A STUDY ON THE EFFECTS OF HEARING-AID FITTING PROCEDURES ON SPATIAL PERCEPTION

Ticiana Matar De Lello

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Luiz Wagner Pereira Biscainho

Rio de Janeiro
Novembro de 2022

A STUDY ON THE EFFECTS OF HEARING-AID FITTING PROCEDURES ON SPATIAL PERCEPTION

Ticiana Matar De Lello

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Orientador: Luiz Wagner Pereