions might reveal interesting patterns also.

- Nelson and Carney (2006) stated "to date, the relative roles of timing (i.e. synchronization to the envelope) and average rate information as neural substrates for AM perception at low modulation depths (near behavioral thresholds) remain unclear." Their objective was to study the brain, but the point stands for the present work: neither rate nor simple synchronization can tell the whole story about a neuron's ability to encode features of a given stimulus. Other researchers have proposed a lot of other complicated metrics to track neural response, e.g. the so called 'phase projected vector strength' which has been used to study AM in cochlear nucleus (Sayles et al., 2013). Such metrics can be applied to better characterize the ANC effect on simulated neurons too.

- All the recorded signals were provided at a fixed level of intensity to the auditory nerve models. It has been shown that the synchronization of AN fibers to amplitude modulated signals may vary across intensities (Heil et al., 2011) (as shown by the so called 'rate-level functions'), hence future work would need to re-examine the results across a wide range of intensities.

- Some researchers have found it useful to distinguish between the acoustic ENV and TFS from the basilar membrane vibration encoding the ENV and TFS, and subsequently the neural ENV and TFS, since the cochlear tuning may imply the neural ENV is correlated with the neural TFS (Swaminathan & Heinz, 2012). Since the ANC effect is probably arising from the interaction of the reduced-noise signal and the basilar membrane via 'upward spread of unmasking', future simulation studies could focus on the these three different kinds of ENV/TFS pairs (physical, basilar membrane, neural) and investigate the differences caused by ANC in each one of them.

- The phenomenological models of inner hair cell and auditory nerve synapse used here is not the only one that have been developed by researchers in computational neuroscience. Other models have been proposed which take different stances regarding the level of detail and the inner structure of the models, for instance the models by Meddis and colleagues (Sumner et al., 2002).

- Neurons in the brainstem have been classified on the basis of their spontaneous rate (SR) of firing, i.e the activity that is observed in the absence of any stimulus, into 'high', 'medium' and 'low' categories (e.g. (Rhode & Greenberg, 1994; Zilany et al., 2009)). In this work all neurons were at the default values of SR, which was 'medium'. Simulated neurons with other SR types could shed new insights on the ANC effect.

# 4

# Intelligibility and listening-effort of shallow-envelope chimera through ANC

## 4.1 Introduction

This chapter is focused on characterizing the ANC headphones with human listeners' speech perception, where audio instructions are presented through the ANC headphones under both 'ANC on' and 'ANC off' conditions, and the listeners carry out a task according to the instruction. By manipulating the modulation depth of the envelope (ENV) we explore whether the ANC causes any differences in the listeners' perception in terms of intelligibility, response time and listening effort.

Intelligibility is commonly defined as the proportion of correctly identified words or sentences in a closed-response speech perception task. By statistically comparing the intelligibility scores across experimental conditions (e.g. settings of a hearing aid), researchers make inferences regarding the effects of those specific manipulations and the potential interactions between experimental conditions, subjects' hearing impairment status and so on. Intelligibility, however, is extremely robust to degradations in the speech signal; it is not until the signal breaks down significantly that the intelligibility begins to drop. Therefore, to better understand the cognitive and emotional difficulties, a measure called listening effort has gained popularity in the last few decades. Although it is a debated and somewhat controversial psychological construct,

a common definition of listening effort is "the allocation of cognitive resources to overcome obstacles or challenges to achieving listening-oriented goals" (Francis & Love, 2020). There are multiple direct and indirect ways of measuring listening effort, including response time, self-report rating scales, skin conductance, pupil dilation and more. These measures do not answer the same questions, i.e. they do not completely overlap with the construct called 'listening effort' but are known to have a high degree of association with it, hence are interpreted as acceptable proxies for it (Alhanbali et al., 2019). However, rating scale based questionnaires are often considered superior because asking listeners explicitly to give ratings on effort for a given listening task is thought to be similar to the conscious introspective experience underlying communication situations; as such, these ratings would be applicable to clinical applications and may even be more sensitive than other listening effort measures (Francis & Love, 2020). Specific types of impairments affect speech perception differently, e.g. reverberation and noise (Rennies et al., 2014), so it is common for studies to combine these measures to implement a specific kind of listening effort measure, e.g. response time under dual task along with skin conductance.

Listeners may use their headphones in noisy areas to listen to speech, such as a phone call or a podcast. In these scenarios the ANC function of the headphone, when enabled, is supposed to offer additional advantage or benefit. As we saw in chapters 2 and 3, the modulation frequencies are enhanced (in a physical sense) and simulated neurons are able to better encode the amplitude envelope in the 'ANC on' condition compared to the 'ANC off' condition. Now, we turn our focus on the other important aspect of amplitude envelope, namely modulation depth. The relevance of amplitude envelope in speech perception, and especially the interaction between noise envelope and speech envelope has been the subject of detailed research in the last two decades. Some of the relevant ideas are summarized below.

### 4.1.1 Amplitude envelope in human speech perception

The role of the amplitude envelope (a.k.a temporal envelope) is currently one of the most heavily studied aspects of the signal in human speech perception (Delgutte et al., 1998; Shamma & Lorenzi, 2013). Any speech signal can be split into the amplitude envelope and temporal fine structure (TFS). It has been shown in multiple studies under different conditions that humans can recognize words with only the ENV, which may be imposed on a noise or tone carrier, e.g. (Shannon et al., 1995). The relative contribution of ENV and TFS can change under different listening conditions (such as heavy masking from noise) or due to hearing

impairment, and the brain might adaptively use these cues in an optimal way to extract the meaning to the best of its ability (Zeng et al., 2005; Shamma & Lorenzi, 2013). The envelope has been hypothesized to carry information on multiple time scales, and includes information on things such as tempo or rhythm, stress, intonation, manner, syllable rate etc. (Rosen, 1992). Psycho-linguists have also worked on the aspects of envelope in relation to speech perception processes, e.g. the amplitude envelope was found to be used by the brain to decode the information of syllable structure and rhythm of speech, which is then used for identification of voicing and manner of articulation of consonants (Jenstad & Souza, 2005), or to distinguish specific consonant-vowel pairs like /ba/ vs. /da/ (Shinn & Blumstein, 1984). The speech envelope can get degraded due to a variety of reasons; for instance, reverberant conditions corrupts the envelope by distorting ('smearing') the shape and also reducing the average modulation depth (Drullman et al., 1994), although there is evidence that neurons in the brain have 'dual-coding' strategies which can sometimes cause better perception of reverberant signals compared to normal (anechoic) signals (Slama & Delgutte, 2015). Füllgrabe et al. (2009) studied speech-on-speech masking in the context of modulations and found that the region around 4 Hz in the amplitude modulation spectrum was critical for speech understanding. Studies with a clinical focus have found psychoacoustic effects of specific speech processing strategies, such as intensity compression on speech intelligibility (Jenstad & Souza, 2005). Hu et al. (2020) studied the performance of automatic speech recognition systems with vocoded speech as input and found a mixed recognition success with envelope-only speech. A common method of quantifying the importance of various frequencies in the speech envelope is to filtering it systematically and studying the perceptual boundaries of the resultant signals. Smith et al. (2002) studied the relative relevance of ENV and TFS, and proposed the chimera concept, which has since been used extensively. Venezia et al. (2016) proposed a concept known as 'auditory bubbles' which is a region in the spectro-temporal modulation pattern, that can be clearly identified as bearing specific function in speech perception such as gender recognition. Swaminathan and Heinz (2012) demonstrated that the envelope of speech is represented in the activity of the auditory nerve, and predicted speech intelligibility with a computational model from the neural ENV and neural TFS; they concluded that neural ENV was the primary cue compared to neural TFS. Bosker and Cooke (2018) studied differences in talking in terms of the amplitude envelope produced by talkers in presence of different noises, and found that people produce speech having deeper modulation depths in the envelopes when speaking in the presence of noise (also known as 'lombard speech').

The concept of studying human speech intelligibility and quality in the presence of additive noise is

also one of the long-standing practical problems in the literature of speech perception. When noise is mixed with speech, interactions between the envelopes of speech and noise reduces the intelligibility and quality of the speech signal, because the noise amplitude envelope is superposed with the speech amplitude envelope, and the resulting 'chaotic' modulations are not conducive to speech perception (Greenberg & Arai, 2001; Dubbelboer & Houtgast, 2008). These and other acoustic to perceptual mappings of speech, including the specific aspects of envelope and speech perception, are reviewed by (Blumstein, 2021; Johnson & Sjerps, 2021). Hilkhuysen et al. (2014) reviewed the literature on noise suppression and how it may relate to speech intelligibility, with detailed consideration of the envelope such as signal to noise ratios of envelope power and a number of intelligibility prediction algorithms utilizing these quantities. Predicting speech intelligibility has also been a practical concern with most audio technology such as telephone networks, and one of the most commonly used methods is to use fidelity of the amplitude envelope. The so called 'Speech Transmission Index' (STI) used a frequency-band wise summation of the envelope distortion to arrive at a representation that could be used to predict speech intelligibility in rooms and enclosed spaces (Steeneken, 1992). However, STI has been shown to lack general prediction ability and several dedicated models (based on different aspects of amplitude envelope) have been proposed to predict speech intelligibility under specific types of distortion, e.g. phase jitter (Elhilali et al., 2003). The auditory model known as the modulation filterbank has been applied to create various predictive indices of speech intelligibility and quality, some of these (such as PEMO-Q (Huber & Kollmeier, 2006), HASPI (Kates & Arehart, 2014a), HASQIv2 (Kates & Arehart, 2021), and GPSMq (Biberger et al., 2018)) are discussed and used in the next chapter. Relaño-Iborra et al. (2019) represents one of the latest examples of application of the modulation filterbank model in predicting speech intelligibility; in this study hearing-impaired listeners' intelligibility are modeled using envelope statistics.

Interestingly, the concept of envelope has come up in the context of understanding ANC devices but never fully explored. Reimes (2019) has proposed an auditory model based on the modulation filterbank architecture (Dau et al., 1997a, 1997b), to objectively predict the listening effort of a user when wearing an ANC headphone, with recordings on the binaural signals of the headphone as input. Bockstael et al. (2011) studied the effect of hearing protector devices on the playback signal, and came close to investigating the change in envelope shape, but did not delve deeper into this aspect. On the other hand, studies focusing on speech intelligibility from practical aspects of ANC, such as privacy preservation in a car cabin (Kim & Park, 2017), do not devote any attention on answering the question of how speech signals' ENV or TFS

might get affected due to the ANC. Hence, the experiment described in this chapter is to bridge the world of ANC with amplitude envelope, using human participants.

## 4.2 Materials and methods

### 4.2.1 Chimera algorithm

The broad goal of the chimera developed in this chapter is to parametrically remove the ENV from the speech signal while keeping its TFS intact, and to do this 'removal' by reducing the modulation depth. Multiplying the ENV by a constant between 0 and 1 changes its modulation depth but this merely changes the signal's power, and any power correction such as rms normalization brings it back to the original. Hence, the method here removes the ENV by adding a small constant (called 'dc bias' in some disciplines) and then multiplying it with the scalar. The algorithm for producing chimera from signal is given in table 4.1 and illustrated as a block diagram in figure 4.1. Some examples of the output signals in time-domain for some values of the modulation depth are shown in figure 4.2.

Table 4.1: Chimera algorithm

| |
|---|
| Inputs: Discrete time representation of speech signal, s<br>Parameters: Modulation depth (m), dc bias (d)<br>Output: Chimera signal |
| **Step 1**: s is split into 17 non-overlapping frequency bands with bandpass filters, with cutoff frequencies between 100 Hz to 9 kHz.<br>Step 2: In each frequency band, the ENV and TFS are extracted with the Hilbert transform. The ENV is given by the absolute value of the Hilbert transform and the TFS is given by the cosine of the phase of the Hilbert transform.<br>Step 3: The ENV is multiplied by m, then d is added to it.<br>Step 4: The modified ENV and the original TFS are combined (multiplied element-wise) to form the intermediate signal (i).<br>Step 5: The i (in each frequency band) are filtered by the same set of bandpass filters as in step 1.<br>Step 6: Results from step 5 are added together to produce another intermediate signal (j).<br>Step 7: Signal j is subjected to squared cosine window function which consists of an 45 millisecond long onset ramp and offset ramps. After this, the resultant signal is padded with 200 milliseconds of silence (zeros) on both ends.<br>Step 8: The result from step 7 is the chimera signal, which is returned. |

### 4.2.2 Stimuli

The Coordinate Response Measure (CRM) dataset introduced by Bolia et al. (2000) was used as the base for constructing the chimera. This dataset consists of English sentences in the form of an instruction that

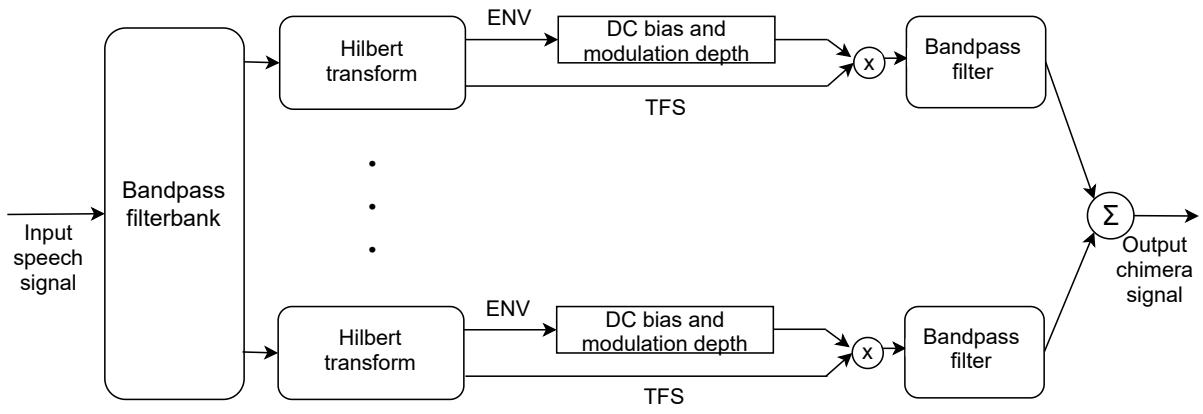



Figure 4.1: Outline of chimera algorithm

asks the listener to choose a color and number combination. An example instruction is "Ready Charlie go to blue seven now", where 'Charlie' is an imaginary call-sign, and 'blue seven' is the target which the listener must choose. In total there are 8 such call-signs, 4 colors (red, green, blue and white) and 8 numbers (1 through 8). This dataset was intended to be used for rapid assessment of speech intelligibility under variety of distortions and other real-world issues by the army or air-force of USA, and is currently widely used in auditory psychophysics research for variety of purposes. Due to the training-free nature of the task and the closed-response design which allows greater control over the experimental design and other factors, this dataset has been used in several studies to study human speech intelligibility (Brungart, 2001; Eddins & Liu, 2012). A publicly available [1] version of this dataset was used which consists of 8 talkers (4 M + 4 F), all of whom uttered all the possible sentence combinations as specified by the dataset.

### 4.2.3 Experiment design

A repeated measures experiment was designed with each participant doing both 'ANC on' and 'ANC off' conditions. To avoid learning or practice effects, participants were allocated in equal number and randomly to have 'ANC on' followed by 'ANC off' conditions, and vice versa. The participants were instructed to take the headphones off during the short break between ANC conditions, and the headphone button was operated by the experimenter discreetly during a break, i.e. without informing the participants about it. The transition between the first and second blocks was signalled by a short message displayed on the screen. In each trial, a chimera signal was played through the headphone to both ears. The participants were seated in front of a computer with a mouse in front of them. The level of the playback was set at around 75 dBA. The participants were instructed to :

[1] available at https://github.com/gallunf/SR2/tree/master/Experiment/CRM_original

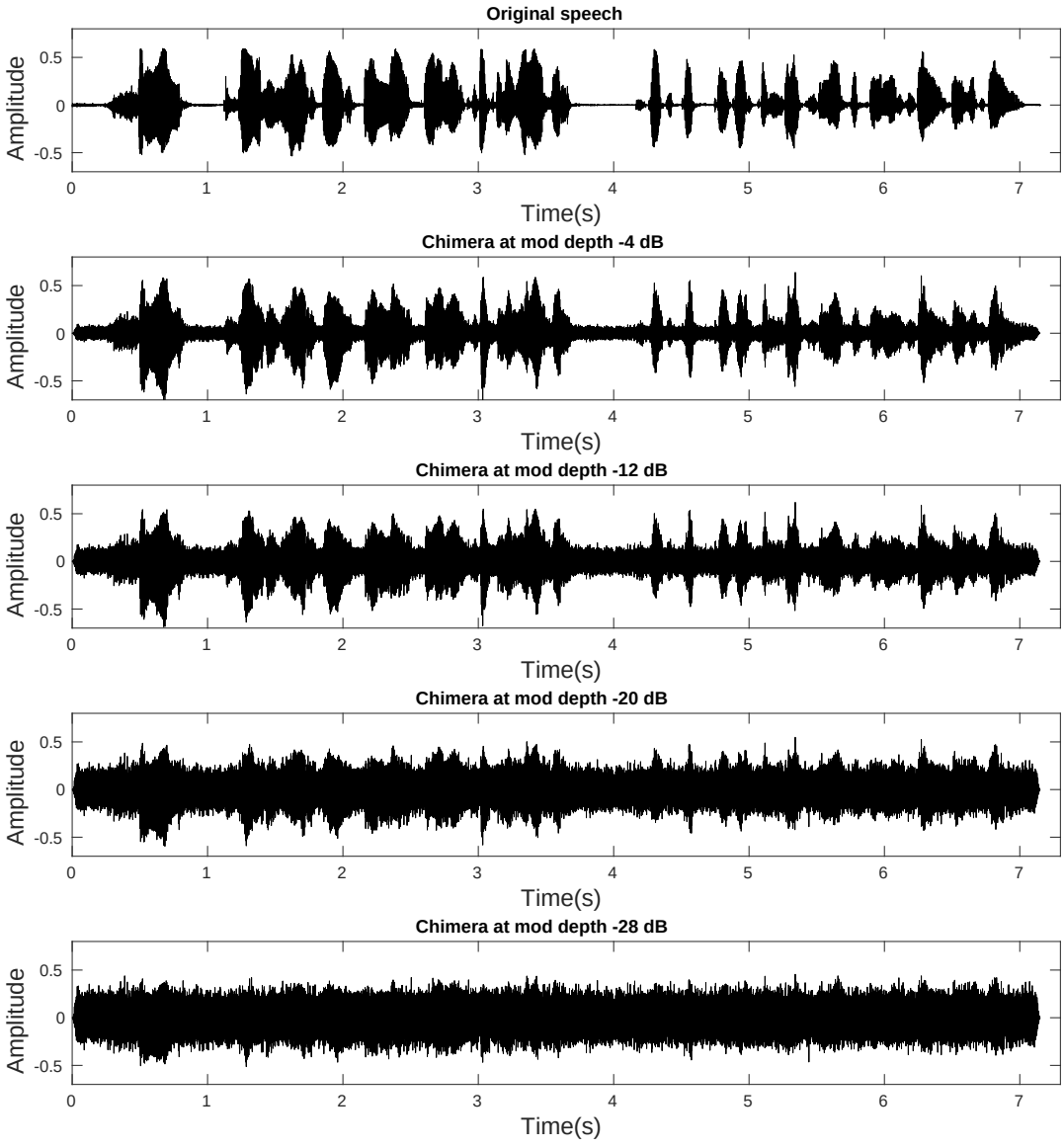



Figure 4.2: Plots of a speech signal and chimera corresponding to various modulation depths

1. Use the mouse to click on a colored square corresponding to the instruction they heard.

2. Perform the task at a pace or speech that feels natural or comfortable.

3. Take a short break, if needed, when the rating scale questionnaire appears.

In the experiment, 36 values of modulation depths were chosen in equal step size from 0 dB to -45 dB. For each modulation depth, 10 unique signals were chosen randomly from this dataset. These chimera signals were presented to the participants in random order, that is, every participant experienced the same set of 360 (=36 modulation depths * 10 repetitions) signals in each ANC condition. The purpose of this design is to allow variations in the waveforms of speech signals to act as a proxy of real-world variability, yet be distributed uniformly across the modulation depths. The maximum limit and spacing of the modulation depths were chosen after conducting a few pilot experiments (not reported here) to keep the overall duration of the task manageable and still obtain sufficient data from each participant.

Listening effort was assessed by the NASA Task Load Index (Hart & Staveland, 1988) (NASA-TLX) at regular intervals during the experiment. After every 30 trials of the CRM task, the NASA-TLX questions were shown in a dialog box on the computer screen. Participants were instructed to carefully read each question and provide an answer by clicking on a radio-button to indicate the degree of agreement or disagreement with the individual questions. The questions and corresponding variable names representing them in outputs, are as follows.

1. mental_demand : How mentally demanding was the task?

2. physical_demand : How physically demanding was the task?

3. temporal_demand : How hurried or rushed was the pace of the task?

4. performance : How successful were you in accomplishing the task?

5. effort : How hard did you have to work to accomplish your level of performance?

6. frustration : How insecure, discouraged, irritated, stressed and annoyed were you?

NASA-TLX has been utilized by speech psychometric experiments to assess listening effort or mental workload in a variety of contexts such as different types of noises and degradations, and generally shows

good agreement (correlation) with other forms of listening effort such as response time or pupil dilation (Alhanbali et al., 2019).

The experiment was conducted in a quiet room with noise absorbing panels attached to walls, and the ambient sound level was around 46 dBA, caused by air conditioning machine. The data from participants were collected on a computer running Ubuntu Linux (version 20.04 LTS) using a custom MATLAB script, which utilized the functions of the Psychophysics Toolbox (Brainard, 1997) (PSYCHTOOLBOX [2]) version 3. The chimera signals were digital-to-analog converted through a two-channel USB audio interface (M-track 2x2 by M-audio, Rhode Island, United States) before being played through the respective headphone.

Prior to the start of the main experiment, participants were given a practice session with relatively easy conditions (modulation depth less than -10 dB and dc bias lower than the actual experiment). An additional purpose of this practice session was also to check whether the participants were able to follow the instruction correctly, e.g. whether any of them had color blindness or other problems that prevented them from doing the task.

The experiment design was approved by the institute ethics committee prior to data collection. Participants were recruited from the university students, who were undergraduate or postgraduate students. Participants signed up to participate in the experiment in response to e-mails sent through a centralized e-mail address. All participants gave informed consent prior to the start of the experiment. Due to the nature of the experiment, participants were allowed to take part only once i.e. each participant reported in the data below is unique.

### 4.2.4 Statistical analysis

For all the three dependent variables, the principal goals were to answer the questions:

- whether or not the dependent variable shows a statistically significant difference across 'ANC on' and 'ANC off'.
- whether any of the headphones perform better or worse than the others.

For each dependent variable (intelligibility, response time and all NASA-TLX questions), a linear model was used to calculate a two-way analysis of variance (ANOVA) with two factors, i.e. ANC condition, and

[2] available from http://psychtoolbox.org/

headphone type. The first factor answered the question of whether the dependent variable varied significantly across the 'ANC on' and 'ANC off' conditions. The second factor answered the question of whether any of the headphones were statistically significantly different from the others. Finally, the interaction factor answered whether any of the headphones' ANC was better or worse than others.

Intelligibility in a given trial is defined as a binary variable, which is averaged across all presentations of a given modulation depth to arrive at a number between 0 and 1. The distribution of raw values of these kinds of measurements are usually skewed, being clustered around either 0.0 or 1.0 depending on the type and difficulty of experiment. Due to this clustering, the data tends to violate the assumptions of linear model based statistical tests which require that the data has equal variances across groups and residuals be normally distributed. To mitigate these problems, researchers dealing with this kind of proportion data (e.g. correct vs. incorrect, alive vs. dead etc.) often use certain non-linear transformations of the data before performing statistical tests on them. There are many such transforms, the most commonly used in speech and language research being the arcsine transform. Arcsine transform is defined as

$$p_a = 2 * \sin^{-1}(\sqrt{p}) \tag{4.1}$$

where $p$ represents the raw intelligibility (between 0 and 1). However, even this transformed version has other problems, such as poor interpretibility and certain other problem with variances. To mitigate these problems, Studebaker (1985) proposed a "rationalized" version of the arcsine transform which was defined as

$$p_{\text{rau}} = 46.47324337 * p_a - 23 \tag{4.2}$$

where $p_a$ refers to the same quantity as in equation 4.1. In this work, the rationalized arcsine transform was applied to the intelligibility data and thus the numbers reported are in RAU (Rationalized Arcsine Units).

Response time in the present study was defined as the number of milliseconds since the end of the audio playback and the down event of the mouse click, as reported by the operating system or MATLAB. For the purposes of this study, only the response time of correct trials were considered. This quantity can sometimes become negative if the participant is too eager and manages to click before the audio stopped playing. It can also have unusually high outliers due to any unpredictable thoughts or distractions. Hence, the pooled response time values (from all participants across all headphones) were first subjected to an outlier removal algorithm with lower threshold of 10 percentile minus the inter-quartile range (IQR) and

upper threshold of 90 percentile plus 2.5 IQR. The exact parameters of these values were chosen after trial and error by trying to minimize the number of rejected samples while maintaining a reasonable upper threshold (about 4500 ms). The resulting values were then submitted to a linear model with two factors as described before.

The NASA-TLX questions were answered on 21-point rating scales with "1" representing "very low" to "21" representing "very high". Any rating scale with more than 12 points can be interpreted as an ordinal level measure (Harpe, 2015), so these data were not subjected to any transformations or other changes. These values were directly forwarded to the analysis functions.

## 4.3 Results and discussions

Total 80 participants (55 M + 25 F) took part in the experiment, who were distributed equally over the four headphones. Participants were aged between 19 and 26 years, with a mean age of 21.2 years. None of them had any hearing impairment or hearing loss according to self-report. The mean duration of the experiment was 74 minutes including breaks taken in between, plus some additional time for the training session.

**A. Response time**

According to the outlier removal criterion defined above, the upper cutoff of the response time was found to be 3825 ms and the lower cutoff was found to be 49 ms, and a total of 46573 trials were identified as valid (pooled over all headphones and all participants). Plots of the response time against modulation depths are shown in figure 4.3, and statistical comparisons of response time across ANC conditions and across headphones are given in table 4.3. Mean and standard deviations of response time are given in table 4.2. Response time is seen to have an overall slight trend of higher values at lower modulation depths, thus weakly supporting the idea that more degraded stimuli takes longer to decide on. However, in terms of the ANC effects the plots show that for most modulation depths, the response time of the 'ANC on' conditions are higher than the 'ANC off' conditions. This is counter-intuitive because the ANC is known to be increasing the intelligibility at least slightly; whether this is at a cost of increased cognitive load is unclear. This may be useful to interpret in terms of theories of cognitive processing that integrate the concept of 'decision making' along with sensory aspects, which is outside the scope of the present work.

**B. Intelligibility**

The statistical F-test results for comparing intelligibility across ANC conditions and across headphones

are shown in table 4.3, and the mean and standard deviations of intelligibility are given in table 4.2. The intelligibility (in RAU units as per (Studebaker, 1985)) plotted across modulation depth for each headphone is shown in figure 4.4. These plots show some interesting patterns. First, there is an overall negative slope in the plot of each headphone and ANC condition, thus validating the idea that removing the envelope does reduce the intelligibility of speech. Second, the intelligibility is not always higher or lower across the ANC conditions, i.e. the relative difference between the ANC condition is not the same across the modulation depths; this may indicate an inconsistent effect of the ANC across different amounts of degradation, or adaptation by the brain, or simply an artifact of the specific sentences that were presented in those particular modulation depths. Third, there is sharp drop and recovery at about -28 dB (figure 4.4). This is most likely due to a kind of 'phase shift' where the brain stops relying on the envelope and switches to using TFS for listening to the speech; this shifting balance of speech intelligibility on TFS vs. ENV has been studied for some time (Shamma & Lorenzi, 2013). It is possible that the contribution of TFS under normal listening conditions is more towards quality and 'goodness' aspects such as understanding the emotional content and similar non-articulated aspects of speech, but under stressful acoustic conditions such as noise or other types of distortion the brain prioritizes TFS over ENV (Zeng et al., 2005).

**C. NASA-TLX**

The distribution of ratings obtained for all six questions of the NASA-TLX are shown in figure 4.5. In each box-plot, stars represent mean, horizontal lines represent median, and notches represent 95% confidence interval around median. NASA-TLX is usually considered a proxy for the general workload experienced by a user of a new device, product or interface. Here, the question of physical demand had little applicability and should be treated as semi-irrelevant. The rest of the questions, taken together, represent the overall effort experienced by the participant to carry out the listening task. The differences between 'ANC on' and 'ANC off' are non-significant in all but the perceived performance question, which is easily explained given that the participants received correct/incorrect feedback after each trial. More interestingly, the headphones were seen to be eliciting different ratings even though the ANC was failing to make a difference; this alone could be an interesting finding for manufacturers and serve as a stepping stone for future studies in quality control and product design.

Overall, the results suggest that this particular experiment is unsuitable for assessing ANC headphones because the effect sizes are generally low and the trends are hard to interpret as quality index of any kind. Further research with other types of chimera, other external noise conditions and other experimental de-

signs (e.g. using a different questionnaire for listening effort) may make it possible to design a better characterization of the ANC headphones in terms of the envelope perception.

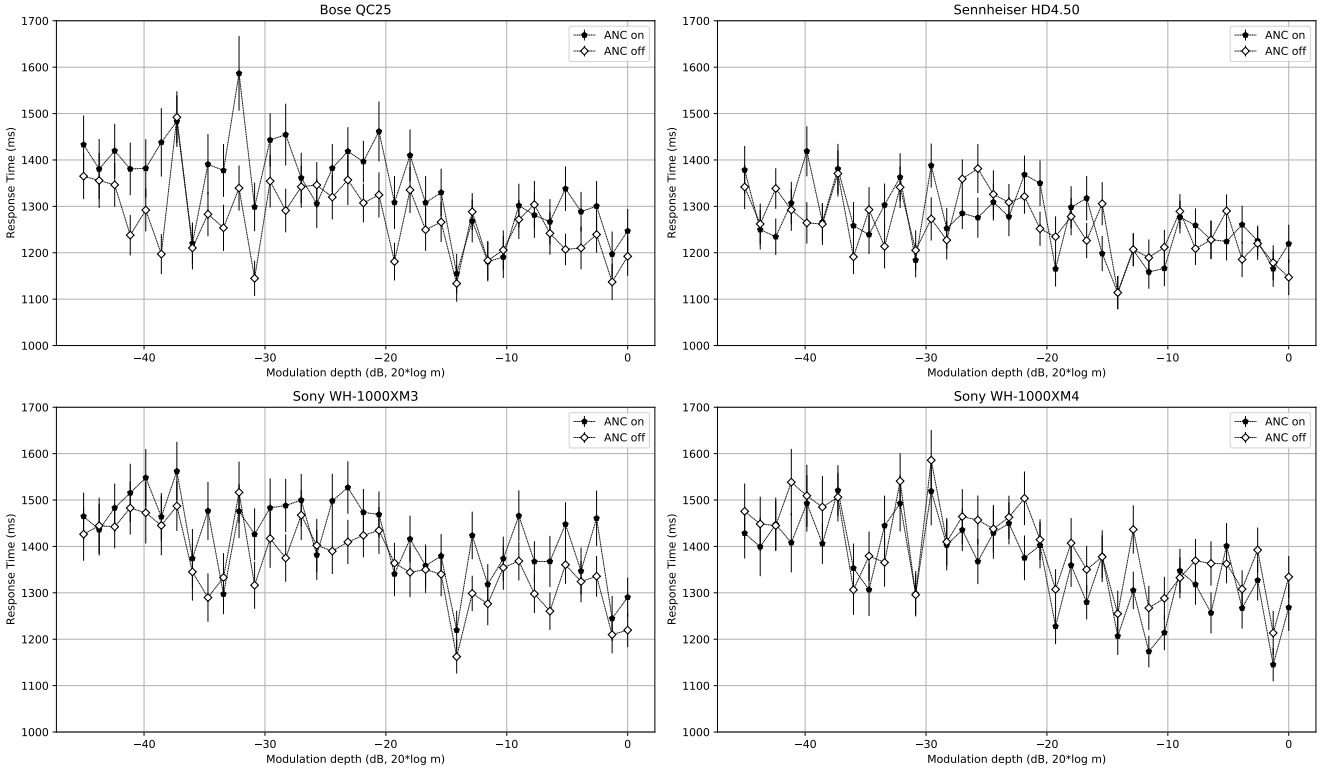



Figure 4.3: Response time against modulation depth in chimera experiment

Table 4.2: Mean and standard deviation of response time and intelligibility across headphones

| | ANC | HD 4.50 | | QC25 | | WH-1000XM3 | | WH-1000XM4 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Stdev | Mean | Stdev | Mean | Stdev | Mean | Stdev |
| Intelligibility (RAU units) | On | 82.8917 | 21.11 | 84.7207 | 19.56 | 85.1606 | 19.96 | 82.6909 | 20.56 |
| | Off | 82.6715 | 21.93 | 83.6995 | 20.05 | 82.7458 | 21.82 | 81.7538 | 20.65 |
| Response Time (millisecond) | On | 1258.91 | 537.9 | 1293.11 | 593.4 | 1381.52 | 615.5 | 1318.8 | 550.2 |
| | Off | 1251.22 | 522.6 | 1256.93 | 569.3 | 1339.33 | 589.5 | 1356.58 | 576.8 |

## 4.4 Chapter summary

The effect of ANC on human speech perception was studied with the help of chimeras that had their envelope parametrically removed. The speech intelligibility, response time and listening effort were measured from human listeners through four commercially available headphones. When compared across 'ANC on' and 'ANC off', both the intelligibility and response time showed mixed results when seen across modulation depths. The listening effort ratings were non-significant due to ANC across ANC conditions, but interestingly they were significantly different across headphones in four out of six questions.

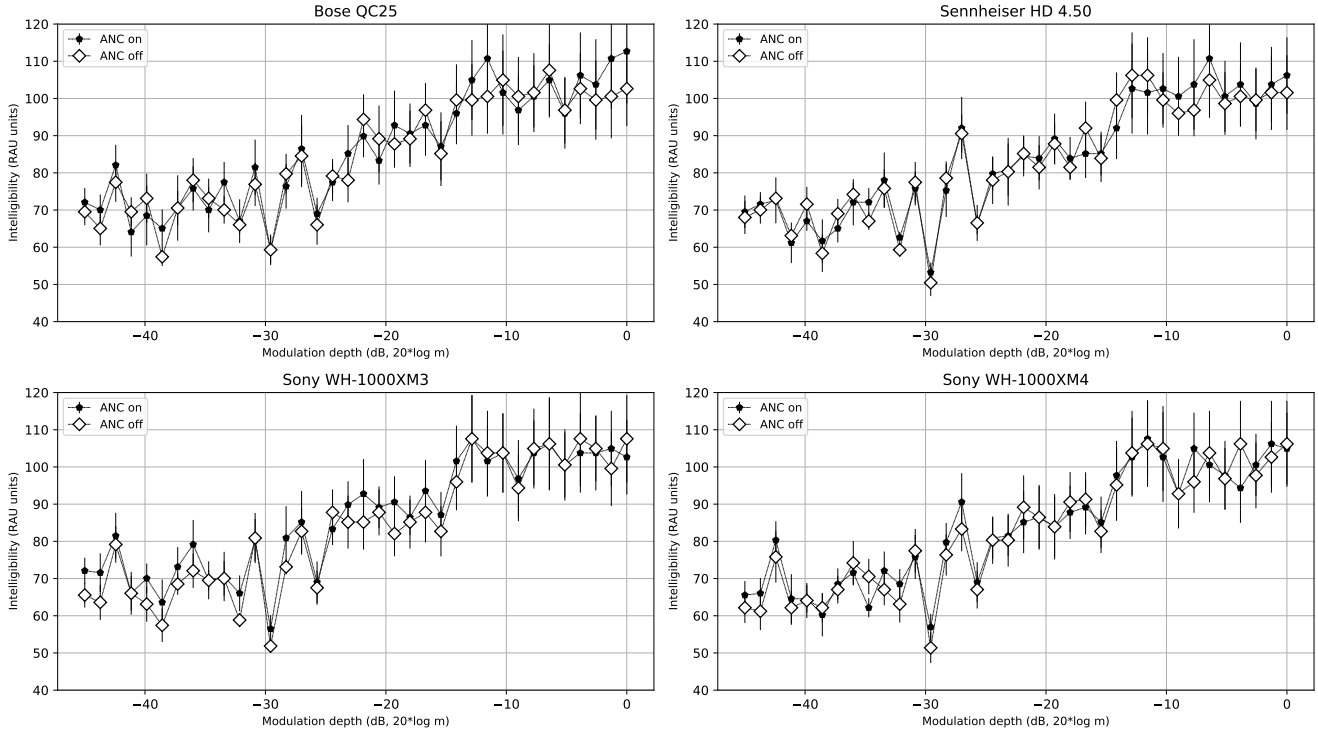
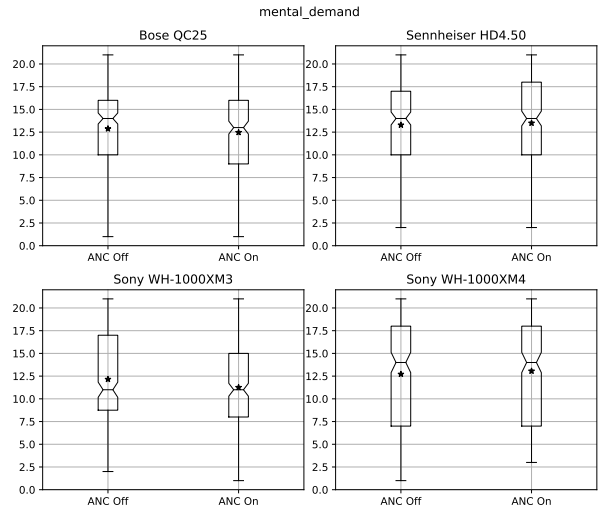


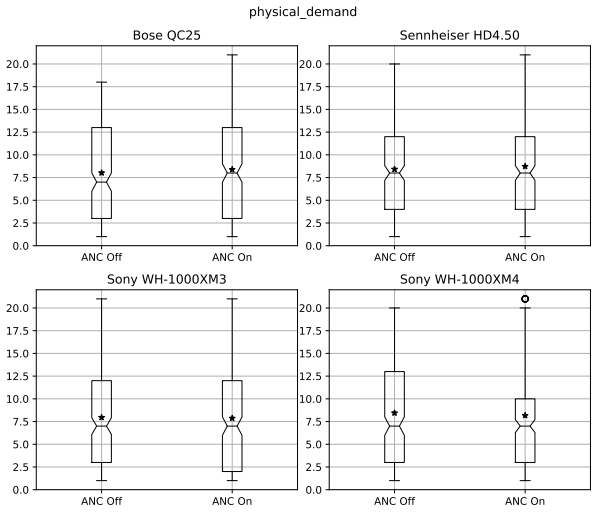

Figure 4.4: Intelligibility (rationalized) against modulation depth in chimera experiment

Table 4.3: Statistical comparison (two-way ANOVAs) of intelligibility and response time

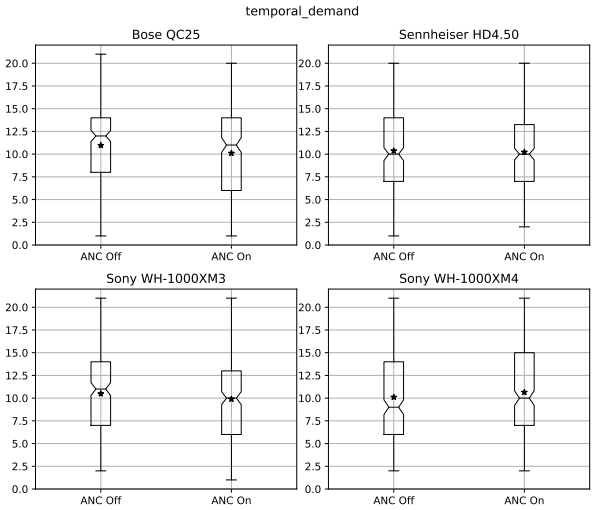
| | factor | F-value | p-value | Effect size ($\eta_p^2$) |
|---|---|---|---|---|
| Intelligibility | Headphone Type | F(1,5752)=3.001 | **0.029** | 0.0016 |
| | ANC condition | F(1,5752)=4.423 | **0.035** | 0.0008 |
| | Interaction | F(3,5752)=0.706 | 0.548 | 0.0004 |
| Response time | Headphone Type | F(3,46565)=89.78 | **<0.001** | 0.0058 |
| | ANC condition | F(1,46565)=5.747 | **0.016** | 0.0001 |
| | Interaction | F(3,46565)=11.86 | **<0.001** | 0.0008 |



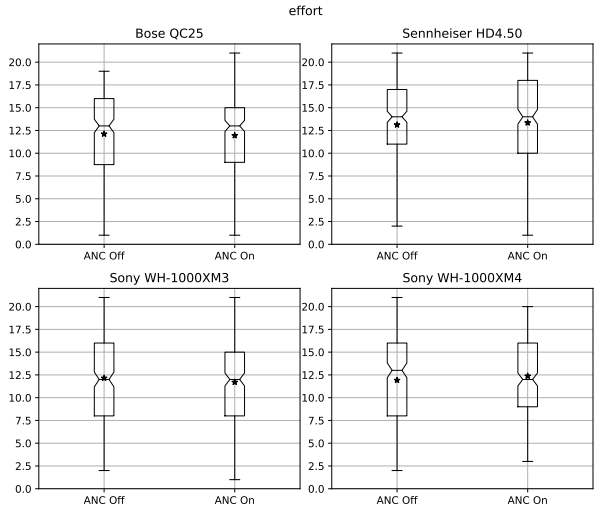

(a) "How mentally demanding was the task?"



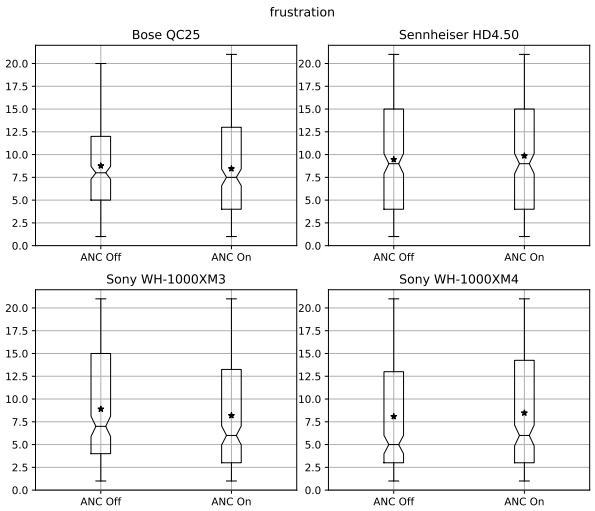

(b) "How physically demanding was the task?"



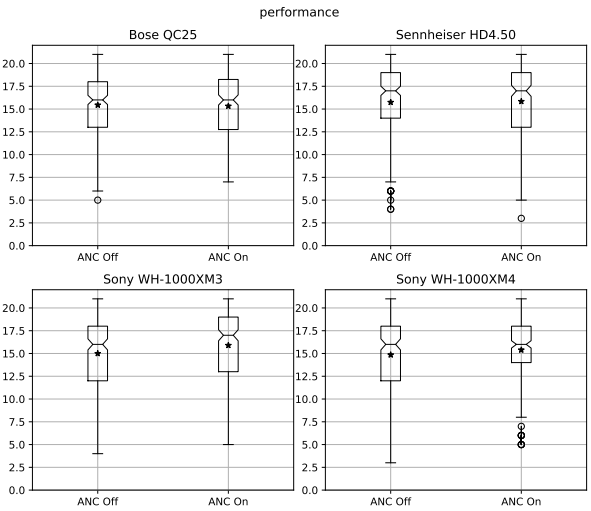

(c) "How hurried or rushed was the pace of the task?"




(d) "How hard did you have to work to accomplish your level of performance?"




(e) "How insecure, discouraged, irritated, stressed and annoyed were you?"




(f) "How successful were you in accomplishing the task?"

Figure 4.5: Distribution of NASA-TLX ratings in chimera experiment

Table 4.4: Statistical comparison (two-way ANOVAs) of NASA-TLX ratings

| Dimension | factor | F-value | p-value | Effect size ($\eta_p^2$) |
|---|---|---|---|---|
| Mental demand | Headphone Type | F(3,1912)=9.103 | **<0.001** | **0.0100** |
| | ANC condition | F(1,1912)=0.612 | 0.434 | 0.0003 |
| | Interaction | F(3,1912)=1.512 | 0.209 | 0.0024 |
| Physical demand | Headphone Type | F(3,1912)=1.189 | 0.312 | 0.0019 |
| | ANC condition | F(1,1912)=0.086 | 0.769 | 0.0000 |
| | Interaction | F(3,1912)=0.382 | 0.766 | 0.0006 |
| Temporal demand | Headphone Type | F(3,1912)=0.389 | 0.761 | 0.0006 |
| | ANC condition | F(1,1912)=1.306 | 0.253 | 0.0007 |
| | Interaction | F(3,1912)=1.765 | 0.152 | 0.0027 |
| Effort | Headphone Type | F(3,1912)=7.909 | **<0.001** | **0.0100** |
| | ANC condition | F(1,1912)=0.007 | 0.935 | 0.0000 |
| | Interaction | F(3,1912)=0.894 | 0.443 | 0.0014 |
| Frustration | Headphone Type | F(3,1912)=4.835 | **0.002** | **0.0075** |
| | ANC condition | F(1,1912)=0.044 | 0.834 | 0.0000 |
| | Interaction | F(3,1912)=1.002 | 0.391 | 0.0016 |
| Perceived performance | Headphone Type | F(3,1912)=2.494 | **0.058** | **0.0039** |
| | ANC condition | F(1,1912)=4.136 | **0.042** | **0.0022** |
| | Interaction | F(3,1912)=1.701 | 0.165 | 0.0027 |

In a headphone, ANC performance depends on the placement of the device on the head of the person wearing it, due to variability in the transfer functions of the gap between headphone and ear, i.e. the acoustic seal between the device and the head (Hansen et al., 2012, p. 802). These differences, coupled with other factors such as eyeglasses or hair length/thickness can lead to large differences in outcomes. Additionally, some participants could have prior knowledge of ANC from personal experience, that may lead to biases or differences in listening strategies which would again lead to differences in outcomes in experiments like the one reported here.

### 4.4.1 Future work

The work reported here can be extended in various ways, such as:

- The chimera used here are based on modulation depth degradation. Other types of chimera can be implemented, with the goals of degrading the envelope in other ways. For instance, using comb filters one can produce chimeras with spectral holes in the amplitude envelope. Such signals would be able to probe questions such as whether the ANC helps perceive specific regions of the modulation frequency domain.

- Continuing from the above, chimera can be also implemented by using the family of techniques

known as 'vocoding', where the original TFS is removed completely and a tone or noise carrier replaces it. Perceptual aspects of vocoded speech is widely studied because vocoding roughly corresponds to the processing done by cochlear implants (Rosen et al., 2015); such studies have extensively proved the relative contributions of ENV and TFS in human speech perception. Applying them in the context of ANC could reveal the nature and extent of the ANC effect in more detail.

- The listening effort can be assessed with other, more direct questionnaire as well as some kind of dual task during the speech intelligibility experiment. Pupil dilation or other kinds of objective (non-task based) measures could also be applied for this purpose.

- The ANC on and ANC off conditions can be presented multiple times each to assess learning effects. The order in which participants experienced the 'ANC on' and 'ANC off' conditions may have a learning effect. In the present work, this problem was mitigated by evenly distributing the participants among the two alternatives to cancel out such an effect, i.e. each participant experienced each ANC condition exactly once. Future studies could be designed to extend this design and use other patterns such as A-B-A-B or A-A-B-A-B-B or other permutations, instead of a simple A-B vs. B-A as done here.

- Custom headphones or ANC algorithms can be implemented in real-time signal processing systems (such as Speedgoat) and effects of various parameters of the ANC system can be studied.

# 5

# Instrumental audio quality of shallow-envelope chimera through ANC

## 5.1 Introduction

This chapter is an extension of chapter 4, where the goal is to use instrumental audio metrics a.k.a 'virtual listeners'. Shallow-envelope chimeras recorded with dummy-head microphone were evaluated with various instrumental metrics, to see the extent of the ANC effect in terms of each metric.

There has been a rapid growth in the number of audio related technologies in the last two decades. As audio devices or technologies form a cascade that can uniquely determine the quality of a human listeners' experience, the concept of evaluating the sound quality of audio processing pipelines have also become increasingly important. In a real-time audio telecommunication application (e.g. Zoom or Skype), it is common to have digital signal processing pipeline that perform separate tasks like echo cancellation, packet loss concealment, and so on. The net result of all these interacting modules on the perception of the output by a human listener can be tricky to estimate. Thus, the concept of perceptual evaluation is applied to find some estimate of the goodness or meet certain performance expectations (Bech & Zacharov, 2007). In audio or speech quality evaluation, the gold standard is a listening test a.k.a 'jury test'. In these tests, listeners (trained or untrained, as needed) are presented with various audio samples and asked to give rating scales

on likert-type scales, in terms of absolute categories (e.g. 'pleasantness') or in terms of pairwise comparisons (e.g. 'soft' vs. 'rough'). The goal of such tests is often to choose between audio processing parameters, or to demonstrate the validity of some algorithm such as de-reverberation. However, carrying out such tests with actual human participants requires resources and are time consuming. Moreover, in rapidly evolving audio technologies such as those deployed in telecommunication, speed of execution is necessary to iterate through ideas or prototypes or parameters. To meet these challenges, a variety of speech and audio quality models have been proposed. Broadly, these are signal processing algorithms that fall into two categories: full-reference and no-reference, where the former requires both an original and degraded signal as input and the latter requires only the degraded signal as input (Rix et al., 2006; Campbell et al., 2009). These algorithms are called 'objective' or 'instrumental' models of perception. In many sub-fields of acoustic signal processing such as noise removal, de-reverberation and so on, objective models of human speech or audio intelligibility/quality perception have been developed. In most cases, such models are domain-specific, i.e. the signal flow inside the model is hand-crafted by researchers and tweaked to match the human perceptual scores in some specific kind of distortion. Here we deploy ten somewhat generic models from recent work, most of which perform computations on the amplitude envelope internally, to compare ANC headphones' output recorded with dummy-head microphones, and look for patterns and trends across ANC conditions and headphones.

Accurately assessing the speech quality of an audio device is a difficult problem, because of hidden and poorly understood non-linearities in the interaction between human cognitive factors like attention and physical factors such as distortion, addition of noise etc. (Akhtar et al., 2019). Nevertheless, in the last decade, the concept of using instrumental quality metrics to evaluate headphones have also gained popularity (Rumsey, 2019; Temme, 2019). Hearing aids are also routinely assessed in terms of instrumental models, and the results are said to be 'promising' (Harlander et al., 2014). In the context of ANC headphones or hearables, some work has been done using instrumental models and much more using actual human listeners. Some of these studies based on listening tests based on multidimensional scaling etc. have been described in section 1.2. Steeneken (2005) studied ANC enabled hearing protection devices in the presence of stationary and impulsive noises to predict speech transmission index (which is based on the modulation transmission concept), and found that ANC improves this metric and the effect increases at higher noise levels. Reimes (2019) attempted to build an objective model to predict listening effort of ANC headphones, which is somewhat compared to other predicted variables. Biberger et al. (2021) applied a number

of instrumental quality and intelligibility models on three different hearables recorded audio datasets that included dummy-head recordings from different modes of operation such as 'hear through mode'. They found that amplitude envelope models (including GPSMq and HASQI, used in the present work) generally performed well.

It is important to note that in the present work, we are not interested in evaluating the *performance* of the objective metrics per se, unlike other studies that utilize a large number of instrumental models (e.g. (Hilkhuysen et al., 2014; Biberger et al., 2021)). Moreover, we are also not involved in finding specific values or thresholds using these metrics to claim specific performance improvements of one headphone over another. These and other ideas such as building an 'ANC quality index' may be possible as part of future work. For the present work, the goal is simply to establish a broad set of patterns in terms of the amplitude envelope fidelity in playback signal.

### 5.1.1 Instrumental audio/speech metrics

The instrumental speech quality or intelligibility models used in this chapter are briefly introduced below.

**A. Fidelity of Envelope Shape (FES)**

This metric was proposed by Stone and Moore (2007), and it is described as "…as the degree to which the envelope shape of the target in different frequency channels is preserved following [a signal-processing algorithm]". It is defined as the average of a correlation-based function over a given set of frequency subbands, between the envelope of original and degraded signals. It can be used in a flexible way for speech signals with any carrier-frequency bandwidth (narrow, wide etc). Its original application was intended to be on speech signals after passing through compression which is common in hearing aids, which has inspired other metrics such as envelope difference index (chapter 2). Even outside the context of hearing aids, compression is a common operation used in audio processing software and devices, hence this measure can be interpreted as a generally informative quantity about the nature of the envelope shape.

**B. Generalized Power Spectrum Model for quality (GPSMq)**

This metric is based on the model of modulation perception model based on modulation filterbank concept (Dau et al., 1997a, 1997b). It extracts the amplitude envelope of both the input signals to compute an internal representation. Then signal-to-noise ratios using the power and envelope power of these internal representations are used to derive a perceptual degradation score (Biberger et al., 2018). The authors initially

validated GPSMq by comparing it with four datasets of audio recordings; including audio codecs, noise reduction algorithms, audio source separation and loudspeaker distortions. They found that GPSMq performed better (in terms of accuracy, consistency and monotonicity) than other metrics (including HASQI, PESQ and PEMO-Q) for the loudspeaker distortions; they stated that it would not be able to accurately predict quality degradations in telecommunications such as phase jitter, variable delay, or time warping. GPSMq is a distortion metric and thus higher values indicate worse audio quality, and it has a range of 4.0 to 17.6 (Biberger et al., 2018). GPSMq has been further evaluated against datasets of distortions typically found in hearables devices; it was again found to performing well in terms of consistency, monotonicity and accuracy (Biberger et al., 2021).

**C. Hearing Aid Speech Perception Index (HASPI)**

Proposed by Kates and Arehart (2014a), this metric provides speech intelligibility measure based on the amplitude envelope. It compares the envelope and TFS of the two input signals, and computes a combined measure of coherence and envelope fidelity. Since it is intended to be used to test hearing aid functionality, it also has a scheme for incorporating hearing profile (audiogram) into the calculation. It was validated against certain datasets such as noise and nonlinear distortion, frequency shifting, noise suppression using an ideal binary mask algorithm; it was found to be a better predictor for both normal-hearing and hearing-impaired data than certain other intelligibility metrics that use similar principles (Kates & Arehart, 2014a). The second version of this metric (which is used in the present work) has been updated to utilize an updated auditory model which was an ensemble of ten neural networks. Neural networks were chosen for their ability to approximate an arbitrary function while preventing problems such as 'overfitting', and this model showed better accuracy than HASPI version 1 (Kates & Arehart, 2021).

**D. Hearing Aid Speech Quality Index (HASQI)**

Proposed by Kates and Arehart (2014b), this metric predicts the speech quality, and also contains provision for accounting for hearing-impaired and normal-hearing differences in the form of audiometric profile. It also uses both the amplitude envelope and TFS, by "a nonlinear term sensitive to noise and nonlinear distortion and a linear term sensitive to long-term spectral changes, and the quality prediction is formed by taking the product of these two terms." HASQI has been deployed for purposes beyond its original scope by Kendrick et al. (2015), where it was used as the target score to train some classical machine learning algorithms (decision trees) to predict perceived audio quality of sounds degraded by crowd-generated audio content on portable devices like mobile phones and tablets. Prior to the current version, it was also used to

predict the mean opinion scores of normal-hearing individuals (Kressner et al., 2013), which is related to but distinct from speech quality. HASQI has also recently been evaluated against different datasets of typical distortions introduced by headphones and hearables devices, and found to be one of the well performing metrics in terms of consistency, monotonicity and accuracy (Biberger et al., 2021).

**E. Perceptual Evaluation of Speech Quality (PESQ)**

Proposed by ITU in Recommendation P.862, this model aims to predict speech intelligibility. It utilizes an auditory model that creates an internal representation, and then compares the original and degraded signals on the basis of that internal representation. There is no overt dependence of the PESQ algorithm on amplitude modulation or its properties, however it does utilize a kind of envelope defined as the logarithm of the ratio of energy and a constant in sliding 4 millisecond windows. This concept is used to find the alignment between the two signals. The output of PESQ is a score in the range of −0.5 to 4.5, which represents a mean opinion score like scale (ITU-T, 2001). It is common for speech processing papers to have PESQ as a proxy for human hearing, for instance to evaluate the effect of algorithms such as noise removal, feedback cancellation, network packet loss concealment and so on (Rix et al., 2006; Campbell et al., 2009). Due to its popularity, audio metrics (especially full-reference ones) proposed in later years often keep PESQ in the list of comparisons, e.g. (Hines et al., 2015; Biberger et al., 2018).

**F. Perception Model for Quality (PEMO-Q)**

Introduced by Huber and Kollmeier (2006), this is based on the modulation filterbank model of human auditory modulation processing system (Dau et al., 1997a, 1997b). The model used here returns a number of measures, of which the 'Perceptual Similarity Measure' is being used here. This measure is based on a correlation based function of internal representations, and has been shown to have a high correlation with the mean opinion scores obtained from human listeners (r=0.97) (Huber & Kollmeier, 2006).

**G. Short Time Objective Intelligibility (STOI)**

Proposed by Taal et al. (2011), this is an intelligibility metric that utilizes the correlation between the log of power of the envelopes of the original and degraded speech signals. STOI uses short time-segments (384 or 386 milliseconds) to calculate the correlation between envelopes, which is also based on a simple model that has only two free parameters (Taal et al., 2011). STOI was evaluated with listening tests on Danish sentences, and showed to properly predict the intelligibility of noisy speech signals, and of the outputs from noise-reduction algorithms. It gives a score in the range of 0 to 1, with higher values indicating better intelligibility.

### H. Virtual Speech Quality Objective Listener (VISQOL)

Proposed in 2012 and fine-tuned and customized for various scenarios over the years, (Hines et al., 2012; Hines et al., 2015; Chinen et al., 2020), this model aims to predict both speech intelligibility under various distortions that occur in the Internet-age audio applications such as jitter, latency, packet loss and so on. It is based on the concept of 'neurogram similarity index measure', which is basically an auditory model that relies on auditory nerve responses as an internal representation (Hines, 2012). VISQOL is not explicitly based on temporal envelope, rather, it uses a correlation function on the neurogram to calculate the similarity of the internal representation of the original and degraded audio. The present work is based on the 2015 version of VISQOL (Hines et al., 2015), which implements a sliding window to aggregate similarity over various 'patches' of the internal representation. It produces a score in the range of 1 to 5, to mimic the mean opinion score from jury tests a.k.a listening tests. Different versions of the model were developed to fine-tune and specifically account for VoIP issues such as time-warping and jitter. It was optimized on the NOIZEUS dataset, and is known to generalize beyond the datasets used for fine-tuning the model (Hines et al., 2015).

### I. Weighted Spectro-Temporal Modulation Index (wSTMI)

Proposed by Edraki et al. (2021), it is an intelligibility metric that is tuned to generalize its predictions to multiple distortion types (across multiple datasets). This is based on a spectro-temporal analysis, and is an extension of an earlier concept called 'STMI' (Elhilali et al., 2003); specifically, it aims to remedy the over-simplifying assumption of equal importance of all modulation frequencies by STMI. wSTMI essentially applies a frequency band importance function to the modulation spectrum (hence the 'w' for weighted) and accordingly calculates the final score. It is claimed to be a generalizable metric across untested acoustic conditions and degradations.

### J. Speech to Reverberation Modulation energy Ratio (SRMR)

This is the only no-reference metric used in the present work. Proposed by Falk et al. (2010), it aims to predict the intelligibility of de-reverberated speech signals, i.e. signals that have been recorded in reverberant conditions and then de-reverberated using signal-processing algorithms. Since reverberation smears the amplitude envelope of any signal (Drullman et al., 1994), the challenge of speech de-reverberation algorithms is to restore the speech envelope without introducing perceptible artifacts. SRMR uses the modulation spectrum to compute a six-dimensional 'coloration' space and ultimately computes a ratio of speech energy and modulation energy; it was validated with multiple de-reverberation algorithm outputs and three

other intrusive metrics and performed equally well or better (Falk et al., 2010).

## 5.2 Materials and Methods

The audio quality metrics described above were applied to signals recorded with ANC headphones through the dummy-head microphone. The apparatus and hardware setup used to record the signals from headphones are the same as described in section 1.5, figure 1.3. The headphones were fitted on the dummy head and a computer was used to play the chimera signals through the headphone. This process was repeated for 'ANC on' and 'ANC off' conditions for the headphones. The recorded signals were stored without compression as wav files for offline analysis. During analysis, the recorded signals and the original signals were given to the instrumental audio or speech metric. This scheme is illustrated in figure 5.1.

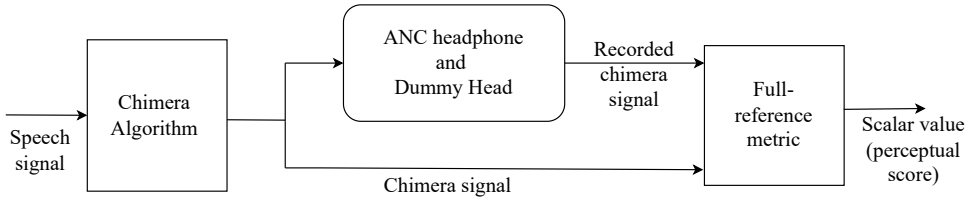



Figure 5.1: Schematic diagram of ANC headphone evaluation with objective metrics

### 5.2.1 Audio signals

The chimera algorithm described in section 4.2.1 was used to produce the signals for this chapter. Two datasets of speech signals were used:

- The clean speech signals from the NOIZEUS dataset (Hu & Loizou, 2007) were used. The dataset provides 30 speech samples each of roughly 3 seconds duration. These speech samples were passed through the chimera at 33 unique values of modulation depth, thereby providing a total 990 signals. The dc-bias was fixed at 0.0058, which was determined after trial and error.

- A subset of speech signals from the CRM dataset (Bolia et al., 2000) from the experiment described in chapter 4 were used. These consisted of 10 unique speech samples for each modulation depth. There were 26 unique values of modulation depth, thereby providing 260 total speech signals. The dc-bias was fixed at 0.008, which was determined on the basis of pilot experiments with participants.

Modulation depths were kept in the range of 0 dB to -32 dB because trial and error showed that this range was sufficient to provide a sufficient range of difficulty for the task of understanding the speech. The dc-bias of the two datasets were kept slightly apart to obtain a somewhat different amount of overall envelope degradation, to see the overall trends in variation across various perceptual measures.

For each instrumental model, the publicly available implementation or the one made available by the author(s) was used in an 'as-is' basis, i.e. no modifications were made to alter the algorithm in any way.

### 5.2.2 Statistical analysis

Similar to chapter 4, the questions of interest are:

1. Are there any differences across 'ANC on' and 'ANC off' conditions?
2. Are these scores different across the four headphones?
3. Is the difference across 'ANC on' and 'ANC off', if significant, more or less pronounced across the modulation depths?
4. Is there any noticeable trend or pattern in the score across the modulation depths?

The first, second and third questions can be answered using a two-way ANOVA (F-test); the effect size of the two factors and their interaction inform us about these answers. However the fourth question is more open ended, and cannot be easily assessed with a specific statistical test. This is because the patterns of the scores are often quite erratic, but not in a completely random way, hinting at deeper emergent patterns due to the complex interactions between headphone, ANC behavior, chimera algorithm, and the objective metric.

## 5.3 Results and discussions

The plots of the various objective metrics across the modulation depth (i.e the modulation depth functions or MDFs) are shown in figures 5.2 to 5.11, and the statistical comparison across 'ANC on' and 'ANC off' are reported in tables 5.1 and 5.2. Since these metrics are supposed to be used as a 'proxy' of human auditory perception, a change in these metrics can be taken as a measure of goodness of a headphone. In the

present work, all but the GPSMq metric have scores that are to be interpreted as 'higher is better'; this fact is represented by the arrows (↑ and ↓) next to the names in the tables 5.3 and 5.4.

Some of the important patterns of variation of the metrics are:

- Consistency of trends and absolute values across datasets: a different trend of the values across the two datasets indicate some non-linear effect arising out of the different dc-bias value across the datasets, and implies there would be even more substantial differences if the dc-bias values were to be more different. This also hints at interesting divergences from the real world applicability of the metric which is beyond the scope of the present work.

- Trend across modulation depth: The scores having a positive or negative slope across the modulation depth axis implies that the metric is able to capture the real-world relevance of the degradation due to the chimera algorithm.

- Variability across speech samples: this is represented by the errorbars in the plots. Larger errorbars indicates more significant dependence on the specific inputs and therefore the possibility that any trend may get reversed with another dataset (e.g. different language).

- Differences between headphones: this is one of the questions addressed by the statistical tests, in line with the statistical tests in the other chapters of this thesis.

- Differences across ANC conditions: this is the other common question similar to the other chapters, i.e. we are characterizing the effect of ANC with objective metrics this time and would therefore like to note any differences across ANC conditions. The question of whether ANC makes a difference and whether such difference (if any) is the same or different across the four headphones, is answered by the main effect and interaction term of the statistical test (two-way ANOVA).

A. **FES** (figure 5.2) shows an inconsistent effect for the Sennheiser headphone. Other than that, it has a fairly large difference across ANC conditions and different trends across the two datasets. It also has a large variability across individual modulation depths implying higher sensitivity to small changes in the nature of the envelope. Overall it can be said to be a poor candidate for a 'ANC index' objective metric.

B. **GPSMq** (figure 5.3) shows the smallest variability across the individual modulation depths (in both datasets), hence the error-bars are almost invisible in the plots. This is a surprising occurrence because there is almost always a small difference in the scores due to inherent differences across different talkers

and speech signal differences. It is consistent in the trend of the 'ANC effect' across both datasets, and the effect magnitude is different across the two datasets. Since GPSMq provides a distortion measure, the score is to be interpreted as 'lower is better'; it is interesting to note that in Sony WH-1000XM4 the absolute values of the 'ANC on' conditions are showing *less* degradation in CRM dataset even though the dc-bias (degradation of envelope) was higher there.

C. **HASPI** (figure 5.4) is showing some radically large changes in the Bose headphone with the NOIZEUS dataset. This could be interpreted as an artifact arising from the specifics of the conditions instead of the metric's capability. However, the effect size of the statistical tests shows that the overall ANC effect is present and of a similar magnitude compared to other metrics. The variability across modulation depth in Sony XM3 headphone shows an interesting crossover point around -8 dB in NOIZEUS dataset but not the other one, this could be informative for future studies.

D. **HASQI** (figure 5.5) shows a strong ANC effect in all cases except with Sony-WH1000XM3 headphone in CRM dataset. The scores are relatively close across the two datasets, showing that the model is perhaps robust to different kinds of real world inputs. HASQI is unable to show any trend across modulation depth, however, thereby indicating that perhaps it is not a good metric for assessing intelligibility of shallow-envelope chimeras, despite being based on the amplitude envelope. HASQI also shows a surprising divergence of the scores for NOIZEUS dataset in Sennheiser HD4.50 and Sony WH-1000XM3 headphones, but not in the other cases; we currently do not have an explanation of this.

E. **PESQ** (figure 5.6) shows a very strange non-monotonic pattern in NOIZEUS dataset that is consistent (in terms of trends, such as rise/fall cutoffs of modulation depth and slope) across all headphones, so this reflects something about the chimera signals rather than ANC or headphones in general. It is possible that PESQ is able to tap into the ENV/TFS dichotomy (mentioned in chapter 4) of the brain, and therefore could be very interesting to study as a phenomenon in its own right, given the ubituity of PESQ in speech processing literature. Interestingly, this pattern is completely absent in the CRM dataset, indicating that the dc-bias effect was strong enough to wipe out the ENV/TFS dichotomy that was probably arising in PESQ output in the NOIZEUS dataset. PESQ is also showing one of the high effect size (0.42) across ANC conditions, reflecting the fact that the ANC effect is real and evident.

F. **PEMO-Q** (figure 5.7) shows a consistent pattern of 'ANC effect' across the two datasets for all the headphones being studied, i.e. the 'ANC on' scores are consistently higher than 'ANC off' scores. This is expected since PEMO is explicitly designed to extract and compare envelope parameters, and thus the

relatively intact envelope when ANC is enabled compels it towards higher scores. Surprisingly, however, PEMO does not show a degradation in intelligibility across the modulation depths, which may be considered a failure of the model.

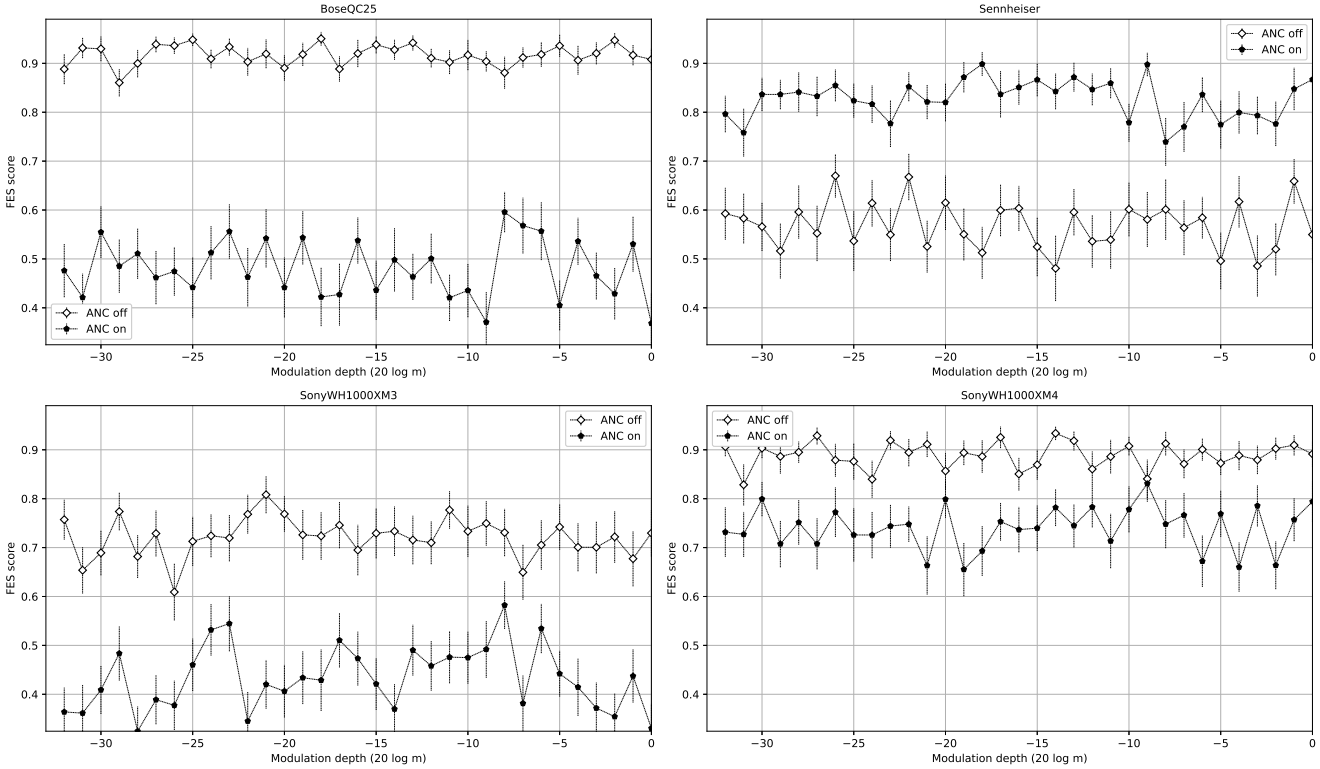
G. **STOI** (figure 5.8) is the only metric that shows non-significant difference across the headphones or the ANC conditions. However, it is also one of the few metrics that shows a positive slope across the modulation depth axis (especially in the NOIZEUS) dataset. This can be taken to imply that STOI is generally able to predict the speech intelligibility but not tuned to the speech quality. STOI is also showing a good agreement in terms of absolute values across the two datasets, implying that the difference in dc-bias did not matter much for its calculations. To the extent that there are differences across ANC conditions, it can be seen that STOI has higher scores for 'ANC off' than 'ANC on'. This can be due to any number of factors, and more detailed study regarding the precise nature of the interaction between envelope, metric and headphone might be able to unveil the reasons for these trends.

H. **VISQOL** (figure 5.9) shows a large difference across the two datasets, and also NOIZEUS shows a smaller variability across modulation depths, coupled with overall lower scores. This may be partly due to the fact that the model was optimized and fine-tuned with the NOIZEUS dataset. The difference caused by ANC also seems to be more pronounced in the CRM dataset than NOIZEUS, and the degradation across modulation depth is not captured by this metric at all. The effect size in the statistical test in this case is misleading, because the plots clearly show that VISQOL does not show a big difference across 'ANC on' and 'ANC off' for most of the headphones, but the one or two that does, dominates. Taken together, this can be interpreted to mean that VISQOL is a poor predictor of the intelligibility of the shallow-envelope chimeras developed here, and hence unsuitable for the task of assessing the goodness (or lack thereof) of ANC headphones in terms of preserving envelope fidelity.

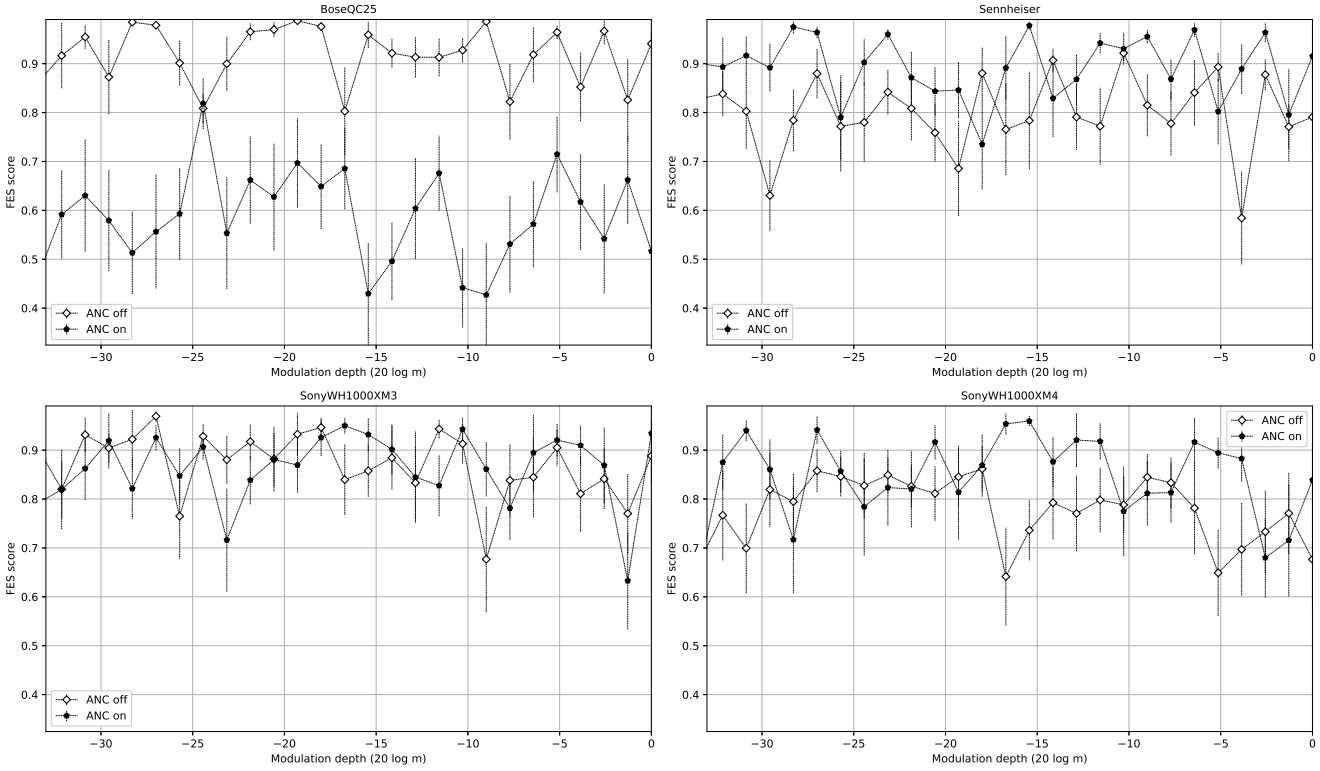
I. **wSTMI** (figure 5.10) shows a contradictory evidence of the 'ANC effect' across the two datasets (comparing Bose QC25 and Sony WH-1000XM4 headphones, each to itself). It shows a relative improvement in two out of four headphones in the NOIZEUS dataset, but the improvement is less evident in the CRM dataset. In addition, the absolute scores are also very different across the two datasets. This suggests that the dc-bias and also the emergent effects arising out of the headphone, wSTMI and chimera processing are unpredictable and unstable. However, wSTMI is able to capture the trend in the score across modulation depth in all headphones, thereby validating it as a useful metric for capturing the intelligibility due to the shallow-envelope chimera.

J. **SRMR** (figure 5.11) shows one of the most clear trend of falling intelligibility across modulation depth, thus validating it for predicting the shallow-envelope chimera. Importantly, it is the only non-intrusive metrics being utilized here, thus its results are not directly comparable to the rest of the metrics. It also has the distinction of being non-significant across ANC conditions in one dataset but not the other (tables 5.1 and 5.2), and that trend is also visible from the plots. Interestingly, SRMR shows 'ANC off' to have higher intelligibility for 3 out of 4 headphones in NOIZEUS dataset. Further, the range of scores across the two datasets are also very similar for all headphones, suggesting that SRMR does not show a lot of dependence of the dc-bias difference across the datasets.

Overall, it can be seen that the 'ANC on' and 'ANC off' conditions are likely different in terms of speech quality. The broad trends in these differences sometimes show unusual patterns across the modulation depths, across headphones and across the four devices under test. The effect sizes of the ANOVA for the ANC condition and headphone type factors provide evidence that these differences are not of the same magnitude across the four headphones being studied. The fact that most of the instrumental models used in the present work extracted and used envelope parameters, yet turned out to produce differences in macro-level trends in the headphone data, suggests that a deeper look is necessary to standardize the outcomes for any effort towards building an 'ANC index'.

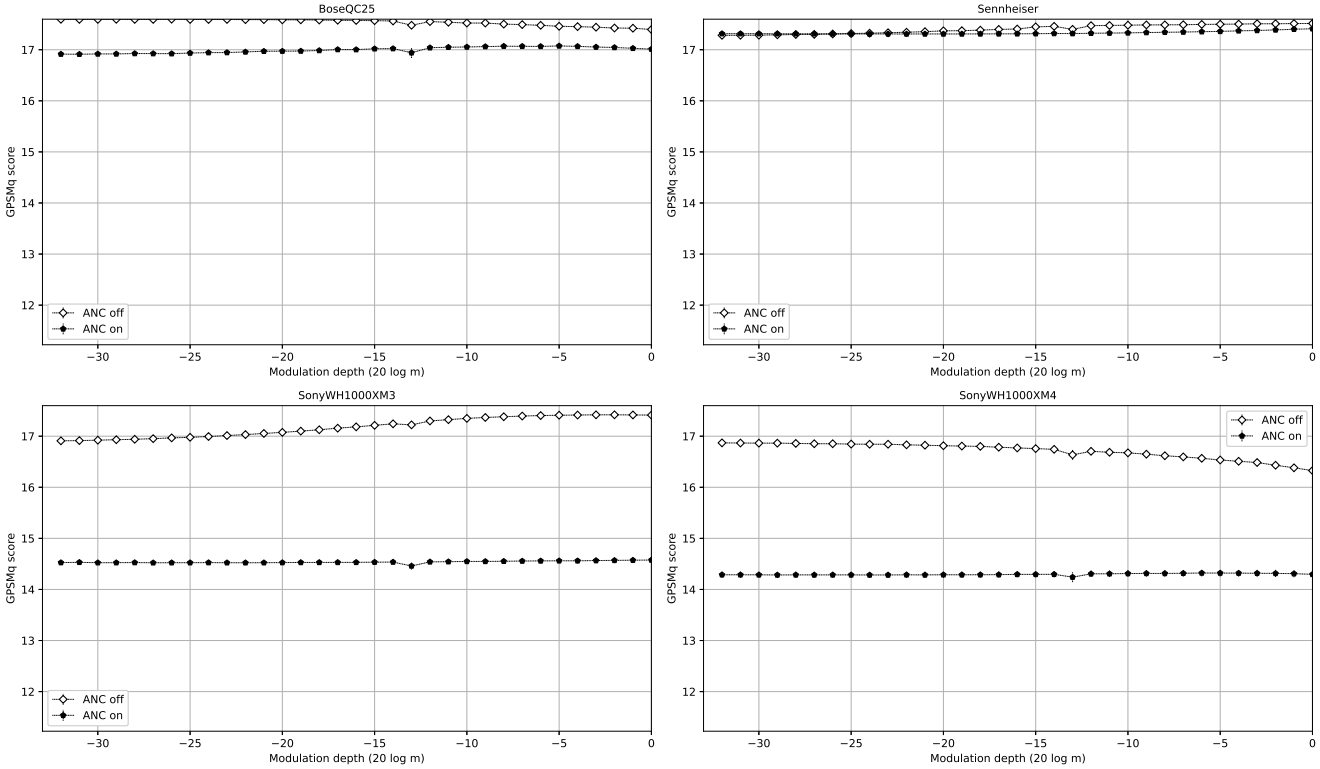



(a) NOIZEUS dataset

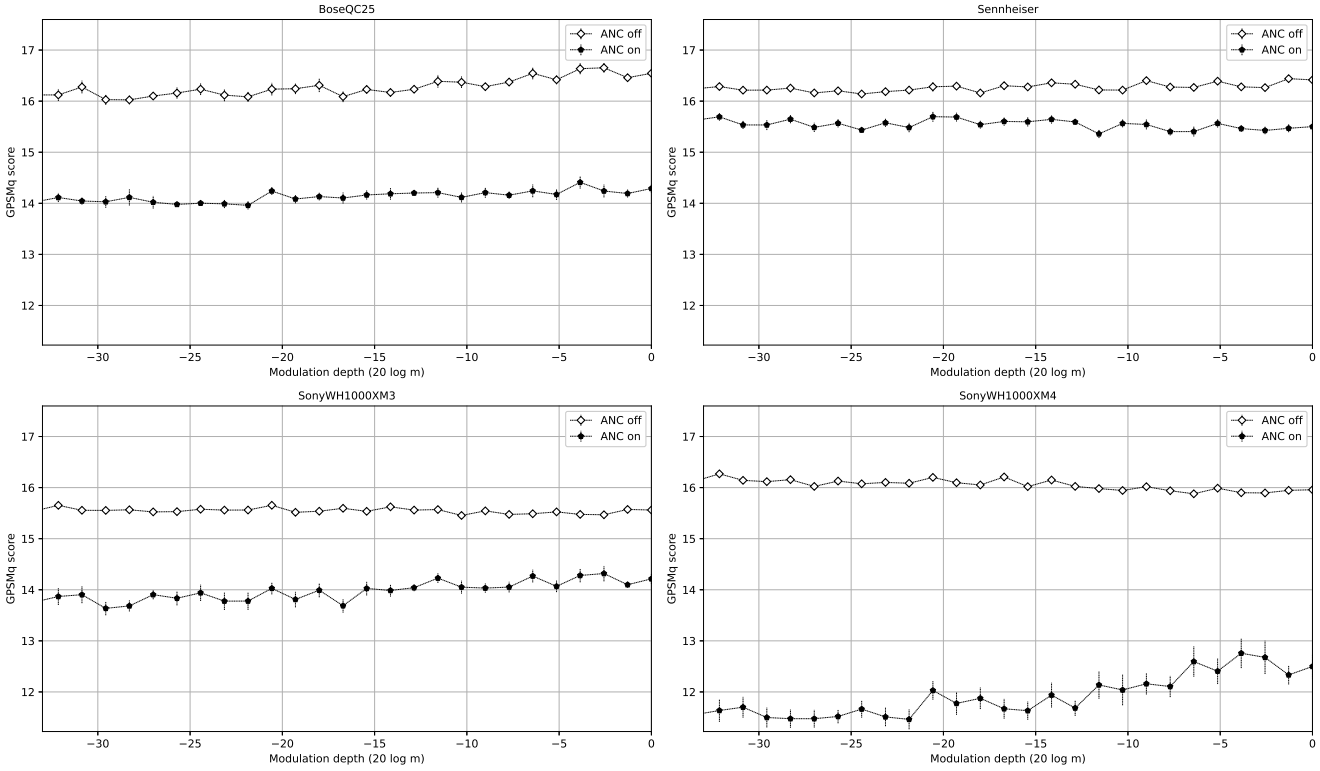



(b) CRM dataset (stimuli used in chapter 4)

Figure 5.2: Chimera audio quality with FES against modulation depth

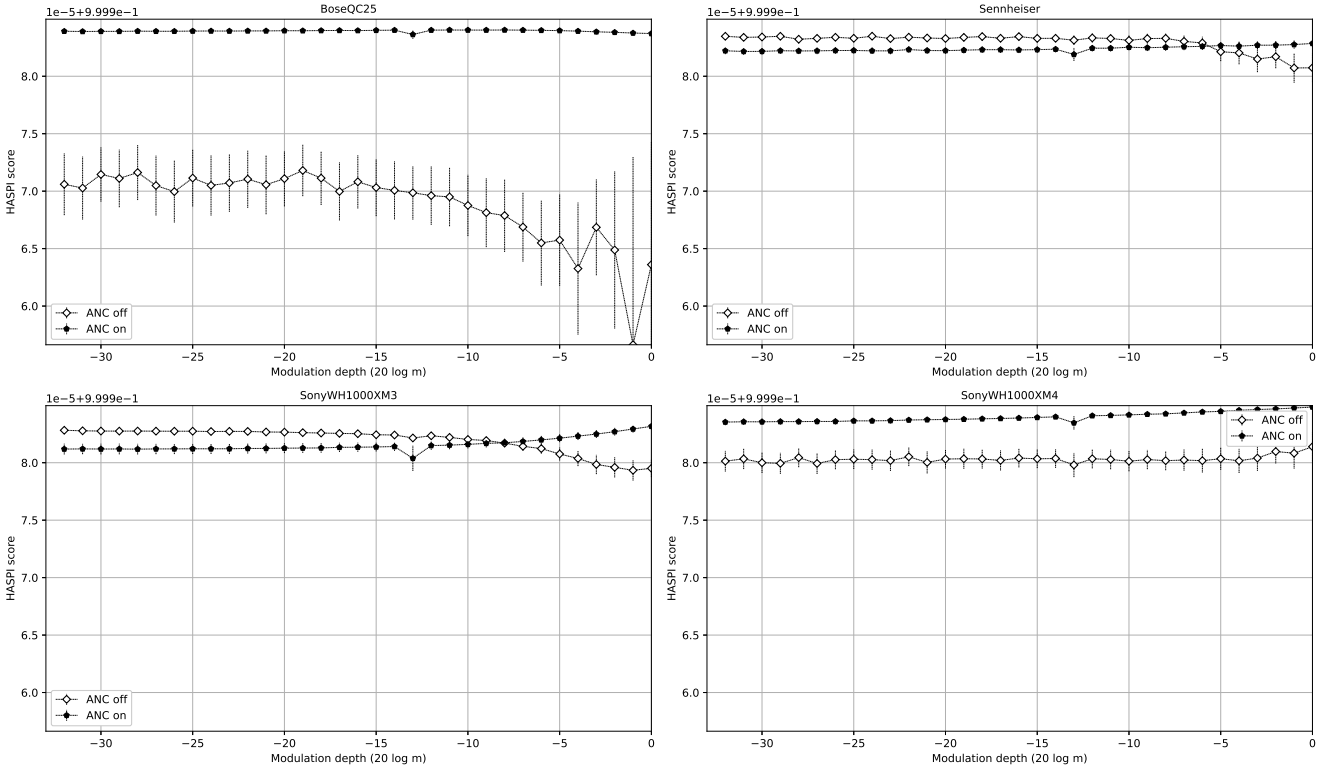



(a) NOIZEUS dataset

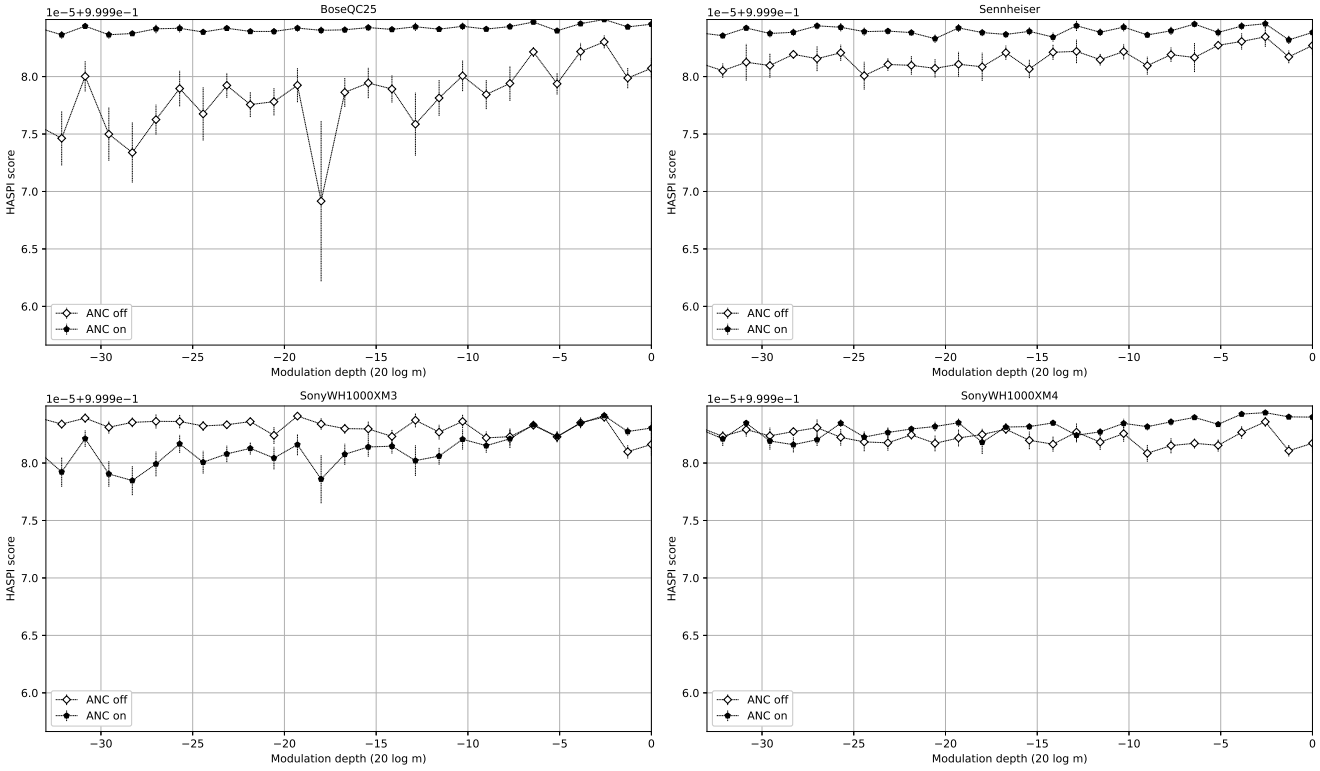



(b) CRM dataset (stimuli used in chapter 4)

Figure 5.3: Chimera audio quality with GPSMq against modulation depth

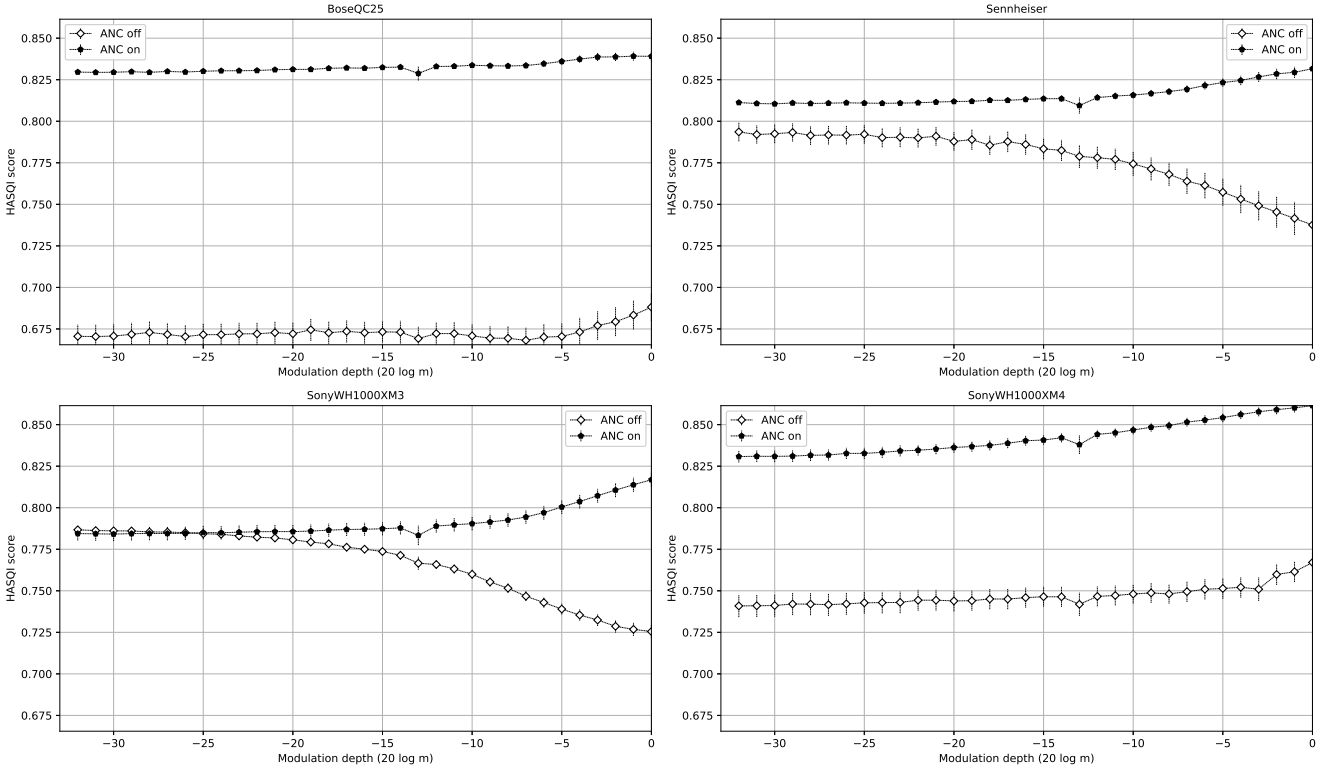



(a) NOIZEUS dataset

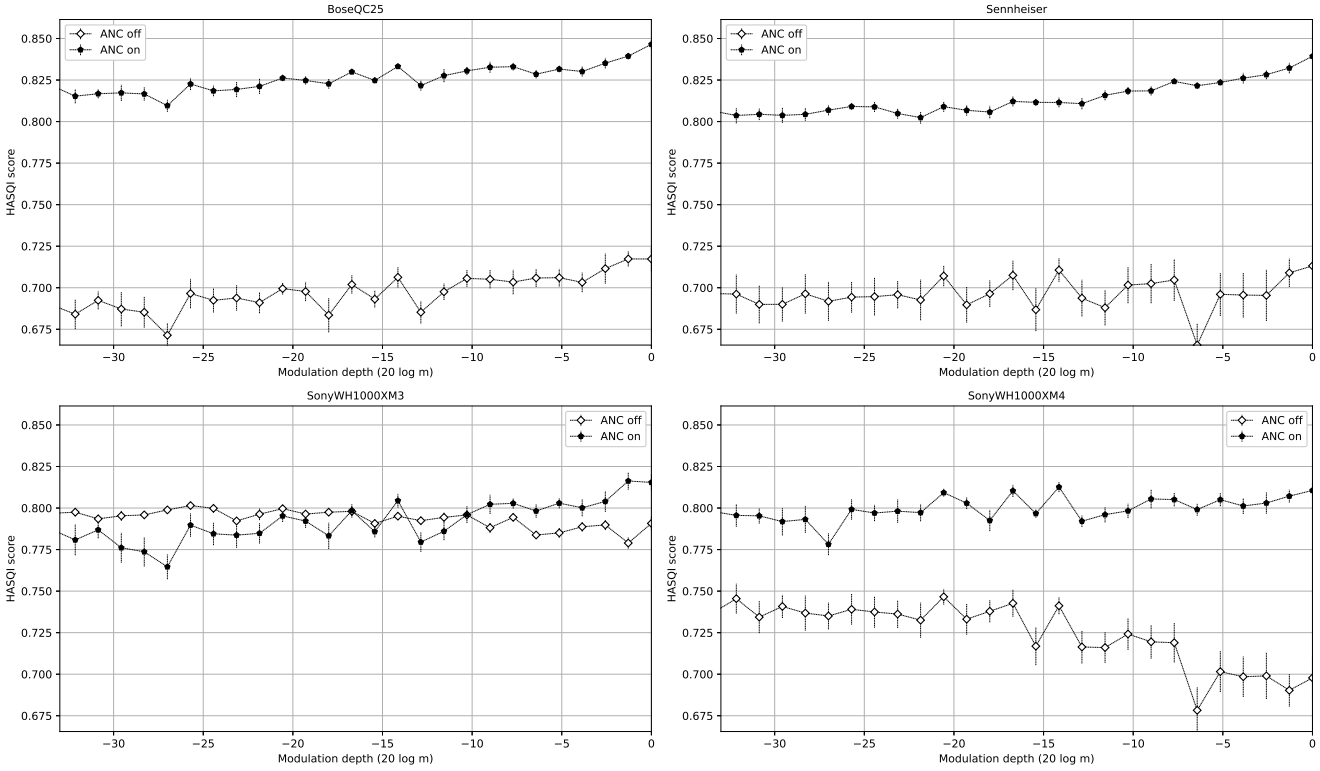



(b) CRM dataset (stimuli used in chapter 4)

Figure 5.4: Chimera audio quality with HASPI against modulation depth

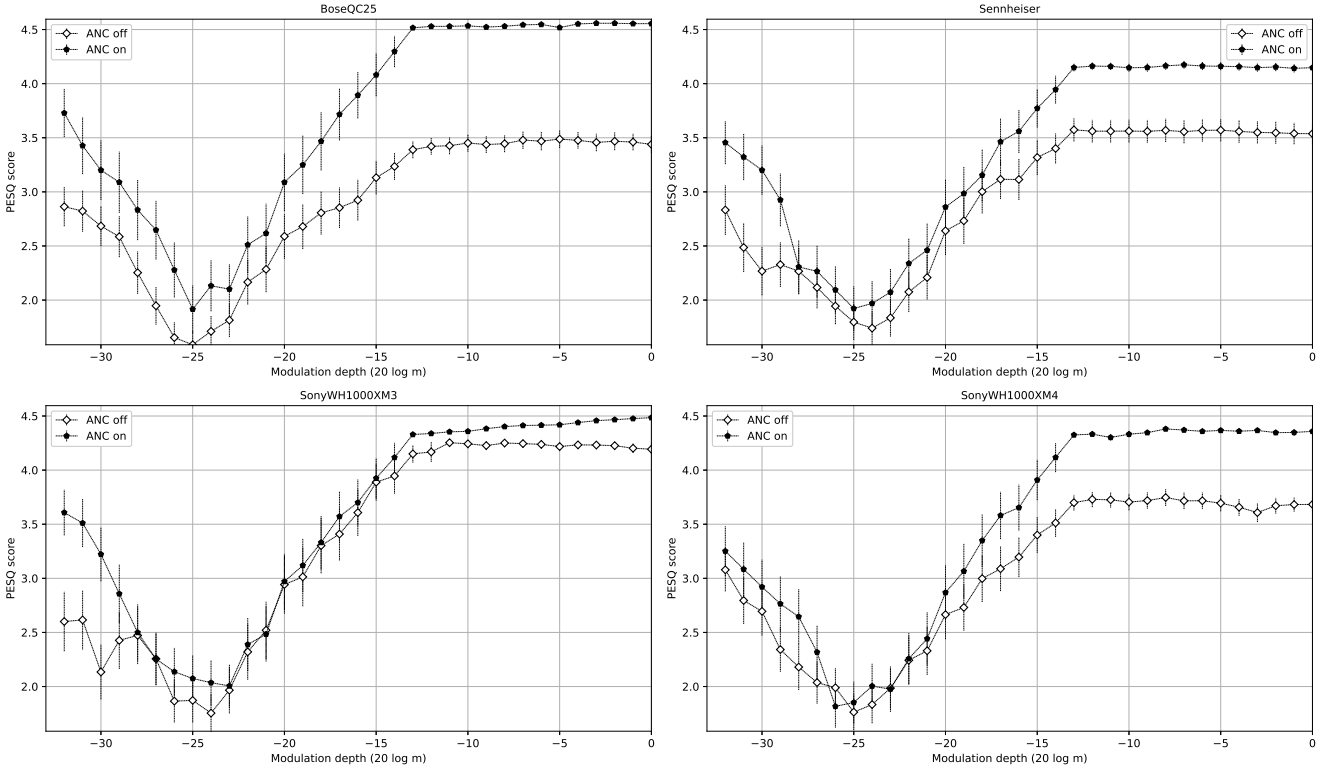



(a) NOIZEUS dataset

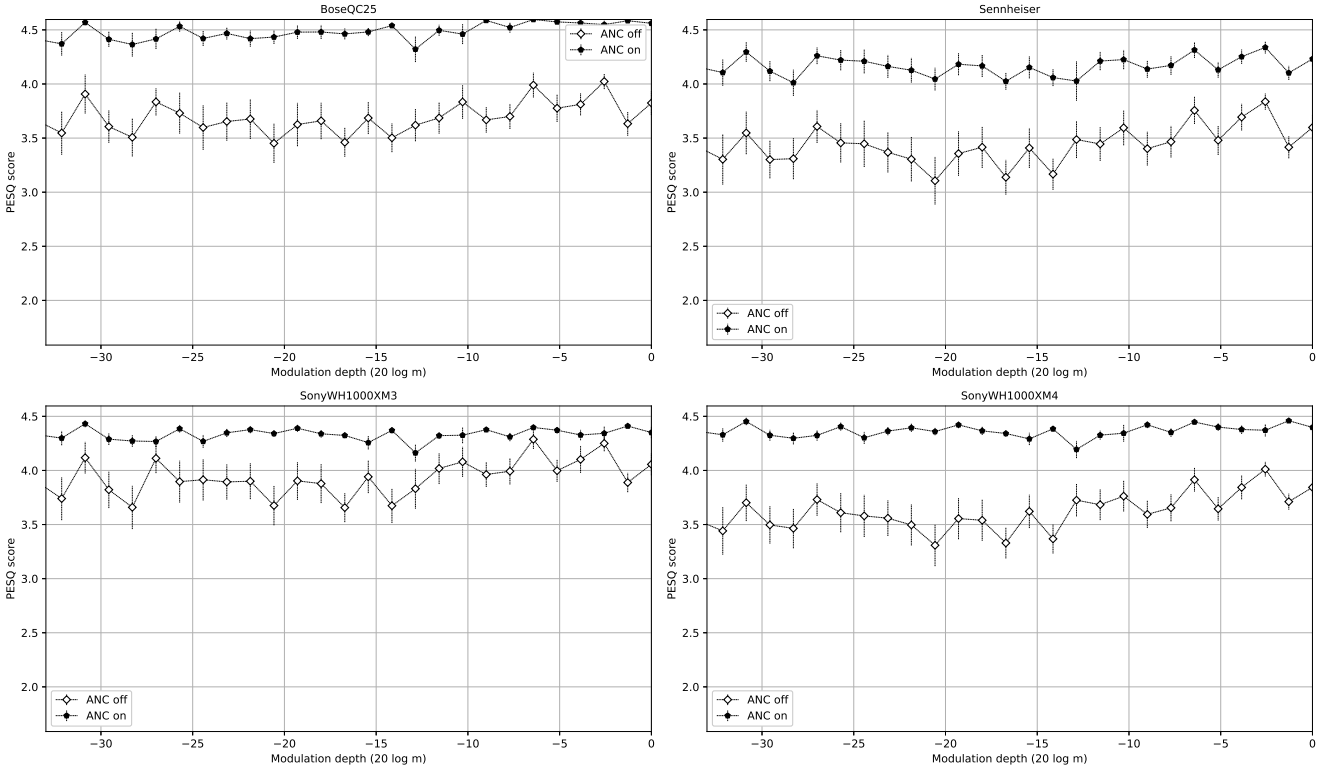



(b) CRM dataset (stimuli used in chapter 4)

Figure 5.5: Chimera audio quality with HASQI against modulation depth

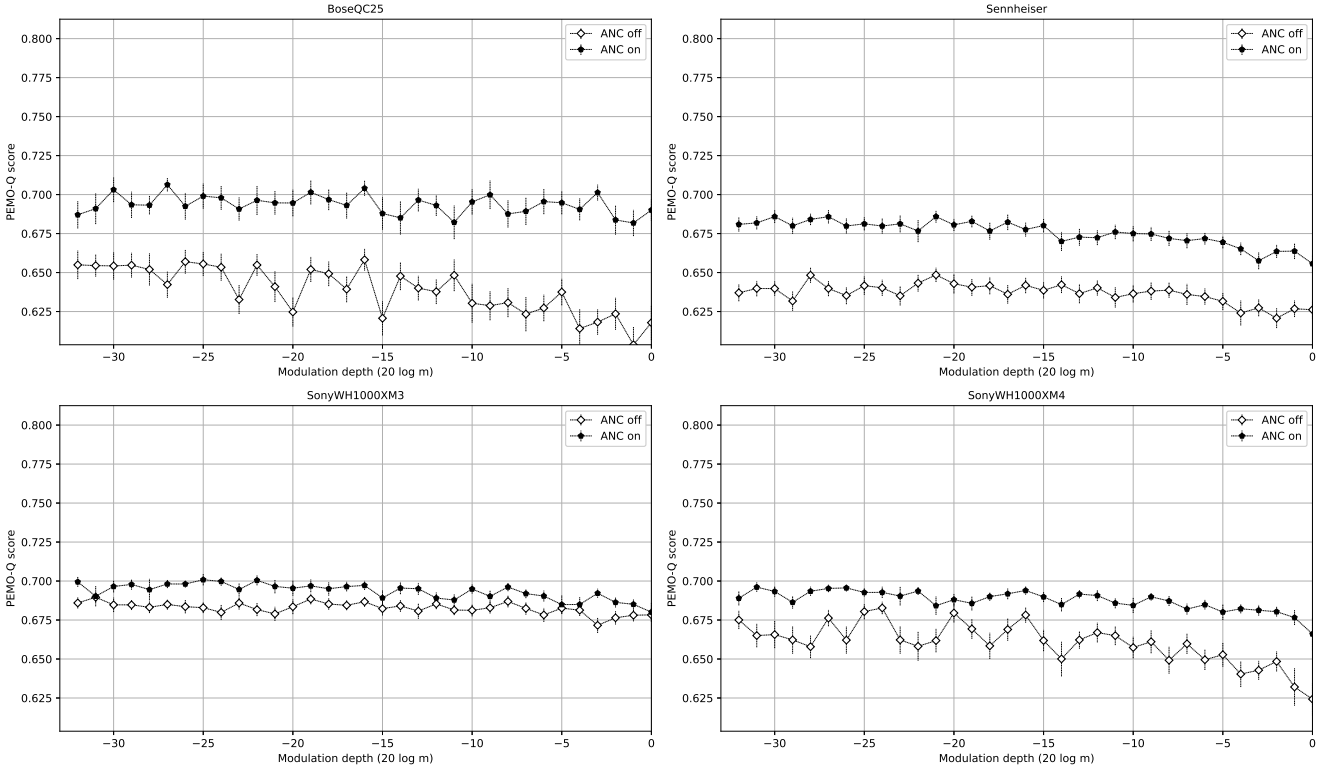



(a) NOIZEUS dataset

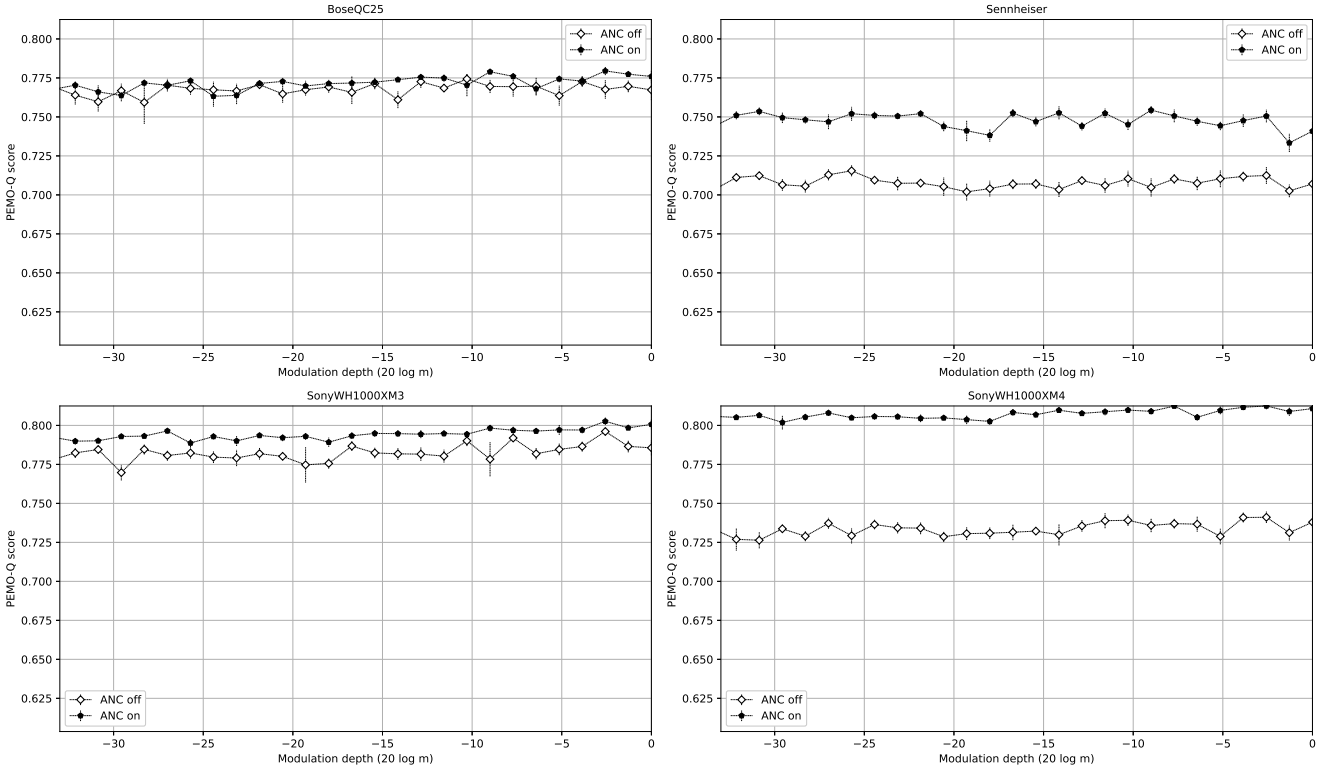



(b) CRM dataset (stimuli used in chapter 4)

Figure 5.6: Chimera audio quality with PESQ against modulation depth

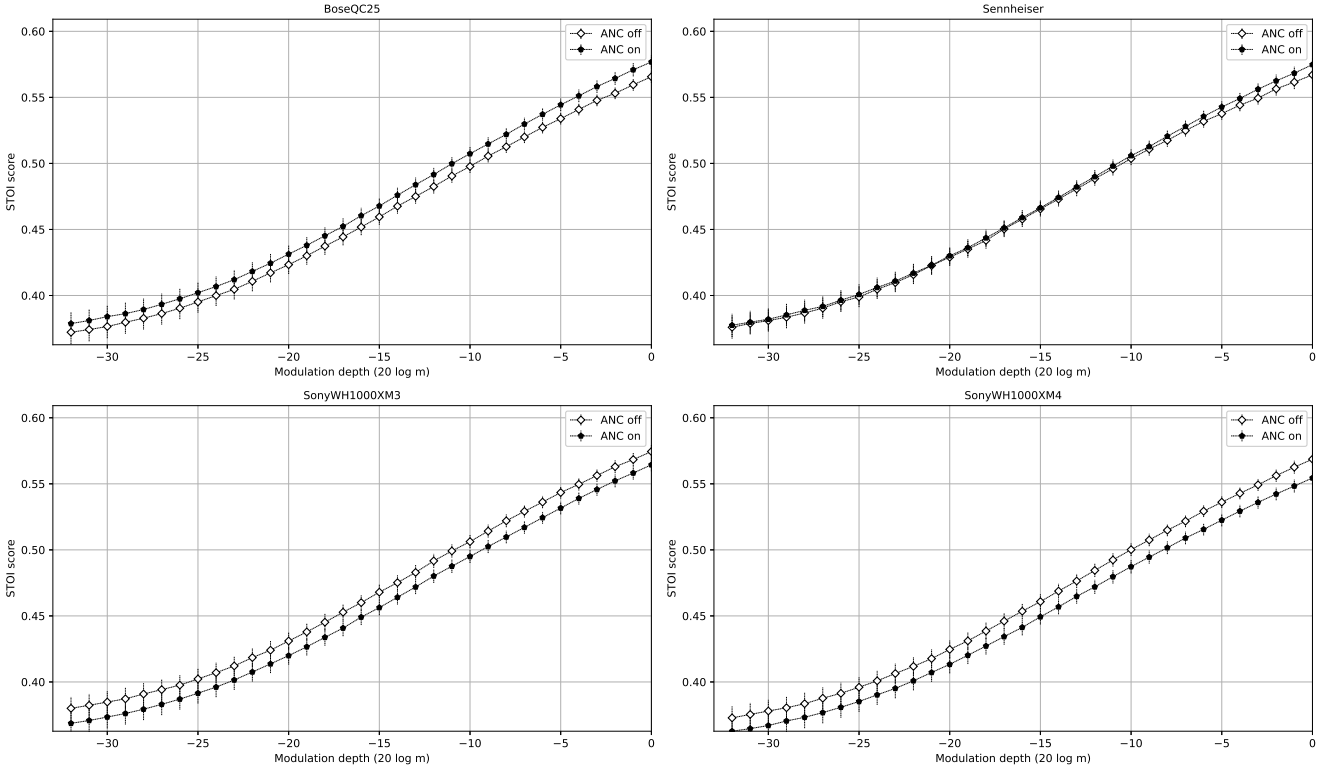



(a) NOIZEUS dataset

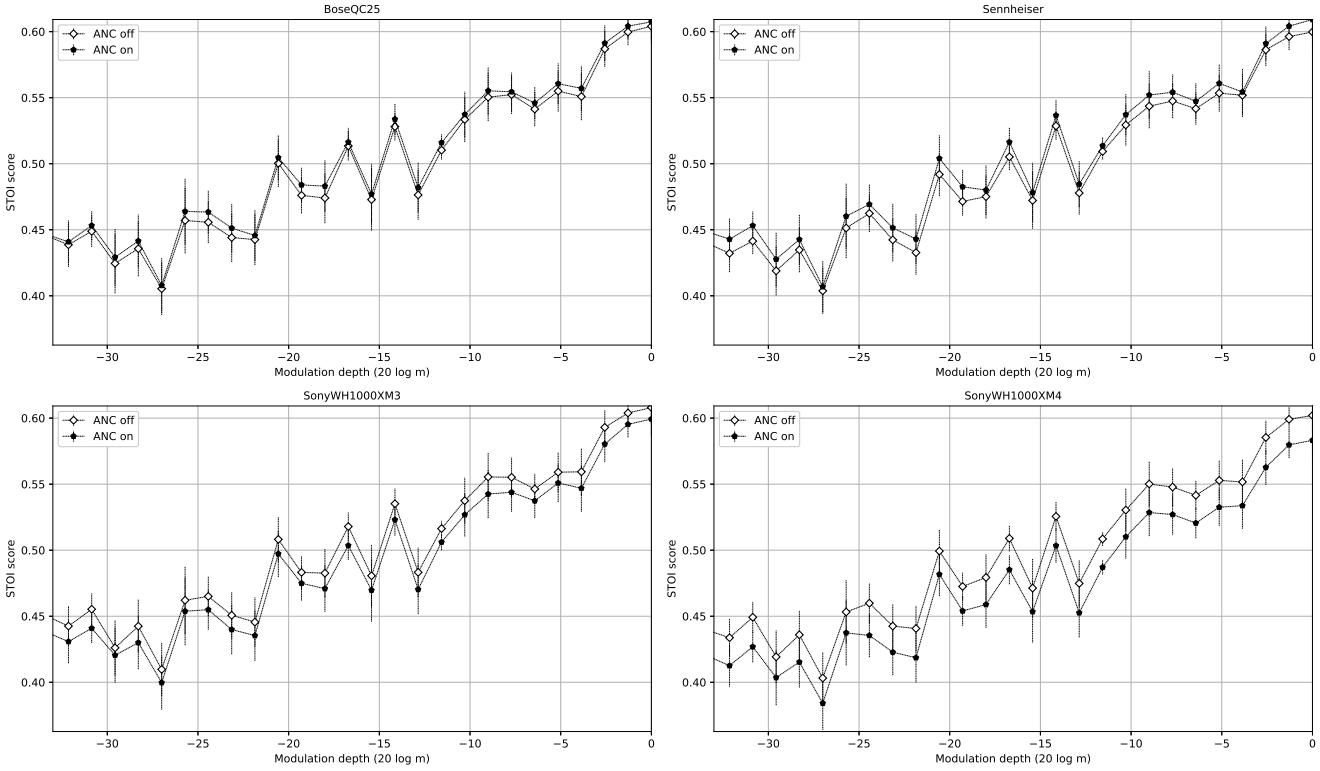



(b) CRM dataset (stimuli used in chapter 4)

Figure 5.7: Chimera audio quality with PEMO against modulation depth

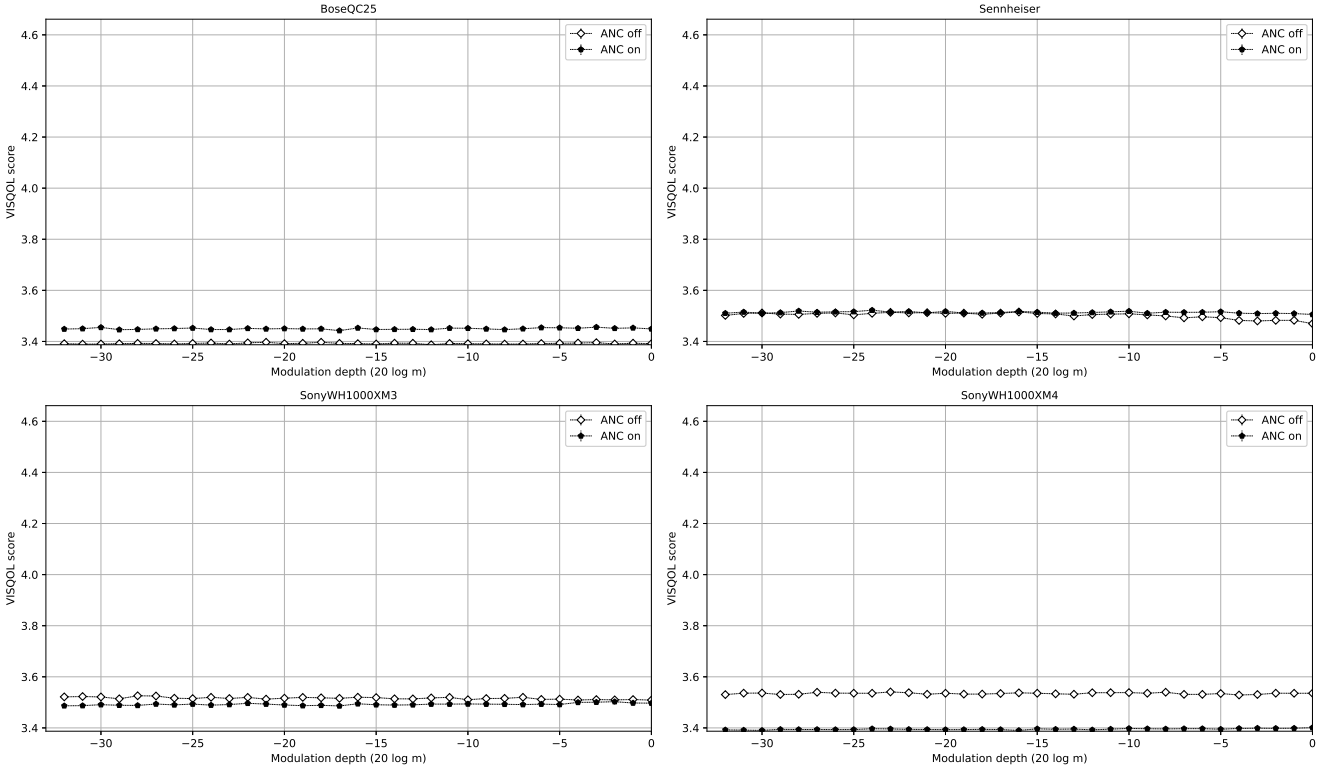



(a) NOIZEUS dataset

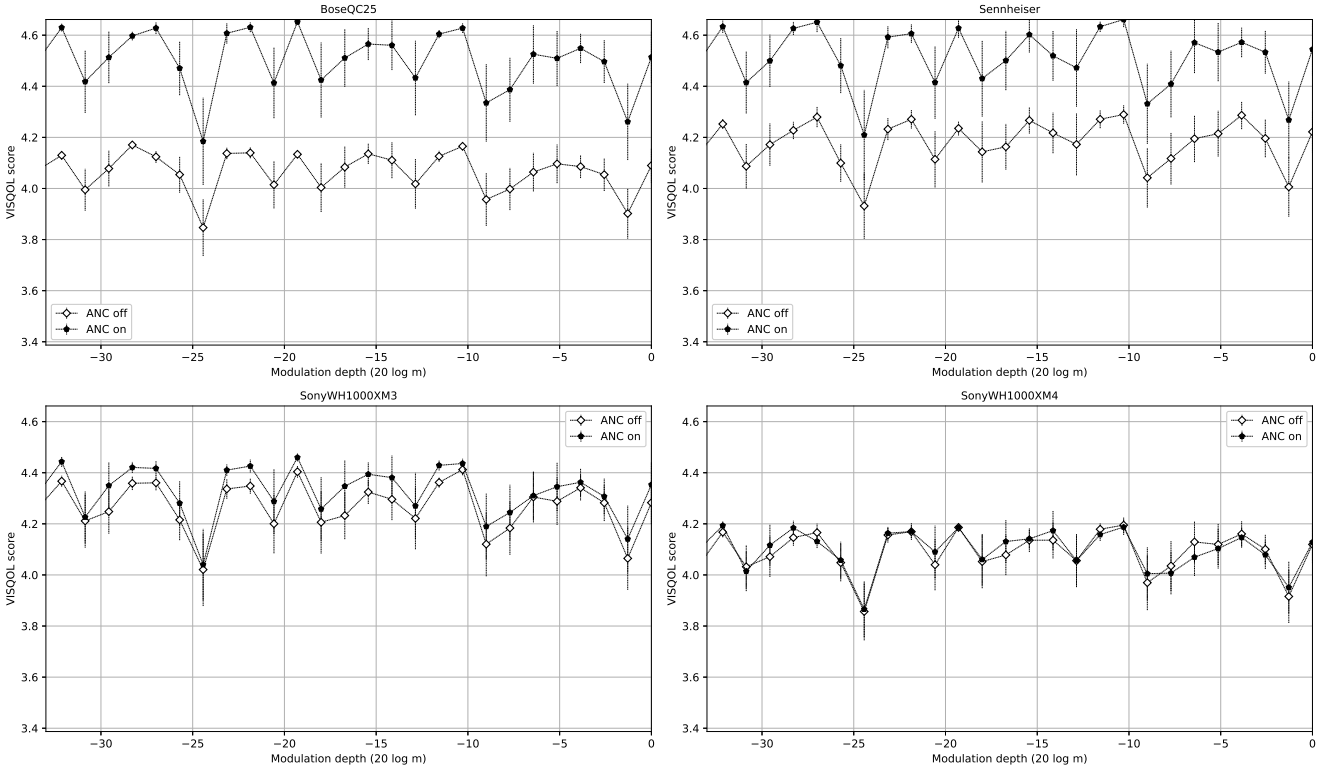



(b) CRM dataset (stimuli used in chapter 4)

Figure 5.8: Chimera audio quality with STOI against modulation depth

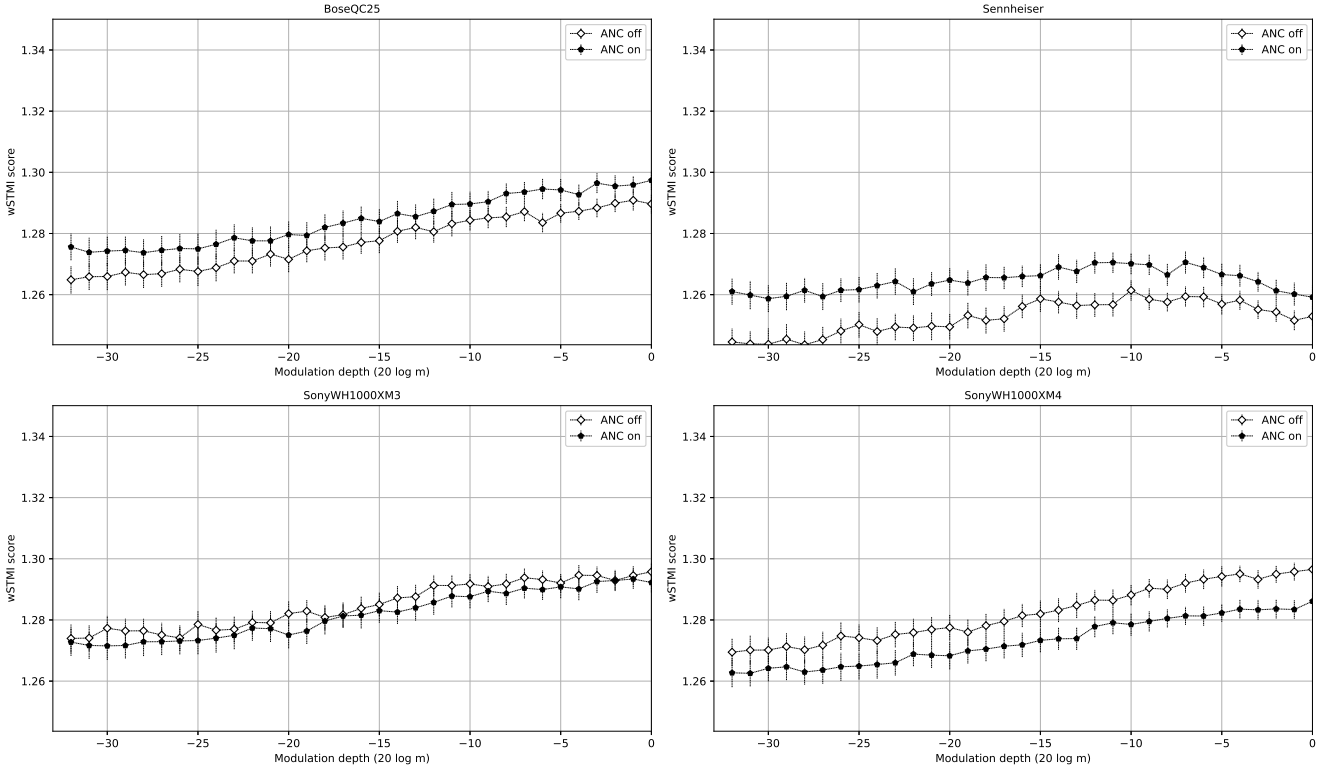



(a) NOIZEUS dataset

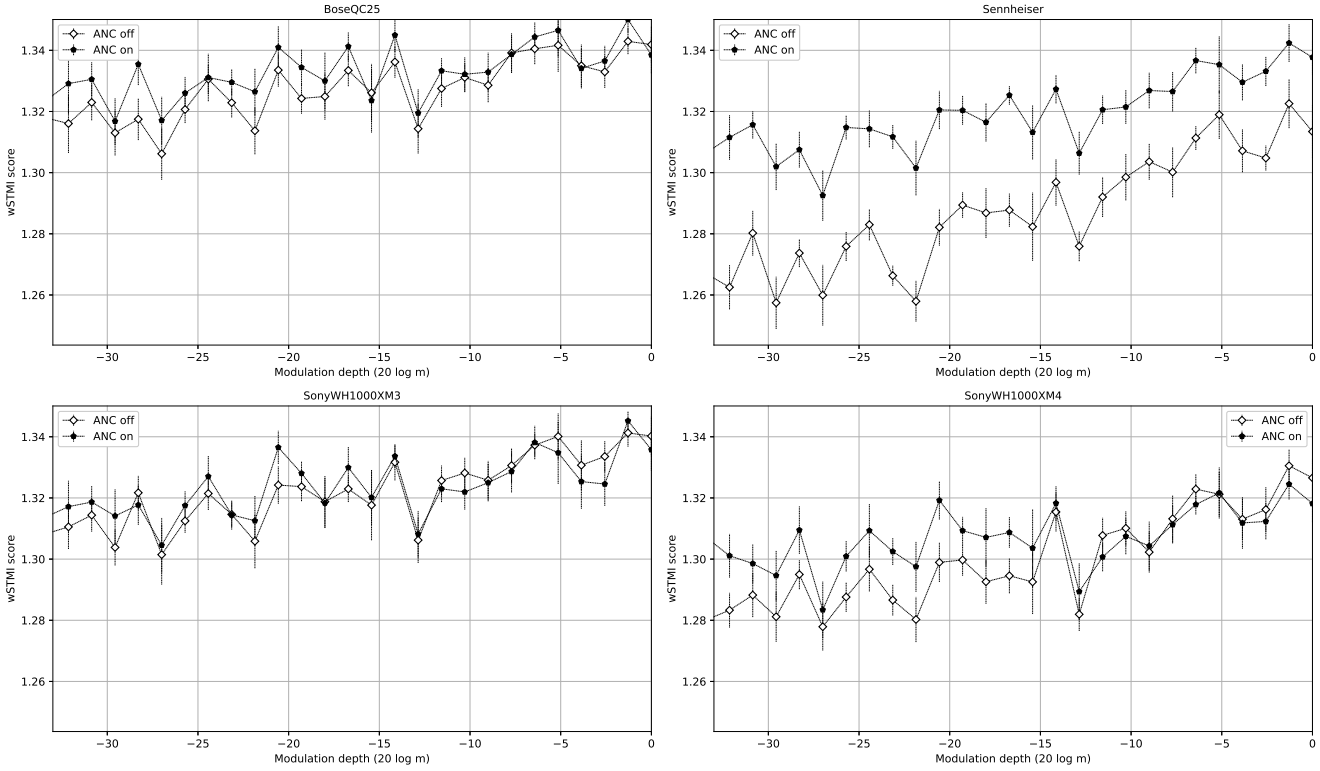



(b) CRM dataset (stimuli used in chapter 4)

Figure 5.9: Chimera audio quality with VISQOL against modulation depth

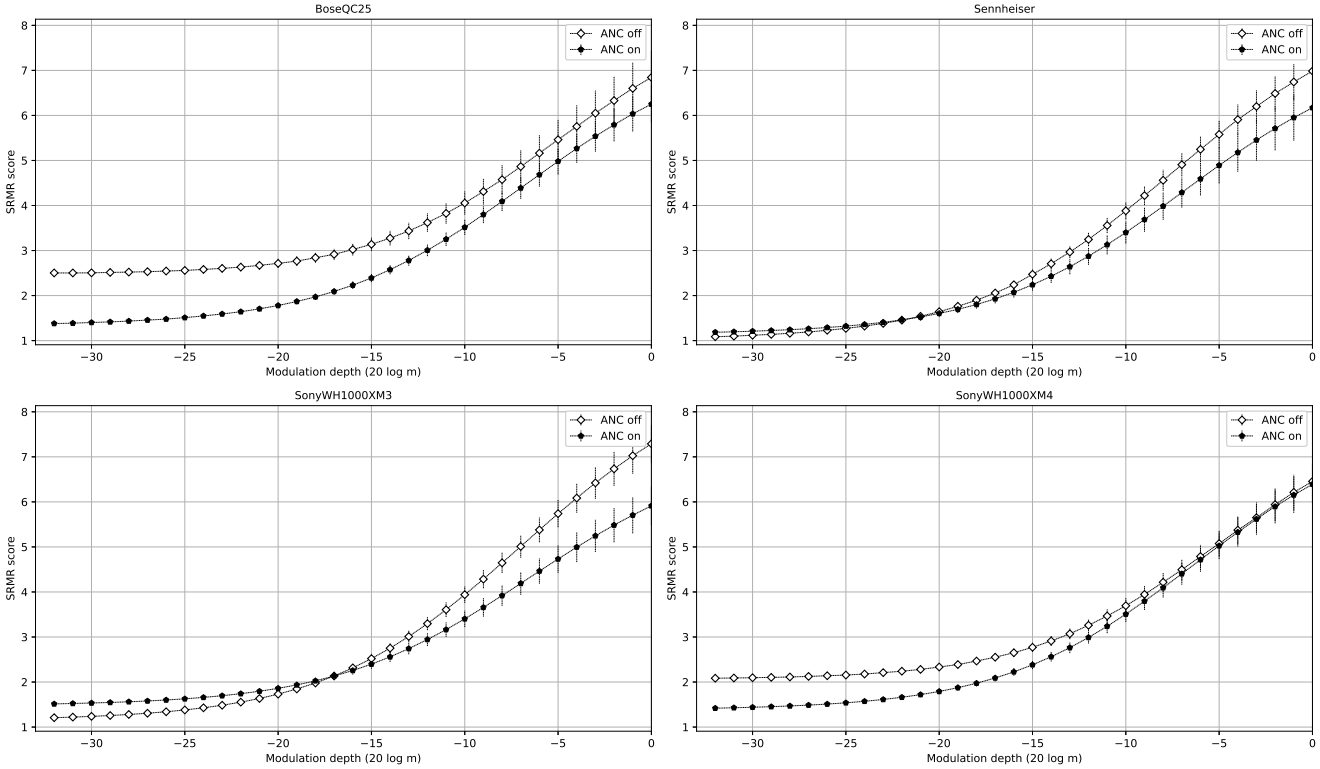



(a) NOIZEUS dataset

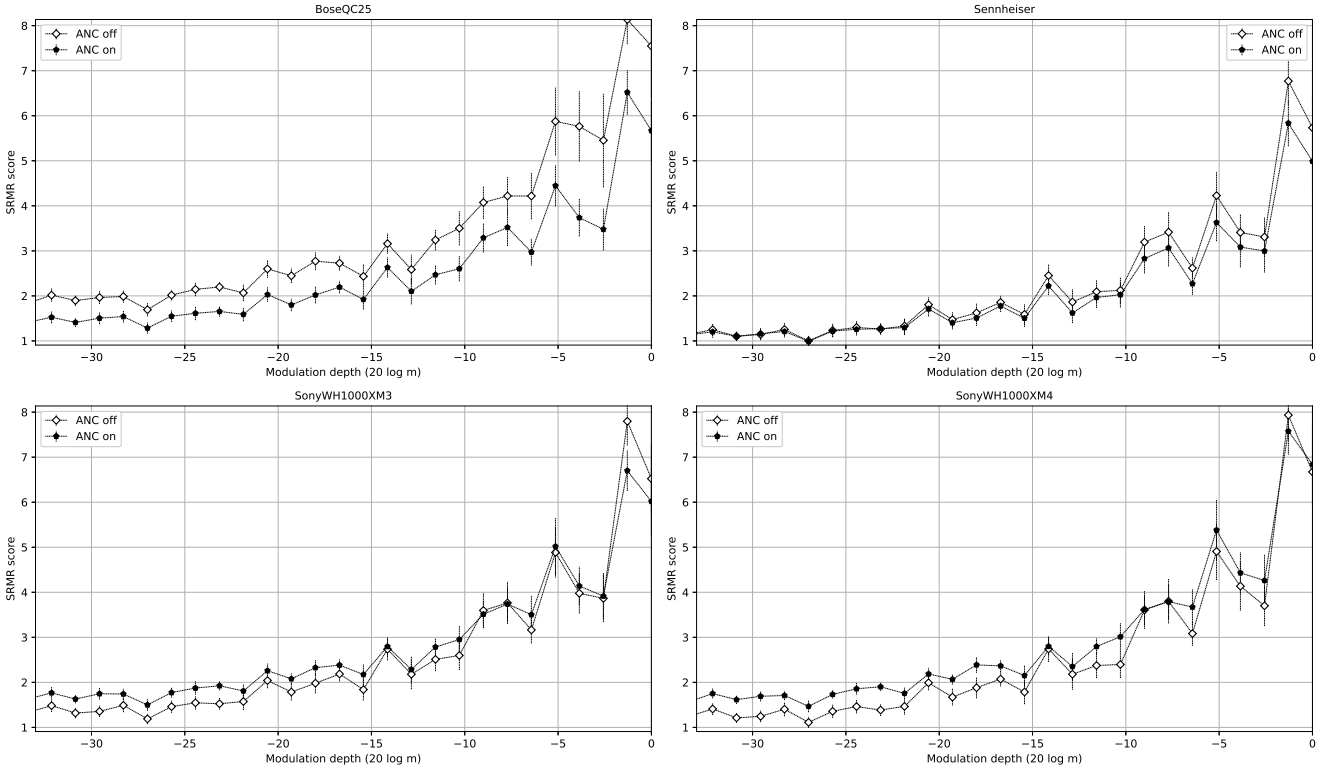



(b) CRM dataset (stimuli used in chapter 4)

Figure 5.10: Chimera audio quality with wSTMI against modulation depth




(a) NOIZEUS dataset




(b) CRM dataset (stimuli used in chapter 4)

Figure 5.11: Chimera audio quality with SRMR against modulation depth

Table 5.1: Statistical comparison (two-way ANOVAs) of objective audio/speech metrics for CRM dataset

| Metric | factor | F-value | p-value | Effect size ($\eta_p^2$) |
|---|---|---|---|---|
| FES | Headphone Type | F(3,2872)=40.35 | **<0.001** | 0.04 |
| | ANC condition | F(1,2872)=26.77 | **<0.001** | 0.01 |
| | Interaction | F(3,2872)=146.3 | **<0.001** | 0.13 |
| GPSMq | Headphone Type | F(3,2872)=3456 | **<0.001** | 0.78 |
| | ANC condition | F(1,2872)=24582 | **<0.001** | 0.90 |
| | Interaction | F(3,2872)=2934 | **<0.001** | 0.75 |
| HASPI | Headphone Type | F(3,2872)=43.45 | **<0.001** | 0.04 |
| | ANC condition | F(1,2872)=191.7 | **<0.001** | 0.06 |
| | Interaction | F(3,2872)=205.8 | **<0.001** | 0.18 |
| HASQI | Headphone Type | F(3,2872)=371.2 | **<0.001** | 0.28 |
| | ANC condition | F(1,2872)=7837 | **<0.001** | 0.73 |
| | Interaction | F(3,2872)=1222 | **<0.001** | 0.56 |
| PESQ | Headphone Type | F(3,2872)=92.38 | **<0.001** | 0.09 |
| | ANC condition | F(1,2872)=2074 | **<0.001** | 0.42 |
| | Interaction | F(3,2872)=32.39 | **<0.001** | 0.03 |
| PEMO-Q | Headphone Type | F(3,2872)=2976 | **<0.001** | 0.76 |
| | ANC condition | F(1,2872)=4940 | **<0.001** | 0.63 |
| | Interaction | F(3,2872)=1137 | **<0.001** | 0.54 |
| STOI | Headphone Type | F(3,2872)=5.029 | **0.002** | 0.005 |
| | ANC condition | F(1,2872)=2.216 | 0.137 | 0.001 |
| | Interaction | F(3,2872)=5.129 | **0.002** | 0.005 |
| VISQOL | Headphone Type | F(3,2872)=125.3 | **<0.001** | 0.12 |
| | ANC condition | F(1,2872)=468.3 | **<0.001** | 0.14 |
| | Interaction | F(3,2872)=111.5 | **<0.001** | 0.10 |
| wSTMI | Headphone Type | F(3,2872)=235.7 | **<0.001** | 0.20 |
| | ANC condition | F(1,2872)=210.7 | **<0.001** | 0.07 |
| | Interaction | F(3,2872)=59.10 | **<0.001** | 0.06 |
| SRMR | Headphone Type | F(3,2872)=23.80 | **<0.001** | 0.02 |
| | ANC condition | F(1,2872)=2.029 | 0.154 | 0.001 |
| | Interaction | F(3,2872)=14.54 | **<0.001** | 0.01 |

Table 5.2: Statistical comparison (two-way ANOVAs) of objective audio/speech metrics for NOIZEUS dataset

| Metric | factor | F-value | p-value | Effect size ($\eta_p^2$) |
|---|---|---|---|---|
| FES | Headphone Type | F(3,7912)=307.3 | **<0.001** | 0.10 |
| | ANC condition | F(1,7912)=772.7 | **<0.001** | 0.09 |
| | Interaction | F(3,7912)=728.3 | **<0.001** | 0.22 |
| GPSMq | Headphone Type | F(3,7912)=29767 | **<0.001** | 0.92 |
| | ANC condition | F(1,7912)=65599 | **<0.001** | 0.89 |
| | Interaction | F(3,7912)=13704 | **<0.001** | 0.84 |
| HASPI | Headphone Type | F(3,7912)=205.6 | **<0.001** | 0.07 |
| | ANC condition | F(1,7912)=472.5 | **<0.001** | 0.06 |
| | Interaction | F(3,7912)=325.1 | **<0.001** | 0.11 |
| HASQI | Headphone Type | F(3,7912)=1037 | **<0.001** | 0.28 |
| | ANC condition | F(1,7912)=15969 | **<0.001** | 0.67 |
| | Interaction | F(3,7912)=2407 | **<0.001** | 0.48 |
| PESQ | Headphone Type | F(3,7912)=17.62 | **<0.001** | 0.006 |
| | ANC condition | F(1,7912)=311.2 | **<0.001** | 0.04 |
| | Interaction | F(3,7912)=19.82 | **<0.001** | 0.007 |
| PEMO-Q | Headphone Type | F(3,7912)=294.6 | **<0.001** | 0.10 |
| | ANC condition | F(1,7912)=1769 | **<0.001** | 0.18 |
| | Interaction | F(3,7912)=141.5 | **<0.001** | 0.05 |
| STOI | Headphone Type | F(3,7912)=7.749 | **<0.001** | 0.002 |
| | ANC condition | F(1,7912)=3.545 | 0.0597 | 0.000 |
| | Interaction | F(3,7912)=9.922 | **<0.001** | 0.003 |
| VISQOL | Headphone Type | F(3,7912)=1621 | **<0.001** | 0.38 |
| | ANC condition | F(1,7912)=548.4 | **<0.001** | 0.06 |
| | Interaction | F(3,7912)=1735 | **<0.001** | 0.40 |
| wSTMI | Headphone Type | F(3,7912)=501.5 | **<0.001** | 0.16 |
| | ANC condition | F(1,7912)=11.16 | **<0.001** | 0.001 |
| | Interaction | F(3,7912)=88.86 | **<0.001** | 0.03 |
| SRMR | Headphone Type | F(3,7912)=15.04 | **<0.001** | 0.006 |
| | ANC condition | F(1,7912)=91.05 | **<0.001** | 0.01 |
| | Interaction | F(3,7912)=7.609 | **<0.001** | 0.003 |

Table 5.3: Mean and standard deviation of objective audio/speech metrics across headphones and ANC conditions for CRM dataset

| Metric | ANC | QC25 | | HD 4.50 | | WH-1000XM3 | | WH-1000XM4 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Stdev | Mean | Stdev | Mean | Stdev | Mean | Stdev |
| FES ↑ | On | 0.599586 | 0.2919 | 0.897744 | 0.1622 | 0.86719 | 0.1904 | 0.84313 | 0.2165 |
| | Off | 0.916114 | 0.1542 | 0.798805 | 0.222 | 0.876646 | 0.1852 | 0.778186 | 0.2264 |
| GPSMq ↓ | On | 14.102 | 0.3041 | 15.5524 | 0.2569 | 13.9097 | 0.4558 | 11.801 | 0.7378 |
| | Off | 16.224 | 0.3606 | 16.2596 | 0.1878 | 15.5585 | 0.1336 | 16.0817 | 0.1723 |
| HASPI ↑ | On | 0.999984 | 9.043e-07 | 0.999984 | 1.043e-06 | 0.999981 | 3.641e-06 | 0.999983 | 2.078e-06 |
| | Off | 0.999977 | 8.069e-06 | 0.999981 | 3.93e-06 | 0.999983 | 1.723e-06 | 0.999982 | 1.961e-06 |
| HASQI ↑ | On | 0.823708 | 0.01314 | 0.812456 | 0.01314 | 0.78887 | 0.0234 | 0.798707 | 0.01884 |
| | Off | 0.694643 | 0.02574 | 0.696349 | 0.03473 | 0.79486 | 0.009383 | 0.728793 | 0.03441 |
| PESQ ↑ | On | 4.47063 | 0.2158 | 4.15273 | 0.3141 | 4.32527 | 0.1476 | 4.3564 | 0.146 |
| | Off | 3.65829 | 0.5203 | 3.40241 | 0.5666 | 3.8885 | 0.5255 | 3.57925 | 0.5304 |
| PEMO-Q ↑ | On | 0.771235 | 0.01119 | 0.748036 | 0.01201 | 0.793386 | 0.007383 | 0.806038 | 0.007593 |
| | Off | 0.766759 | 0.01665 | 0.707815 | 0.01308 | 0.781366 | 0.01478 | 0.732993 | 0.01337 |
| STOI ↑ | On | 0.484677 | 0.07819 | 0.484556 | 0.07735 | 0.474971 | 0.07804 | 0.458453 | 0.07695 |
| | Off | 0.480013 | 0.07862 | 0.476254 | 0.07527 | 0.485547 | 0.07777 | 0.477994 | 0.07611 |
| VISQOL ↑ | On | 4.50657 | 0.3144 | 4.51444 | 0.3234 | 4.33159 | 0.272 | 4.10251 | 0.2217 |
| | Off | 4.06979 | 0.2147 | 4.17368 | 0.2536 | 4.27034 | 0.2632 | 4.08972 | 0.2264 |
| wSTMI ↑ | On | 1.33085 | 0.02344 | 1.31602 | 0.02348 | 1.32144 | 0.02415 | 1.30563 | 0.02445 |
| | Off | 1.3242 | 0.02315 | 1.2816 | 0.02958 | 1.31865 | 0.02367 | 1.29557 | 0.02664 |
| SRMR ↑ | On | 2.2644 | 1.448 | 1.8493 | 1.348 | 2.52716 | 1.509 | 2.59214 | 1.723 |
| | Off | 2.98964 | 2.005 | 1.99777 | 1.551 | 2.33069 | 1.738 | 2.26597 | 1.805 |

Table 5.4: Mean and standard deviation of objective audio/speech metrics across headphones and ANC conditions for NOIZEUS dataset

| Metric | ANC | QC25 | | HD 4.50 | | WH-1000XM3 | | WH-1000XM4 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Stdev | Mean | Stdev | Mean | Stdev | Mean | Stdev |
| FES ↑ | On | 0.480362 | 0.3023 | 0.827032 | 0.2093 | 0.433152 | 0.2948 | 0.740288 | 0.2558 |
| | Off | 0.9168 | 0.1278 | 0.569264 | 0.2966 | 0.720975 | 0.2609 | 0.88869 | 0.1564 |
| GPSMq ↓ | On | 16.9935 | 0.2186 | 17.3301 | 0.07457 | 14.5367 | 0.2282 | 14.2958 | 0.3426 |
| | Off | 17.5405 | 0.1245 | 17.4056 | 0.273 | 17.1797 | 0.3134 | 16.7115 | 0.2756 |
| HASPI ↑ | On | 0.999984 | 8.917e-07 | 0.999982 | 1.664e-06 | 0.999982 | 2.43e-06 | 0.999984 | 1.295e-06 |
| | Off | 0.999969 | 2.468e-05 | 0.999983 | 3.18e-06 | 0.999982 | 2.37e-06 | 0.99998 | 5.031e-06 |
| HASQI ↑ | On | 0.832594 | 0.009833 | 0.815606 | 0.01327 | 0.791066 | 0.02411 | 0.842031 | 0.01953 |
| | Off | 0.672843 | 0.03919 | 0.777549 | 0.04056 | 0.766383 | 0.02521 | 0.746934 | 0.03465 |
| PESQ ↑ | On | 3.69156 | 1.354 | 3.40157 | 1.217 | 3.562 | 1.292 | 3.47786 | 1.304 |
| | Off | 2.87566 | 1.003 | 2.94045 | 1.095 | 3.33317 | 1.348 | 3.04896 | 1.096 |
| PEMO-Q ↑ | On | 0.693624 | 0.04417 | 0.675561 | 0.02699 | 0.693385 | 0.02246 | 0.687574 | 0.02289 |
| | Off | 0.638797 | 0.05197 | 0.636824 | 0.03112 | 0.682681 | 0.02177 | 0.660234 | 0.04289 |
| STOI ↑ | On | 0.466577 | 0.07268 | 0.465047 | 0.07206 | 0.455122 | 0.07197 | 0.447705 | 0.07091 |
| | Off | 0.458048 | 0.07175 | 0.4626 | 0.07085 | 0.466301 | 0.07182 | 0.459644 | 0.07178 |
| VISQOL ↑ | On | 3.44976 | 0.0415 | 3.51363 | 0.04995 | 3.49221 | 0.04755 | 3.39505 | 0.04296 |
| | Off | 3.39191 | 0.03428 | 3.50253 | 0.04539 | 3.51621 | 0.04615 | 3.53478 | 0.05033 |
| wSTMI ↑ | On | 1.28431 | 0.02376 | 1.26448 | 0.02143 | 1.28177 | 0.02316 | 1.27311 | 0.02203 |
| | Off | 1.27738 | 0.0224 | 1.25258 | 0.02205 | 1.28477 | 0.0215 | 1.2822 | 0.02268 |
| SRMR ↑ | On | 2.91614 | 1.872 | 2.76982 | 2.104 | 2.8819 | 1.766 | 2.94318 | 1.897 |
| | Off | 3.70375 | 2.027 | 3.03811 | 2.209 | 3.15373 | 2.237 | 3.31902 | 1.707 |

## 5.4 Chapter summary

This chapter characterized ANC headphones in terms of certain audio or speech quality and intelligibility metrics. This was done by evaluating dummy-head recordings of parametrically degraded-envelope speech signals (chimeras), and comparing the results for 'ANC on' and 'ANC off' conditions for the four commercially available headphones. It led us to discover some strengths and weaknesses of both the audio metrics as well as the particular chimera algorithm in carrying out such a characterization. The objective of the proposed characterization of ANC headphones with the help of these models is not to use the exact values in the outputs, but instead to observe and analyze numerically backed macro-level trends in the data. With increasing bandwidth of speech being transmitted in modern audio applications (such as VoIP call applications), the boundaries between traditional speech quality estimation and audio quality estimation have blurred (Moeller & Heusdens, 2013). Moreover, the precise nature of interaction between various pieces of the technology in hearables are going to become more significant with time. This chapter lays the foundation of some of these analyses that may be conducted on similar devices towards the goal of understanding the perceptual effect of the various algorithms and also perhaps the auditory processing of the brain itself.

The results of this chapter showed that the all the metrics do not agree much, even at the level of broad trends across the ANC conditions or the headphones. This could be due to inherent differences between the metrics, or lack of fine-tuning it for the present type of degradation (shallow envelope chimera). Hence, it is hard to rely on any one of them for fully characterizing ANC headphones. Indeed, this is a common problem. Perceptual-based instrumental models can be shown to perform well for some degradation types, but the performance rapidly degrades for other types of degradation or novel types of signal (e.g a new language) (Akhtar et al., 2019). Thus, any effort towards building an 'ANC index' (similar to HASPI etc.) needs to take these aspects into account.

### 5.4.1 Future work

The work reported in this chapter can be extended in the following ways:

- Given that some of the signals used in chapter 4 were used here, it would be interesting to correlate the metric scores with the results obtained in section 4.3. This kind of work is sometimes used to build a new speech intelligibility metric by combining aspects of two or more existing metrics, or

to explain previously unexplained aspects of the envelope perception (Relaño-Iborra et al., 2019). However, since our goal is more modest; we are simply interested in characterizing the headphones in terms of modulation perception and compare it across ANC conditions.

- Instrumental audio metrics are sometimes combined into another super-metric. For ANC headphones, an 'ANC index' could be created by carefully studying the aspects of a playback signal that are desirable from an ANC headphone. For instance, MoBI-Q proposed by Biberger et al. (2021) is essentially a fusion of two earlier metrics by the same research lab. In turn, MoBI-Q, GPSMq and others (like PEMO-Q) have been based on a 'family' of auditory perception models going back to the original modulation filterbank concept (Dau et al., 1997a, 1997b). The precise nature of the interaction between the headphone algorithm, instrumental metric and the speech envelope can be utilized at a more finer detail, and for the purposes of building an envelope-based assessment protocol or metric for the 'ANC index', these findings can serve as foundational ideas.

- Since the advent of machine learning and particularly deep neural networks, audio researchers have also naturally found all sorts of applications of these tools in their work. One such direction has been a trend to copy the full-reference metrics into no-reference mode, with the reasoning that it would make the metric more widely applicable albeit at the cost of reducing accuracy or reliability. One such model called 'WAWEnets' supposedly produced a trained neural network which is supposedly able to reproduce PEMO-Q, STOI, and PESQ in no-reference mode (Catellier & Voran, 2020). It is possible that such models, especially if they are domain-neutral, could be the future of instrumental audio and speech quality/intelligibility predictions and as such would be interesting to evaluate ANC systems with these.

- It is common for audio signal processing researchers to use datasets like NOIZEUS for evaluating performance of speech processing systems, such as hearing aid algorithms. It is possible that some specific quirks of the instrumental metrics used here and elsewhere arise from all the algorithms using the same signals to test their outcomes. In this context, it is extremely vital to cross-check or validate results (such as those from this chapter) using multiple datasets. It would be interesting to compare multiple speech datasets (e.g. with differences in spoken accent) with the exact same chimera parameters to see whether the ANC effect is different in those cases.

# 6

# Conclusions and Discussions

The objective of this thesis was to elucidate the properties of ANC headphones in terms of their amplitude modulation transmission properties in the playback signal, by measuring physical and psychoacoustic changes of the amplitude envelope of the playback signal. This was done from two angles; the modulation frequency was investigated in chapters 2 and 3, and the modulation depth was investigated in chapters 4 and 5. Various metrics were borrowed or adapted from the literature around acoustics, psychology of hearing, neuroscience and audio processing for these purposes, and the playback signals (recorded with dummy-head microphone) were analyzed with these metrics and compared across 'ANC on' and 'ANC off' conditions. Four commercially available ANC headphones were examined in this way. It was seen that there are generally significant differences caused by the ANC, and these differences are not the same for all the four devices under test. However, it was seen that in some metrics as well as the human performance, the differences caused by ANC were small in magnitude. Taken together, this means that although the ANC system when enabled helps deliver a cleaner envelope in the playback signal, the psychological benefits may be limited and depends on various other factors. Overall, this research was a fruitful initial exploration into an under-explored phenomenon that we hope could yield new research directions in the disciplines of audio signal processing as well as auditory psychology, some of which are outlined in section 6.2.

## 6.1 General comments

ANC technology has come a long way since its inception, and the applications of real-time digital signal processing to a playback signal in the 'hearables' market has started including advanced functions such as 'wake word detection' or different behaviors depending on environment; this makes it increasingly hard to design standardized test protocols (Temme, 2019). Researchers have long since recognized that ANC devices need to be tested with custom evaluation protocols, and standardized methods borrowed from other areas may not be applicable (Wagstaff & Woxen, 1997; Reimes, 2019), and that ANC evaluations need to focus on overall sound quality of their devices instead of just convergence and reduction of decibels (de Oliveira, 2010). Currently, we are in the age of smart audio where researchers and corporate entities are shifting their focus towards holistic and human-oriented acoustic solutions for vehicles, aircrafts, and urban spaces like office buildings, construction sites and so on, where ANC technology can play a significant role (Lee et al., 2019; Sosna, 2019). For instance, Lam et al. (2021) outlined a set of interesting questions regarding proof-of-concept structures and solutions where ANC has to be integrated within a larger set of constraints such as airflow and visual aesthetics. Given the above, it is high time that the ANC technology is integrated more closely with the rest of auditory research (neuroscience and psychology) and we gain a better understanding regarding the features of ANC in terms of psychoacoustic impact. The present work was a step in that direction, in terms of evaluating the changes to the amplitude envelope by ANC headphones.

There is a noticeable lack of communication across disciplines related to acoustics. In disciplines of audio engineering, researchers have long since recognized the need for connecting their work with fundamentals of auditory perception (Cabot, 1997; Priese et al., 2013). For instance, Geddes and Lee (2003) noted in the context of designing accurate distortion metrics that "the relationship of the signal based metrics to subjective impression is virtually nonexistent". Some papers have far-reaching pronouncements like "…objects and machines are built to serve humans and must always be designed with human users in mind. Regarding the ANC technique, it should be noted that the ultimate goal of ANC is to minimize the annoyance of acoustic noise to human hearing. Hence, human hearing sensation is an important factor to be taken into account in the ANC system design" (Bao & Panahi, 2013). Despite the discipline of psychophysics having rich history of dealing with the concept of quantifying human sensation, tools and methods from psychophysics are relatively lacking in the disciplines around audio engineering. State-

of-the-art ANC research still disregard the human hearing system and solely rely on concepts like convergence or performance of the controller (Lu et al., 2021a, 2021b; Pradhan, 2021). When designing standards, recommendations, tests and other tools for assessing audio quality, engineers tend to have a bias towards 'objective' measurements and ignore the subjective aspects and the complexity of the phenomena involved. It is interesting to see statements like "monitoring audio is not only a question of using visual displays. Listening is just as important – at least!" (Brixen, 2020) in cutting-edge research, which indicates a growing awareness that the ear cannot be understood solely in terms of things that the eye perceives. To complicate matters further, human subjective biases can play a significant role in the overall acceptance of an ANC system. For instance, consider Guicking (2007, section 2.9): "Many other car manufacturers develop their own systems, and some of them have successfully built prototypes, but all of them are hesitating to install the ANC systems in series production (e.g., within a "comfort package" at extra cost). One argument is that customers would complain if they pay for noise reduction, and there still remains some disturbing noise." This example highlights the need for perceptual and psychological evaluation of ANC systems in addition to acoustic evaluation.

Modulation is surprisingly often ignored when building experiments to understand the human brain, or when building tools or softwares based on audio. Wakefield and Viemeister (1990) stated "we often represent acoustic signals in purely temporal or spectral terms, despite the observation that our perception of these signals reflects variations in spectral energy over time which neither a temporal nor spectral representation can capture adequately". This point has been repeatedly made by researchers working in a wide variety of contexts, such as auditory scene analysis and categorization (Bregman, 1994, chap. 3). The tendency to implement auditory stimuli without specifying relevant (or any!) details about the amplitude envelope has drawn criticism; Schutz and Gillard (2020) pointed out that researchers are "…significantly more likely to include the exact model of delivery device than any information about amplitude envelope". On the other hand, there are still many aspects of modulation perception that are poorly understood. For instance, after working on modulation and temporal integration processes of the auditory system for over 30 years, Viemeister (2014) wrote "… models based on modulation filtering do not adequately capture some fundamental aspects of temporal processing because they minimize, or ignore, the temporal structure of the envelopes of complex stimuli. Time-reversed waveforms such as speech tokens and ramped vs. damped broadband noise can sound considerably different even though their amplitude spectra in both the AM domain and the audio frequency domain are identical. Modulation filters that ignore phase are missing

a crucial aspect of temporal processing. However, temporally based models such as those based on cross correlation and template matching have not been able to provide a good quantitative account of AM frequency selectivity." Although the concept of amplitude modulation or amplitude envelope is not new, it is not widely known or appreciated. Some authors express reservations about its relevance or usefulness in the overall understanding of the human auditory system. For instance, after reviewing a significant amount of literature on the phenomena related to modulation perception, Moore (2012, p. 189) said "The modulation filter bank has been suggested as a possible explanation for certain perceptual phenomena, some of which are described below. It should be emphasized, however, that this is still a controversial concept." The processes of the brain that are responsible for decoding various aspects of amplitude modulations from sound are not strictly auditory but also partially cognitive (Viemeister, 2014). In practical terms (for acoustic engineering), this implies that any perceptual effect of an AM signal is determined significantly by the properties of the audio device which produces it, which may have a nonlinear relationship with the parameters of the signal.

## 6.2 Limitations and future work

This thesis was intended as a broad exploration into a relatively uncommon concept. Hence, the effort was made to neither exhaust the depth nor the breadth of any particular topic. Rather, the goal was to find out useful connections between phenomena that may be of a general interest to disciplines such as audio signal processing, psychology, ergonomics and so on. Obviously, this means that the present work suffers from many limitations. Some general ideas are summarized below that could serve as future research directions.

- Auditory effects of ANC beyond amplitude and into frequency modulation (FM) could be studied. Given that auditory researchers have moved away from studying AM and FM in isolation and have been advocating for investigation into the joint spectro-temporal modulation space (Elliott & Theunissen, 2009; Elhilali, 2019), it would be enriching to re-create a parallel set of experiments described in this thesis in the context of psychoacoustics of FM.

- In chapter 3, only one cochlear model was (implicitly) used to study the brain's response to the AM stimuli, which was part of the model of auditory nerve synapse. However, the cochlear filter models that mimic the basilar membrane vibrations are equally important (as the auditory nerve model) in

understanding auditory transduction at the inner hair cells. Saremi et al. (2016) has conducted an extensive evaluation of seven of these models by providing a set of standardized stimuli; it would be straightforward and interesting to extend their work by analyzing the ANC recorded signals in a similar fashion and studying the pre-neural cochlear filter outputs.

- All the AM signals investigated in chapters 2 and 3 were sinusoids imposed on other sinusoids. This may provide the basis for all potential signals from a Fourier perspective, but given the non-linear interactions between ANC and playback signals, it may be worthwhile investigating other (non-sinusoidal) forms of signals as modulators and also as carriers.

- All the speech signals investigated in chapters 4 and 5 were derived with a single type of chimera. The literature around vocoding and chimera (and temporal envelope of speech in general) have proposed and utilized many other types of these signals to understand the human auditory system. Some of these could be adapted to better characterize the effect of ANC, e.g. using comb filtering or bandpass filtering to investigate relative importance of modulation frequencies in the envelope.

- ANC was investigated from a black-box perspective here, where it was only 'on' or 'off'. While this forms the baseline of measurements, the next logical step would be to 'open' the black-box and tweak various parameters of the ANC algorithm while conducting the AM measurements described in chapters 2, 3, 4 and 5.

- Due to the adaptive properties and plasticity of neurons in the brain, using an ANC headphone for a prolonged duration may also cause lingering effects after the user takes off the headphones. The literature on ANC has not thoroughly investigated these possibilities, and neither is the purpose of this thesis.

- Humans have two ears, and the so called 'binaural' aspects of hearing form a relatively less explored discipline under auditory psychology and neuroscience. All the models used in the present work (and largely, elsewhere) are of a monoaural nature, that is, they do not consider the interactions between two ears and take only a single channel of audio as input. In addition, it remains an open question whether ANC headphones have any differences in delivering the playback signal to the two ears, i.e. whether the two ears get a dedicated 'control signal' and playback signal (figure 1.1). Hence, a good place to start would be to use binaural models of hearing and apply them to dummy-head

microphone recorded signals, which already provide an excellent approximation of the 'head related transfer function' and serve as a proxy for the sound pressure waveform available at the eardrum. In fact, ANC earphones have been assessed using objective (insertion loss) and subjective (perceived loudness) methods in a binaural hearing context (Watanabe & Hamada, 2011).

- The participants' task in chapter 4 was a simple closed-response one, with no additional cognitive load such as dual-task (as done by e.g. (Molesworth, Burgess, & Kwon, 2013)). Increasing the task complexity and cognitive load would allow future researchers to narrow down the situations where the interaction between 'ANC effect' and AM perception is enhanced or suppressed. Similar concepts have been the focus of many aviation related studies, e.g. by Valimont et al. (2006), but the general phenomena of interaction between ANC and cognition remains under-explored.

- The quality of the physical seal of an ANC headphone around the ear makes a substantial difference in the leakage signal and thereby the quality of the ANC performance and beneficial effects; this phenomenon is widely investigated in the context of hearing protection devices but less so in the ANC literature. All the experiments of the present work were done with circumaural headphones, i.e. they completely covered the ear from all sides. Since there are earplug type and 'on-the-ear' type ANC headphones in the market as well, these devices should be investigated in future studies.

- The auditory system is known to process naturalistic stimuli in different ways than simple/artificial stimuli (such as modulated narrowband noises or tones) (Eggermont, 2001; Schutz & Gillard, 2020), so it would be useful and practical to characterize the ANC effect using such stimuli, whether using models or human listeners.

- The possibility of using ANC as a tool to study human hearing is also interesting. Valimont (2006) speculated about the possibility of a phenomenon that is the opposite of 'upward spread of masking' caused by ANC. In this way, an ANC headphone allows auditory researchers to probe the mechanisms of audition related to the inverse, i.e. whether the 'unmasking' has any significant impact on the human auditory system. Fabry et al. (1993) investigated such possibilities in the context of hearing aids, and including ANC in the arsenal for similar studies would provide researchers with the opportunity to come up with novel models and explanations of missing pieces of the puzzle regarding the overall functioning of the auditory system.

- Leroux et al. (2018) wrote a detailed report on the effectiveness of hearing aids and associated technologies while being used in noisy workplaces by hearing impaired users. Among other things, they observed that "the maximum output of hearing aids is typically regulated according to the individual's discomfort threshold, or on the basis of discomfort thresholds measured with relatively brief sound signals. However, discomfort thresholds measured in a clinic may not be very representative of tolerance over longer durations (for example, an eight-hour work shift) and are often higher than the sound levels that could damage hearing when continued over a long period." This aspect of prolonged exposure is also applicable to the work presented here, i.e, it should be investigated whether the prolonged use of ANC devices induces any changes (beneficial or harmful) that lingers well beyond the usage period.

- ANC systems are increasingly being deployed as part of larger systems, under the broad umbrella term of 'hearables'. This means any effect of ANC on the playback signal is going to be further affected by the other digital signal processing blocks, and the final perception of such a device needs to be separately evaluated. For instance, a system with voice activity detection and ANC may interact in unusual ways in reverberant environments (Canli & Doğan, 2019). Given the rapid pace of advancement of audio technology, such complexity is difficult to manage while carrying out human subject based listening tests. Hence, a broader family of characterization methods for assessing such devices are needed.

# Bibliography


Abel, S. M., & Spencer, D. L. (1995). Active noise cancellation vs. passive sound attenuation: Help or hindrance? *Canadian Acoustics*, *23*(3), 23–24.

Abel, S. M., & Spencer, D. L. (1997). Active noise reduction versus conventional hearing protection relative benefits for normal-hearing and impaired listeners. *Scandinavian Audiology*, *26*(3), 155–167.

Abel, S. M., Tsang, S., & Boyne, S. (2007). Sound localization with communications headsets: Comparison of passive and active systems. *Noise and Health*, *9*(37).

Akhtar, S., Weigle, C. G., Cheng, E. Y., Toohill, R., & Berens, R. J. (2000). Use of active noise cancellation devices in caregivers in the intensive care unit. *Critical Care Medicine*, *28*(4), 1157–1160.

Akhtar, Z., Siddique, K., Rattani, A., Lutfi, S. L., & Falk, T. H. (2019). Why is multimedia quality of experience assessment a challenging problem? *IEEE Access*, *7*, 117897–117915.

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing*, *40*(5), 1084.

Anderson, W., & Garinther, G. (1997). Effects of active noise reduction in armor crew headsets. *Audio Effectiveness in Aviation*, *15*, 20–1.

Ang, L. Y. L., Koh, Y. K., & Lee, H. P. (2017). The performance of active noise-canceling headphones in different noise environments. *Applied Acoustics*, *122*, 16–22.

Baby, D., & Verhulst, S. (2018). Machines hear better when they have ears. *arXiv preprint arXiv:1806.01145*.

Backman, J. (1997). A method for measuring modulation transmission in speech transmitted via a nonlinear channel. *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, *3*, 1715–1718.

Backman, J. (2010). Complex modulation transfer function and its applications in transducer and room acoustics measurements. *Audio Engineering Society Convention 129*.

Bagha, S., Tripathy, R., Nanda, P., Preetam, C., & Das, D. P. (2018). Understanding perception of active noise control system through multichannel EEG analysis. *Healthcare Technology Letters*, *5*(3), 101–106.

Bao, H., & Panahi, I. M. (2010a). A novel feedforward active noise control structure with spectrum-tuning for residual noise. *IEEE Transactions on Consumer Electronics*, *56*(4), 2093–2097.

Bao, H., & Panahi, I. M. (2010b). Psychoacoustic active noise control with ITU-r 468 noise weighting and its sound quality analysis. *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 4323–4326.

Bao, H., & Panahi, I. M. (2013). A perceptually motivated active noise control design and its psychoacoustic analysis. *ETRI Journal*, *35*(5), 859–868.

Basu, S., Tackett, J., Trumpy, D., Walt, A., & Adari, S. (2021). *Study of the effects of active noise cancellation on music playback*. SAE Technical Paper.

Bech, S., & Zacharov, N. (2007). *Perceptual audio evaluation-theory, method and application*. John Wiley & Sons.

Behn, M. (2013). Psychoacoustical evaluation of active noise control in headphones. *AIA-DAGA 2013 Proceedings of the International Conference on Acoustics*, 1707.

Belyi, V., & Gan, W.-S. (2017). Psychoacoustic subband active noise control algorithm. *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 457–463.

Belyi, V., & Gan, W.-S. (2019). Integrated psychoacoustic active noise control and masking. *Applied Acoustics*, *145*, 339–348.

Benois, P. R., Papantoni, V., & Zölzer, U. (2018). Psychoacoustic hybrid active noise control structure for application in headphones. *2018 Proceedings of the 25th International Congress on Sound and Vibration*.

Berg, B. G. (1996). On the relation between comodulation masking release and temporal modulation transfer functions. *The Journal of the Acoustical Society of America*, *100*(2), 1013–1023.

Berger, E. (2002). Active noise reduction (ANR) in hearing protection: Does it make sense for industrial applications. *27th Conference of the National Hearing Conservation Association*.

Bhattacharjya, A. (2019, December). Future of noise cancellation category products in the market. Retrieved February 28, 2022, from https://www.businessworld.in/article/Future-Of-Noise-Cancellation-Category-Products-In-The-Market/21-12-2019-180885/

Biberger, T., Fleßner, J.-H., Huber, R., & Ewert, S. D. (2018). An objective audio quality measure based on power and envelope power cues. *Journal of the Audio Engineering Society*, *66*(7), 578–593.

Biberger, T., Schepker, H., Denk, F., & Ewert, S. D. (2021). Instrumental quality predictions and analysis of auditory cues for algorithms in modern headphone technology. *Trends in Hearing*, *25*, 23312165211001219.

Blumstein, S. E. (2021). Features in speech perception and lexical access. *The Handbook of Speech Perception, Wiley*, 123–144.

Bockstael, A., De Coensel, B., Botteldooren, D., D'Haenens, W., Keppler, H., Maes, L., Philips, B., Swinnen, F., & Bart, V. (2011). Speech recognition in noise with active and passive hearing protectors: A comparative study. *The Journal of the Acoustical Society of America*, *129*(6), 3702–3715.

Bolia, R. S. (2003). Effects of spatial intercoms and active noise reduction headsets on speech intelligibility in an AWACS environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*, 100–103.

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, *107*(2), 1065–1066.

Bolia, R. S., Nelson, W. T., Vidulich, M. A., Simpson, B. D., & Brungart, D. S. (2005). Communications research for command & control: Human-machine interface technologies supporting effective air battle management. *Air Force Research Laboratory*.

Bosker, H. R., & Cooke, M. (2018). Talkers produce more pronounced amplitude modulations when speaking in noise. *The Journal of the Acoustical Society of America*, *143*(2), EL121–EL126.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Brammer, A. J., Peterson, D. R., Cherniack, M. G., & Gullapalli, S. (2005). Improving the effectiveness of communication headsets with active noise reduction: Influence of control structure. *New Directions for Improving Audio Effectiveness*.

Brammer, A. J., Yu, G., Bernstein, E. R., Cherniack, M. G., Peterson, D. R., & Tufts, J. B. (2014). Understanding speech when wearing communication headsets and hearing protectors with subband processing. *The Journal of the Acoustical Society of America*, *136*(2), 671–681.

Brammer, A. J., Yu, G., Peterson, D. R., Bernstein, E. R., & Cherniak, M. (2008). Hearing protection and communication in an age of digital signal processing: Progress and prospects. *Proceedings of the 9th International Congress on Noise as a Public Health Problem*, 21–25.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Brixen, E. B. (2020). *Audio metering: Measurements, standards, and practice*. Focal Press.

Bromwich, M. A., Parsa, V., Lanthier, N., Yoo, J., & Parnes, L. S. (2008). Active noise reduction audiometry: A prospective analysis of a new approach to noise management in audiometric testing. *The Laryngoscope*, *118*(1), 104–109.

Brungart, D. S. (2001). Evaluation of speech intelligibility with the coordinate response measure. *The Journal of the Acoustical Society of America*, *109*(5), 2276–2279.

Burgess, M., & Molesworth, B. (2012). Investigating the effects of active noise cancelling headphones on speech intelligibility in aviation. *41st International Congress and Exposition on Noise Control Engineering*, *2012*, 8723–8726.

Burgess, M., & Molesworth, B. (2013). Investigation of the role for noise cancelling headphones to assist speech recall in noise. *Proceedings of Meetings on Acoustics*, 040012–040012.

Butterbaugh, S., & Gordon-Pershey, M. (2019). The usefulness of active noise cancelling headphones when listening to music in the presence of ambient noise. *eHearsay*, *9*, 17.

Cabot, R. C. (1997). Fundamentals of modern audio measurement. *Audio Engineering Society Conference: UK 12th Conference: The Measure of Audio (MOA)*.

Campbell, D., Jones, E., & Glavin, M. (2009). Audio quality assessment techniques—a review, and recent developments. *Signal Processing*, *89*(8), 1489–1500.

Canevet, G., & Mangiante, G. (2004). Psychoacoustic assessment of active noise control. *Proceedings of ACTIVE 04*, 1–10.

Canli, Ö., & Doğan, H. (2019). Performance evaluation of a speech protected active noise control system. *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 1–4.

Carney, L. H. (1993). A model for the responses of low-frequency auditory-nerve fibers in cat. *The Journal of the Acoustical Society of America*, *93*(1), 401–417.

Casali, J. G., & Gower, D. W. (1993). Communications headset augmentation via active noise cancellation: Attenuation and speech intelligibility performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *37*, 554–558.

Casali, J. G., Robinson, G. S., Dabney, E. C., & Gauger, D. (2004). Effect of electronic ANR and conventional hearing protectors on vehicle backup alarm detection in noise. *Human Factors*, *46*(1), 1–10.

Catellier, A. A., & Voran, S. D. (2020). Wawenets: A no-reference convolutional waveform-based approach to estimating narrowband and wideband speech quality [ISSN: 2379-190X]. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 331–335.

Chang, H.-Y., Luo, C.-H., Lo, T.-S., & Tai, C.-C. (2019). Compensated active noise cancellation earphone for audiometric screening tests in noisy environments. *International journal of audiology*, *58*(11), 747–753.

Chinen, M., Lim, F. S., Skoglund, J., Gureev, N., O'Gorman, F., & Hines, A. (2020). ViSQOL v3: An open source production ready objective speech and audio metric. *2020 twelfth international conference on quality of multimedia experience (QoMEX)*, 1–6.

Clark, J. G., Brady, M., Earl, B. R., Scheifele, P. M., Snyder, L., & Clark, S. D. (2017). Use of noise cancellation earphones in out-of-booth audiometric evaluations. *International Journal of Audiology*, *56*(12), 989–996.

Crabtree, R. (1997). Constraints in the application of personal active noise reduction systems. *Audio Effectiveness in Aviation*, 15–1.

Darlington, P., & Guiu, P. (2010). Active noise reduction in personal audio delivery systems; assessment using loudness balance methods. *Audio Engineering Society Convention 128*.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997a). Modeling auditory processing of amplitude modulation. I. detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, *102*(5), 2892–2905.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997b). Modeling auditory processing of amplitude modulation. II. spectral and temporal integration. *The Journal of the Acoustical Society of America*, *102*(5), 2906–2919.

Dau, T., Püschel, D., & Kohlrausch, A. (1996). A quantitative model of the "effective"signal processing in the auditory system. i. model structure. *The Journal of the Acoustical Society of America*, *99*(6), 3615–3622.

Davis, R. I., Qiu, W., & Hamernik, R. P. (2009). Role of the kurtosis statistic in evaluating complex noise exposures for the protection of hearing. *Ear and Hearing*, *30*(5), 628–634.

de Diego, M., Gonzalez, A., Pinero, G., Ferrer, M., & Garcia-Bonito, J. (2001). Subjective evaluation of actively controlled interior car noise. *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, *5*, 3225–3228 vol.5.

Delgutte, B., Hammond, B., & Cariani, P. (1998). Neural coding of the temporal envelope of speech: Relation to modulation transfer functions. *Psychophysical and Physiological Advances in Hearing*, 595–603.

Denk, F., Vogl, S., Schepker, H., Kollmeier, B., Blau, M., & Doclo, S. (2017). The acoustically transparent hearing device: Towards integration of individualized sound equalization, electro-acoustic modeling and feedback cancellation. *Proc. 1st International Workshop on Challenges in Hearing Assistive Technology (CHAT-2017)*, 89–94.

de Oliveira, L. P. (2010). Review and future perspectives on active sound quality control. *Proceedings of the International Conference on Noise and Vibration Engineering–ISMA*, *2*, 2–4.

de Oliveira, L. P., Janssens, K., Gajdatsy, P., Van der Auweraer, H., Varoto, P. S., Sas, P., & Desmet, W. (2009). Active sound quality control of engine induced cavity noise. *Mechanical systems and signal processing*, *23*(2), 476–488.

de Oliveira, L. P., Stallaert, B., Janssens, K., Van der Auweraer, H., Sas, P., & Desmet, W. (2010). NEX-LMS: A novel adaptive control scheme for harmonic sound quality control. *Mechanical Systems and Signal Processing*, *24*(6), 1727–1738.

Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America*, *95*(2), 1053–1064.

Dubbelboer, F., & Houtgast, T. (2008). The concept of signal-to-noise ratio in the modulation domain and speech intelligibility. *The Journal of the Acoustical Society of America*, *124*(6), 3937–3946.

Durant, E., Xiao, J., Xu, B., McKinney, M., & Zhang, T. (2013). Perceptually motivated ANC for hearing-impaired listeners. *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1–4.

Eddins, D. A., & Liu, C. (2012). Psychometric properties of the coordinate response measure corpus with various types of background interference. *The Journal of the Acoustical Society of America*, *131*(2), EL177–EL183.

Edraki, A., Chan, W.-Y., Jensen, J., & Fogerty, D. (2021). Speech intelligibility prediction using spectro-temporal modulation analysis. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 210–225.

Eggermont, J. J. (2001). Between sound and perception: Reviewing the search for a neural code. *Hearing Research*, *157*(1), 1–42.

Elhilali, M. (2019). Modulation representations for speech and music. *Timbre: Acoustics, perception, and cognition* (pp. 335–359). Springer.

Elhilali, M., Chi, T., & Shamma, S. A. (2003). A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. *Speech Communication*, *41*(2), 331–348.

Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLOS Computational Biology*, *5*(3), e1000302.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge university press.

Fabry, D. A., Leek, M. R., Walden, B. E., & Cord, M. T. (1993). Do adaptive frequency response (AFR) hearing aids reduce 'upward spread' of masking? *Journal of Rehabilitation Research and Development*, *30 3*, 318–25.

Falk, T. H., Zheng, C., & Chan, W.-Y. (2010). A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*(7), 1766–1774.

Fortune, T. W., Woodruff, B. D., & Preves, D. A. (1994). A new technique for quantifying temporal envelope contrasts. *Ear and Hearing*, *15*(1), 93–99.

Francis, A. L., & Love, J. (2020). Listening effort: Are we measuring cognition or affect, or both? *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(1), e1514.

Füllgrabe, C., Stone, M. A., & Moore, B. C. (2009). Contribution of very low amplitude-modulation rates to intelligibility in a competing-speech task. *The Journal of the Acoustical Society of America*, *125*(3), 1277–1280.

Gabrielsson, A., & Sjögren, H. (1979). Perceived sound quality of sound-reproducing systems. *The Journal of the Acoustical Society of America*, *65*(4), 1019–1033.

Galambos, R., & Davis, H. (1943). The response of single auditory-nerve fibers to acoustic stimulation. *Journal of neurophysiology*, *6*(1), 39–57.

Gallacher, S., Enki, D., Stevens, S., & Bennett, M. J. (2017). An experimental model to measure the ability of headphones with active noise control to reduce patient's exposure to noise in an intensive care unit. *Intensive care medicine experimental*, *5*(1), 1–8.

Geddes, E. R., & Lee, L. W. (2003). Auditory perception of nonlinear distortion-theory. *Audio Engineering Society Convention 115*.

George, N. V., & Panda, G. (2013). Advances in active noise control: A survey, with emphasis on recent nonlinear techniques. *Signal Processing*, *93*(2), 363–377.

Ghido, F., Disch, S., Herre, J., Reutelhuber, F., & Adami, A. (2017). Coding of fine granular audio signals using high resolution envelope processing (HREP). *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 701–705.

Gonzalez, A., Ferrer, M., De Diego, M., Pinero, G., & Garcia-Bonito, J. (2003). Sound quality of low-frequency and car engine noises after active noise control. *Journal of Sound and Vibration*, *265*(3), 663–679.

Goodman, D. F., Winter, I. M., Léger, A. C., de Cheveigné, A., & Lorenzi, C. (2018). Modelling firing regularity in the ventral cochlear nucleus: Mechanisms, and effects of stimulus level and synaptopathy. *Hearing Research*, *358*, 98–110.

Gower, D. W., & Casali, J. G. (1994). Speech intelligibility and protective effectiveness of selected active noise reduction and conventional communications headsets. *Human Factors*, *36*(2), 350–367.

Greenberg, S., & Arai, T. (2001). The relation between speech intelligibility and the complex modulation spectrum. *Seventh European Conference on Speech Communication and Technology*.

Guérin, A., Jeannès, R. L. B., Bès, J., Faucon, G., & Lorenzi, C. (2006). Evaluation of two computational models of amplitude modulation coding in the inferior colliculus. *Hearing Research*, *211*(1), 54–62.

Guicking, D. (2007). Active control of sound and vibration: History–fundamentals–state of the art. *Oscillations, Waves and Interactions–60 Years Drittes Physikalisches Institute*, 107–138.

Guldenschuh, M., Sontacchi, A., Perkmann, M., & Opitz, M. (2012). Assessment of active noise cancelling headphones. *2012 IEEE Second International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, 299–303.

Hagood, M. (2011). Quiet comfort: Noise, otherness, and the mobile production of personal space. *American Quarterly*, *63*(3), 573–589.

Hall, D. A., Chambers, J., Akeroyd, M. A., Foster, J. R., Coxon, R., & Palmer, A. R. (2009). Acoustic, psychophysical, and neuroimaging measurements of the effectiveness of active cancellation during auditory functional magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *125*(1), 347–359.

Hamacher, D., Schley, F., Hollander, K., & Zech, A. (2018). Effects of manipulated auditory information on local dynamic gait stability. *Human Movement Science*, *58*, 219–223.

Hamernik, R. P., Qiu, W., & Davis, B. (2003). The effects of the amplitude distribution of equal energy exposures on noise-induced hearing loss: The kurtosis metric. *The Journal of the Acoustical Society of America*, *114*(1), 386–395.

Hamschin, B., & Loughlin, P. J. (2013). Time and frequency constrained sonar signal design for optimal detection of elastic objects. *The Journal of the Acoustical Society of America*, *133*(4), 2169–2179.

Hansen, C. H. (2002). *Understanding active noise cancellation*. CRC Press.

Hansen, C. H., Snyder, S. D., Qiu, X., Brooks, L. A., & Moreau, D. J. (2012). *Active control of noise and vibration: Volume 1* (Vol. 1). CRC Press.

Harlander, N., Huber, R., & Ewert, S. D. (2014). Sound quality assessment using auditory models. *Journal of the Audio Engineering Society*, *62*(5), 324–336.

Harpe, S. E. (2015). How to analyze likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, *7*(6), 836–850.

Harris, L. (2015). *Development of a modulation transfer function-based method for evaluating bass reproduction accuracy in professional monitoring loudspeakers* (Doctoral dissertation). University of Southampton. England.

Harris, L., Newell, P., & Holland, K. (2018). The 'bass transmission index': A new concept for evaluating loudspeaker performance. *Proceedings of the Institute of Acoustics*. https://eprints.soton.ac.uk/426110/

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in psychology* (pp. 139–183). Elsevier.

Hasegawa, R., Yamashita, H., & Kajikawa, Y. (2018). Effectiveness of active noise control system for non-stationary noise in consideration of psychoacoustic properties. *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 1256–1261.

Hauenstein, M. (1998). Application of meddis' inner hair-cell model to the prediction of subjective speech quality. *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, *1*, 545–548.

Heil, P., Neubauer, H., & Irvine, D. R. (2011). An improved model for the rate–level functions of auditory-nerve fibers. *Journal of Neuroscience*, *31*(43), 15424–15437.

Hilkhuysen, G., Gaubitch, N., Brookes, M., & Huckvale, M. (2014). Effects of noise suppression on intelligibility. II: An attempt to validate physical metrics. *The Journal of the Acoustical Society of America*, *135*(1), 439–450.

Hines, A. (2012). *Predicting speech intelligibility* (Doctoral dissertation). Trinity College Dublin.

Hines, A., Skoglund, J., Kokaram, A., & Harte, N. (2012). ViSQOL: The virtual speech quality objective listener. *IWAENC 2012; International Workshop on Acoustic Signal Enhancement*, 1–4.

Hines, A., Skoglund, J., Kokaram, A. C., & Harte, N. (2015). ViSQOL: An objective speech quality model. *EURASIP Journal on Audio, Speech, and Music Processing*, *2015*(1), 1–18.

Hirahara, T. (2004). Physical characteristics of headphones used in psychophysical experiments. *Acoustical Science and Technology*, *25*(4), 276–285.

Hoover, E. C., Souza, P. E., & Gallun, F. J. (2012). The consonant-weighted envelope difference index (cEDI): A proposed technique for quantifying envelope distortion. *Journal of Speech, Language, and Hearing Research*.

Houtgast, T. (1989). Frequency selectivity in amplitude-modulation detection. *The Journal of the Acoustical Society of America*, *85*(4), 1676–1680.

Howard, D. M., & Angus, J. A. S. (2017). *Acoustics and psychoacoustics*. Routledge.

Hu, G., Determan, S. C., Dong, Y., Beeve, A. T., Collins, J. E., & Gai, Y. (2020). Spectral and temporal envelope cues for human and automatic speech recognition in noise. *Journal of the Association for Research in Otolaryngology*, *21*(1), 73–87.

Hu, Y., & Loizou, P. C. (2007). Subjective comparison and evaluation of speech enhancement algorithms. *Speech Communication*, *49*(7), 588–601.

Huber, R., & Kollmeier, B. (2006). PEMO-q—a new method for objective audio quality assessment using a model of auditory perception. *IEEE Transactions on audio, speech, and language processing*, *14*(6), 1902–1911.

Hunn, N. (2016). *The market for hearable devices 2016–2020*. WiFore Wireless Consulting.

Hunn, N. (2020). *The market for hearable devices 2020–2025*. WiFore Wireless Consulting.

Imaizumi, K., Priebe, N. J., Sharpee, T. O., Cheung, S. W., & Schreiner, C. E. (2010). Encoding of temporal information by timing, rate, and place in cat auditory cortex. *PLoS One*, *5*(7), e11531.

ISO. (2017). ISO 532-1 acoustics–methods for calculating loudness–part 1: Zwicker method. *International Organization for Standardization*.

ITU-R. (1986). Recommendation 468-4: Measurement of audio frequency noise in broadcasting, sound recording systems and on sound programme circuits. *International Telecommunications Union*, *10*(1), 285–291.

ITU-T. (2001). P.862 perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. *International Telecommunications Union*.

ITU-T. (2003). P.380 electro-acoustic measurements on headsets. *International Telecommunications Union*.

ITU-T. (2015). BS.1116 methods for the subjective assessment of small impairments in audio systems. *International Telecommunications Union*.

ITU-T. (2016). P.381 technical requirements and test methods for the universal wired headset or headphone interface of digital mobile terminals. *International Telecommunications Union*.

ITU-T. (2020). P.501 test signals for use in telephony and other speech-based applications. *International Telecommunications Union*.

Ives, D. T., Calcus, A., Kalluri, S., Strelcyk, O., Sheft, S., & Lorenzi, C. (2013). Effects of noise reduction on AM and FM perception. *Journal of the Association for Research in Otolaryngology*, *14*(1), 149–157.

Ives, D. T., Kalluri, S., Strelcyk, O., Sheft, S., Miermont, F., Coez, A., Bizaguet, E., & Lorenzi, C. (2014). Effects of noise reduction on AM perception for hearing-impaired listeners. *Journal of the Association for Research in Otolaryngology*, *15*(5), 839–848.

Jang, R., Molesworth, B., Burgess, M., & Estival, D. (2014). Improving communication in general aviation through the use of noise cancelling headphones. *Safety Science*, *62*, 499–504.

Jeffress, L. A. (1948). A place theory of sound localization. *Journal of comparative and physiological psychology*, *41*(1), 35.

Jenstad, L. M., & Souza, P. E. (2005). Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *Journal of Speech, Language, and Hearing Research*, *48*(3), 651–667.

Jepsen, M. L., Ewert, S. D., & Dau, T. (2008). A computational model of human auditory signal processing and perception. *The Journal of the Acoustical Society of America*, *124*(1), 422–438.

Jiang, J., & Li, Y. (2018). Review of active noise control techniques with emphasis on sound quality enhancement. *Applied Acoustics*, *136*, 139–148.

Johnson, K., & Sjerps, M. J. (2021). Speaker normalization in speech perception. *The handbook of speech perception, Wiley*, 145–176.

Joris, P., Schreiner, C., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews*, *84*(2), 541–577.

Karalar, M., Keles, I., Doğantekin, E., Kahveci, O. K., & Sarici, H. (2016). Reduced pain and anxiety with music and noise-canceling headphones during shockwave lithotripsy. *Journal of Endourology*, *30*(6), 674–677.

Kari, T., Makkonen, M., & Frank, L. (2017). The effect of using noise cancellation earplugs in open-plan offices on the offices on the work well-being and work performance of software professionals. In A. Harfouche, A. Pateli, A. Polyviou, N. P. A. Kokkinaki, & T. Torre (Eds.), *MCIS 2017 : 11th mediterranean conference on information systems* (p. 36). MCIS.

Kass, R. E., Amari, S.-I., Arai, K., Brown, E. N., Diekman, C. O., Diesmann, M., Doiron, B., Eden, U. T., Fairhall, A. L., Fiddyment, G. M., Fukai, T., Grün, S., Harrison, M. T., Helias, M., Nakahara, H., Teramae, J.-n., Thomas, P. J., Reimers, M., Rodu, J., … Kramer, M. A. (2018). Computational neuroscience: Mathematical and statistical perspectives. *Annual Review of Statistics and Its Application*, *5*(1), 183–214.

Kates, J. M., & Arehart, K. H. (2014a). The hearing-aid speech perception index (HASPI). *Speech Communication*, *65*, 75–93.

Kates, J. M., & Arehart, K. H. (2014b). The hearing-aid speech quality index (HASQI) version 2. *Journal of the Audio Engineering Society*, *62*(3), 99–117.

Kates, J. M., & Arehart, K. H. (2021). The hearing-aid speech perception index (HASPI) version 2. *Speech Communication*, *131*, 35–46.

Kendrick, P., Jackson, I. R., Li, F. F., Fazenda, B., Cox, T., et al. (2015). Perceived audio quality of sounds degraded by non-linear distortions and single-ended assessment using HASQI. *Journal of the Audio Engineering Society*, *63*(9), 698–712.

Kim, Y., & Park, Y. (2017). Effect of active noise control and masking sound on speech intelligibility. *Applied Acoustics*, *123*, 152–157.

Krebs, B., Lesica, N. A., & Grothe, B. (2008). The representation of amplitude modulations in the mammalian auditory midbrain. *Journal of Neurophysiology*, *100*(3), 1602–1609.

Kressner, A. A., Anderson, D. V., & Rozell, C. J. (2013). Evaluating the generalization of the hearing aid speech quality index (HASQI). *IEEE Transactions on Audio, Speech, and Language Processing*, *21*(2), 407–415.

Krishna, B. S., & Semple, M. N. (2000). Auditory temporal processing: Responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *Journal of Neurophysiology*, *84*(1), 255–273.

Lam, B., Gan, W.-S., Shi, D., Nishimura, M., & Elliott, S. (2021). Ten questions concerning active noise control in the built environment. *Building and Environment*, *200*, 107928.

Lee, H. M., Wang, Z., Lim, K. M., & Lee, H. P. (2019). A review of active noise control applications on noise barrier in three-dimensional/open space: Myths and challenges. *Fluctuation and Noise Letters*, *18*(4), 1930002.

Lei, S.-F., Ahroon, W. A., & Hamernik, R. P. (1994). The application of frequency and time domain kurtosis to the assessment of hazardous noise exposures. *The Journal of the Acoustical Society of America*, *96*(3), 1435–1444.

Lelo de Larrea-Mancera, E. S., Stavropoulos, T., Hoover, E. C., Eddins, D. A., Gallun, F. J., & Seitz, A. R. (2020). Portable automated rapid testing (PART) for auditory assessment: Validation in a young adult normal-hearing population. *The Journal of the Acoustical Society of America*, *148*(4), 1831–1851.

Leroux, T., Laroche, C., Giguère, C., & Voix, J. (2018). Hearing aid use in noisy workplaces.

Li, H., Chen, K., Wang, X., Gao, Y., & Yu, W. (2018). A perceptual dissimilarities based nonlinear sound quality model for range hood noise. *The Journal of the Acoustical Society of America*, *144*(4), 2300–2311.

Liang, M., Zhao, F., French, D., & Zheng, Y. (2012). Characteristics of noise-canceling headphones to reduce the hearing hazard for MP3 users. *The Journal of the Acoustical Society of America*, *131*(6), 4526–4534.

Liebich, S., Fabry, J., Jax, P., & Vary, P. (2018). Signal processing challenges for active noise cancellation headphones. *Speech Communication; 13th ITG-Symposium*, 1–5.

Liebich, S., Richter, J.-G., Fabry, J., Durand, C., Fels, J., & Jax, P. (2018). Direction-of-arrival dependency of active noise cancellation headphones. *Noise Control and Acoustics Division Conference*, *51425*, V001T08A003.

Liu, L., Yu, Y., Qiu, Y., Zheng, X., Dai, Z., & Cui, Y. (2022). Reproduction of sound field inside an aircraft mock-up for evaluating ANC devices. *Applied Acoustics*, *188*, 108588.

Lorenzi, C., Berthommier, F., & Demany, L. (1999). Discrimination of amplitude-modulation phase spectrum. *The Journal of the Acoustical Society of America*, *105*(5), 2987–2990.

Lorho, G. (2010). *Perceived quality evaluation: An application to sound reproduction over headphones* (Doctoral dissertation). Helsinki University of Technology. https://aaltodoc.aalto.fi/handle/123456789/4807

Lu, L., Yin, K.-L., de Lamare, R. C., Zheng, Z., Yu, Y., Yang, X., & Chen, B. (2021a). A survey on active noise control in the past decade–part i: Linear systems. *Signal Processing*, 108039.

Lu, L., Yin, K.-L., de Lamare, R. C., Zheng, Z., Yu, Y., Yang, X., & Chen, B. (2021b). A survey on active noise control in the past decade–part II: Nonlinear systems. *Signal Processing*, *181*, 107929.

Lueg, P. (1936). Process of silencing sound oscillations. *US Patent 2043416*.

Lyon, R. F. (2017). *Human and machine hearing: Extracting meaning from sound*. Cambridge University Press.

Mangiante, G. (2004). Limitations on the performance of active noise control systems due to subjective effects. *The Journal of the Acoustical Society of America*, *115*(5), 2498–2498.

Mangiante, G., & Canevet, G. (2007). Noise annoyance reduction using active control. *The Journal of the Acoustical Society of America*, *121*(5), 3179–3179.

Marchesiello, S., Fasana, A., & Garibaldi, L. (2015). Best parameter choice of stochastic resonance to enhance fault signature in bearings. *ICEDYN 2015*.

Marozeau, J. (2011). Models of loudness. *Loudness* (pp. 261–284). Springer.

Marple, L. (1999). Computing the discrete-time" analytic" signal via FFT. *IEEE Transactions on signal processing*, *47*(9), 2600–2603.

Martin, D. W., & Anderson, L. J. (1945). Headphone measurements and their interpretation. *The Journal of the Acoustical Society of America*, *17*(1), 101–101.

Martin, S. B., Lucke, K., & Barclay, D. R. (2020). Techniques for distinguishing between impulsive and non-impulsive sound in the context of regulating sound exposure for marine mammals. *The Journal of the Acoustical Society of America*, *147*(4), 2159–2176.

McDonald, J. H. (2009). *Handbook of biological statistics* (Vol. 2). sparky house publishing Baltimore, MD.

McJury, M., Stewart, R., Crawford, D., & Toma, E. (1997). The use of active noise control (ANC) to reduce acoustic noise generated during MRI scanning: Some initial results. *Magnetic Resonance Imaging*, *15*(3), 319–322.

McKinley, R. L., Morris, L. J., & Nixon, C. W. (1997). Special applications of active noise reduction headsets. *Audio Effectiveness in Aviation*, 21–1.

Meddis, R., & Lopez-Poveda, E. A. (2010). Auditory periphery: From pinna to auditory nerve. *Computational models of the auditory system* (pp. 7–38). Springer.

Meddis, R., Lopez-Poveda, E. A., Fay, R. R., & Popper, A. N. (2010). *Computational models of the auditory system*. Springer.

Moeller, S., & Heusdens, R. (2013). Objective estimation of speech quality for communication systems. *Proceedings of the IEEE*, *101*(9), 1955–1967.

Molesworth, B., & Burgess, M. (2013). Improving intelligibility at a safety critical point: In flight cabin safety. *Safety Science*, *51*(1), 11–16.

Molesworth, B., Burgess, M., & Chung, A. (2013). Using active noise cancelling headphones to reduce the effects of masking in commercial aviation. *Acta Acustica united with Acustica*, *99*(5), 822–827.

Molesworth, B., Burgess, M., Gunnell, B., & Duong, K. (2014). Examining the benefits of active noise control and supplementary pictorial information in improving the recall of information in a noisy environment. *Aviation Psychology and Applied Human Factors*.

Molesworth, B., Burgess, M., Gunnell, B., Löffler, D., & Venjakob, A. (2014). The effect on recognition memory of noise cancelling headphones in a noisy environment with native and nonnative speakers. *Noise and Health*, *16*(71), 240.

Molesworth, B., Burgess, M., & Kwon, D. (2013). The use of noise cancelling headphones to improve concurrent task performance in a noisy environment. *Applied Acoustics*, *74*(1), 110–115.

Møller, A. R. (2012). *Hearing: Anatomy, physiology, and disorders of the auditory system*. Plural Publishing.

Moore, B. C. (2012). *An introduction to the psychology of hearing*. Koninklijke Brill NV.

Müller, B. J., Liebl, A., & Martin, N. (2019). Influence of active-noise-cancelling headphones on cognitive performance and employee satisfaction in open space offices. *Proceedings of the 23rd International Congress on Acoustics*.

Nelson, P. C., & Carney, L. H. (2004). A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. *The Journal of the Acoustical Society of America*, *116*(4), 2173–2186.

Nelson, P. C., & Carney, L. H. (2006). Cues for masked amplitude-modulation detection. *The Journal of the Acoustical Society of America*, *120*(2), 978–990.

Nelson, P. C., & Carney, L. H. (2007). Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus. *Journal of Neurophysiology*, *97*(1), 522–539.

Nielsen, K., Kreisman, B., Pallett, S., & Kreisman, N. (2011). Effects of noise attenuation devices on screening distortion product otoacoustic emissions in different levels of background noise. *Journal of Educational Audiology*, *17*.

Nyre, L., Tessem, B., Wendelbo, B., & Øvreås, S. H. (2019). The perceptual mechanics of noise-cancelling headphones. *Self-published*. Retrieved February 6, 2022, from https://teklab.uib.no/artikler/the-perceptual-mechanics-of-noise-cancelling-headphones/

Obleser, J., Herrmann, B., & Henry, M. J. (2012). Neural oscillations in speech: Don't be enslaved by the envelope. *Frontiers in human neuroscience*, *6*, 250.

Ogonowski, Z. (2003). Benchmarking of active noise control. *Pomiary Automatyka Kontrola*, *49*, 8–11.

Owen, D. (2019). *Volume control: Hearing in a deafening world*. Riverhead Books.

Peelle, J. E. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Frontiers in neuroscience*, *8*, 253.

Perala, C. H. (2006). *Active noise reduction headphone measurement: Comparison of physical and psychophysical protocols and effects of microphone placement* (Doctoral dissertation). Virginia Polytechnic Institute and State University. United States.

Plack, C. J. (2018). *The sense of hearing*. Routledge.

Plack, C. J., & Carlyon, R. P. (1995). Loudness perception and intensity coding. *Hearing* (2nd ed., pp. 123–160). Academic Press.

Plomp, R., Houtgast, T., & Steeneken, H. J. (1984). The modulation transfer function in audition. *Limits in perception* (pp. 117–138). CRC Press.

Powell, J. A., Kimball, K. A., Mozo, B. T., & Murphy, B. A. (2003). Improved communications and hearing protection in helmet systems: The communications earplug. *Military Medicine*, *168*(6), 431–436.

Pradhan, S. (2021). *Computationally efficient adaptive algorithms for active control systems* (Doctoral dissertation). University of Technology Sydney. Austalia.

Priese, S., Bruhnken, C., Voss, D., Peissig, J., & Reithmeier, E. (2013). The need for psychoacoustics in active noise cancellation. *AIA-DAGA 2013 Proceedings of the International Conference on Acoustics*, 4.

Pulkki, V., & Karjalainen, M. (2015). *Communication acoustics: An introduction to speech, audio and psychoacoustics*. John Wiley & Sons.

Qiu, X., Lu, J., & Pan, J. (2014). A new era for applications of active noise control. *43rd International Congress on Noise Control Engineering*.

Rabinowitz, N. C., Willmore, B. D., King, A. J., & Schnupp, J. W. (2013). Constructing noise-invariant representations of sound in the auditory pathway. *PLoS Biology*, *11*(11), e1001710.

Radun, J., Kontinen, V., Keränen, J., Tervahartiala, I.-K., & Hongisto, V. (2021). Benefits of active noise-cancelling headphones in offices. *ICBEN Congress on Noise as a Public Health Problem*.

Rees, A., Sarbaz, A., Malmierca, M. S., & Le Beau, F. E. (1997). Regularity of firing of neurons in the inferior colliculus. *Journal of Neurophysiology*, *77*(6), 2945–2965.

Reimes, J. (2019). Evaluation of predicted listening effort for active noise cancelling headsets. *2019 AES International Conference on Headphone Technology*.

Relaño-Iborra, H., Zaar, J., & Dau, T. (2019). A speech-based computational auditory signal processing and perception model. *The Journal of the Acoustical Society of America*, *146*(5), 3306–3317.

Rennies, J., Schepker, H., Holube, I., & Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *The Journal of the Acoustical Society of America*, *136*(5), 2642–2653.

Rhode, W. S., & Greenberg, S. (1994). Encoding of amplitude modulation in the cochlear nucleus of the cat. *Journal of Neurophysiology*, *71*(5), 1797–1825.

Rhode, W. S., Roth, G. L., & Recio-Spinoso, A. (2010). Response properties of cochlear nucleus neurons in monkeys. *Hearing Research*, *259*(1), 1–15.

Rix, A. W., Beerends, J. G., Kim, D.-S., Kroon, P., & Ghitza, O. (2006). Objective assessment of speech and audio quality—technology and applications. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*(6), 1890–1901.

Robinson, G. S., & Casali, J. G. (2000). Issues relating to the conduct of empirical research into the detection of auditory stimuli in noise when wearing nontraditional hearing protectors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*, 722–725.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *336*(1278), 367–373.

Rosen, S., Zhang, Y., & Speers, K. (2015). Spectral density affects the intelligibility of tone-vocoded speech: Implications for cochlear implant simulations. *The Journal of the Acoustical Society of America*, *138*(3), EL318–EL323.

Rudnicki, M., Schoppe, O., Isik, M., Völk, F., & Hemmert, W. (2015). Modeling auditory coding: From sound to spikes. *Cell and Tissue Research*, *361*(1), 159–175.

Rudzyn, B., & Fisher, M. (2012). Performance of personal active noise reduction devices. *Applied Acoustics*, *73*(11), 1159–1167.

Rumsey, F. (2019). Headphone technology: Hear-through, bone conduction, and noise canceling. *Journal of the Audio Engineering Society*, *67*(11), 914–919.

Rybicka, W., & Sęk, A. (2003). Detection of the amplitude modulation for modulating signals characterised by different crest factors. *Archives of Acoustics*, *28*(4).

Sabin, A. T., Gallun, F. J., & Souza, P. E. (2013). Acoustical correlates of performance on a dynamic range compression discrimination task. *The Journal of the Acoustical Society of America*, *134*(3), 2136–2147.

Saiz-Alía, M., & Reichenbach, T. (2020). Computational modeling of the auditory brainstem response to continuous speech. *Journal of Neural Engineering*, *17*(3), 036035.

Saremi, A., Beutelmann, R., Dietz, M., Ashida, G., Kretzberg, J., & Verhulst, S. (2016). A comparative study of seven human cochlear filter models. *The Journal of the Acoustical Society of America*, *140*(3), 1618–1634.

Sayles, M., Füllgrabe, C., & Winter, I. M. (2013). Neurometric amplitude-modulation detection threshold in the guinea-pig ventral cochlear nucleus. *The Journal of Physiology*, *591*(13), 3401–3419.

Sayles, M., & Heinz, M. G. (2017). Afferent coding and efferent control in the normal and impaired cochlea. *Understanding the cochlea* (pp. 215–252). Springer.

Schlieper, R., Li, S., Preihs, S., & Peissig, J. (2019). The effect of active noise cancellation on the acoustic impedance of headphones. *2019 AES International Conference on Headphone Technology*.

Schlieper, R., Li, S., Preihs, S., & Peissig, J. (2021). Psychoacoustic optimization of a robust feedback active noise controller for headphones. *JASA Express Letters*, *1*(12), 124801.

Schlittenlacher, J., & Moore, B. C. (2016). Discrimination of amplitude-modulation depth by subjects with normal and impaired hearing. *The Journal of the Acoustical Society of America*, *140*(5), 3487–3495.

Schroeder, M. R. (1981). Modulation transfer functions: Definition and measurement. *Acta Acustica united with Acustica*, *49*(3), 179–182.

Schutz, M., & Gillard, J. (2020). On the generalization of tones: A detailed exploration of non-speech auditory perception stimuli. *Scientific Reports*, *10*(1), 1–14.

Shamma, S., & Lorenzi, C. (2013). On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. *The Journal of the Acoustical Society of America*, *133*(5), 2818–2833.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303–304.

Shiavi, R. (2010). *Introduction to applied statistical signal analysis: Guide to biomedical and electrical engineering applications*. Elsevier.

Shinn, P., & Blumstein, S. E. (1984). On the role of the amplitude envelope for the perception of [b] and [w]. *The Journal of the Acoustical Society of America*, *75*(4), 1243–1252.

Simpson, C., & King, R. (1997). Active noise reduction flight tests in military helicopters. *Audio Effectiveness in Aviation*, 22–1.

Slama, M. C., & Delgutte, B. (2015). Neural coding of sound envelope in reverberant environments. *Journal of Neuroscience*, *35*(10), 4452–4468.

Smith, S. W. (1999). *The scientist and engineer's guide to digital signal processing, second edition*. California Technical Publishing.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*(6876), 87–90.

Snyder, S. D. (2000). *Active noise control primer*. Springer Science & Business Media.

Sommerfeldt, S. D., & Samuels, T. O. (2001). Incorporation of loudness measures in active noise control. *The Journal of the Acoustical Society of America*, *109*(2), 591–599.

Song, S., & Abbott, L. (2000). Temporally asymmetric hebbian learning and neuronal response variability. *Neurocomputing*, *32*, 523–528.

Sosna, N. (2019). Micro-ecologies of sound: Elemental dialog at work. *RE: SOUND 2019–8th International Conference on Media Art, Science, and Technology 8*, 145–151.

Souza, P., Hoover, E., & Gallun, F. (2012). Application of the envelope difference index to spectrally sparse speech. *Journal of Speech, Language, and Hearing Research*, *55*(3), 824–837.

Steeneken, H. J. (1992). Quality evaluation of speech processing systems. *Digital speech processing: Speech coding, synthesis and recognition* (pp. 127–160). Springer.

Steeneken, H. J. (2005). *Assessment and standardization of personal hearing protection including active noise reduction*. Human Factors Research Institute TNO Soesterberg (Netherlands).

Steeneken, H. J., & Verhave, J. (1997). Personal active noise reduction with integrated speech communication devices: Development and assessment. *Audio Effectiveness in Aviation*, 18–1.

Stone, M. A., & Moore, B. C. J. (2007). Quantifying the effects of fast-acting compression on the envelope of speech. *The Journal of the Acoustical Society of America*, *121*(3), 1654–1664.

Struck, C. J. (2019). Objective measurements of headphone active noise cancelation performance. *2019 AES International Conference on Headphone Technology*.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech, Language, and Hearing Research*, *28*(3), 455–462.

Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., & Meddis, R. (2002). A revised model of the inner-hair cell and auditory-nerve complex. *The Journal of the Acoustical Society of America*, *111*(5), 2178–2188.

Sun, C., Liu, Y., & Wang, X. (2019). An automated hearing test equipment based on active noise control technology. *2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 1–5.

Swaminathan, J., & Heinz, M. G. (2012). Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. *Journal of Neuroscience*, *32*(5), 1747–1756.

Swift, S. H., Gee, K. L., Neilsen, T. B., Downing, J. M., & James, M. M. (2017). Exploring the use of time-sensitive sound quality metrics and related quantities for detecting crackle. *Proceedings of Meetings on Acoustics 173EAA*, *30*, 040003.

Taal, C. H., Hendriks, R. C., Heusdens, R., & Jensen, J. (2011). An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(7), 2125–2136.

Temme, S. (2019). Testing audio performance of hearables. *2019 AES International Conference on Headphone Technology*.

Temme, S., Olive, S., Tatarunis, S., Welti, T., & McMullin, E. (2014). The correlation between distortion audibility and listener preference in headphones. *Audio Engineering Society Convention 137*.

Toole, F. E. (1984). The acoustics and psychoacoustics of headphones. *Journal of the Audio Engineering Society*.

Toole, F. E. (1986a). Loudspeaker measurements and their relationship to listener preferences: Part 1. *Journal of the Audio Engineering Society*, *34*(4), 227–235.

Toole, F. E. (1986b). Loudspeaker measurements and their relationship to listener preferences: Part 2. *Journal of the Audio Engineering Society*, *34*(5), 323–348.

Trahiotis, C., & Bernstein, L. R. (2014). Advances in the understanding of binaural information processing: Consideration of the stimulus as processed. *Perspectives on auditory research*, 585–600.

Tran, B. W., Nowrouz, M. Y., Dhillon, S. K., Xie, K. K., Breslin, K. M., & Golladay, G. J. (2020). The impact of music and noise-cancellation on sedation requirements during total knee replacement: A randomized controlled trial. *Geriatric Orthopaedic Surgery & Rehabilitation*, *11*, 2151459320910844.

Trappenberg, T. (2009). *Fundamentals of computational neuroscience*. Oxford University Press.

Tsivian, M., Qi, P., Kimura, M., Chen, V. H., Chen, S. H., Gan, T. J., & Polascik, T. J. (2012). The effect of noise-cancelling headphones or music on pain perception and anxiety in men undergoing transrectal prostate biopsy. *Urology*, *79*(1), 32–36.

Turner, R. E. (2010). *Statistical models for natural sounds* (Doctoral dissertation). University College London. England.

Ueda, K., & Akagi, M. (1990). Sharpness and amplitude envelopes of broadband noise. *The Journal of the Acoustical Society of America*, *87*(2), 814–819.

Valimont, R. B., Casali, J. G., & Lancaster, J. A. (2006). ANR vs. passive communications headsets: Investigation of speech intelligibility, pilot workload, and flight performance in an aircraft simulator. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*, 2143–2147.

Valimont, R. B. (2006). *Active noise reduction versus passive designs in communication headsets: Speech intelligibility and pilot performance effects in an instrument flight simulation* (Doctoral dissertation). Virginia Polytechnic Institute and State University. United States.

van de Par, S., Disch, S., Niedermeier, A., Burdiel Pérez, E., & Edler, B. (2019). Temporal envelope-based psychoacoustic modelling for evaluating non-waveform preserving audio codecs. *Audio Engineering Society Convention 147*.

Vecchi, A. O., Varnet, L., Carney, L. H., Dau, T., Bruce, I. C., Verhulst, S., & Majdak, P. (2021). A comparative study of eight human auditory models of monaural processing. *arXiv preprint arXiv:2107.01753*.

Venezia, J. H., Hickok, G., & Richards, V. M. (2016). Auditory "bubbles": Efficient classification of the spectrotemporal modulations essential for speech intelligibility. *The Journal of the Acoustical Society of America*, *140*(2), 1072–1088.

Verhey, J. L., Pressnitzer, D., & Winter, I. M. (2003). The psychophysics and physiology of comodulation masking release. *Experimental Brain Research*, *153*(4), 405–417.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *The Journal of the Acoustical Society of America*, *66*(5), 1364–1380.

Viemeister, N. F. (2014). Temporal processing: Observations on the psychophysics and modeling of temporal integration and temporal resolution. In A. N. Popper & R. R. Fay (Eds.), *Perspectives on auditory research* (pp. 601–610). Springer.

Vinay, S., & Moore, B. C. (2010). Effects of the use of personal music players on amplitude modulation detection and frequency discrimination. *The Journal of the Acoustical Society of America*, *128*(6), 3634–3641.

Vorländer, M. (2000). Acoustic load on the ear caused by headphones. *The Journal of the Acoustical Society of America*, *107*(4), 2082–2088.

Wagstaff, A., & Woxen, O. (1997). Effects of active noise reduction on noise levels at the tympanic membrane. *Audio Effectiveness in Aviation*, 17–1.

Wakefield, G. H., & Viemeister, N. F. (1990). Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise. *The Journal of the Acoustical Society of America*, *88*(3), 1367–1373.

Wang, H., Kawai, T., Iwai, K., & Nishiura, T. (2019). Discomfort reduction for residual noise of active noise control system using noise shaping filter [Issue: 7]. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, *259*, 2645–2653.

Wang, X. (2007). Neural coding strategies in auditory cortex. *Hearing Research*, *229*(1), 81–93.

Watanabe, Y., & Hamada, H. (2011). Binaural technique used for active noise control assessment. *Principles and applications of spatial hearing* (pp. 494–503). World Scientific.

Wirtzfeld, M. R., Ibrahim, R. A., & Bruce, I. C. (2017). Predictions of speech chimaera intelligibility using auditory nerve mean-rate and spike-timing neural cues. *Journal of the Association for Research in Otolaryngology*, *18*(5), 687–710.

Yin, P., Johnson, J. S., O'Connor, K. N., & Sutter, M. L. (2011). Coding of amplitude modulation in primary auditory cortex. *Journal of Neurophysiology*, *105*(2), 582–600.

Zacharov, N. (2019). *Sensory evaluation of sound*. CRC Press.

Zacharov, N., Ramsgaard, J., Le Ray, G., & Jørgensen, C. V. (2010). The multidimensional characterization of active noise cancelation headphone perception. *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)*, 130–135.

Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., & Cao, K. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings of the National Academy of Sciences*, *102*(7), 2293–2298.

Ziemer, T., Kiattipadungkul, P., & Karuchit, T. (2020). Music recommendation based on acoustic features from the recording studio. *The Journal of the Acoustical Society of America*, *148*(4), 2701–2701.

Zilany, M. S., Bruce, I. C., & Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical Society of America*, *135*(1), 283–286.

Zilany, M. S., Bruce, I. C., Nelson, P. C., & Carney, L. H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. *The Journal of the Acoustical Society of America*, *126*(5), 2390–2412.

Zuccatti, C., & Bandiera, M. (2009). Dynamics distortion: Loudspeaker sensitivity modulation generated by audio signals. *Journal of the Audio Engineering Society*, *57*(5), 338–348.

Zwicker, E., & Fastl, H. (2013). *Psychoacoustics: Facts and models* (Vol. 22). Springer Science & Business Media.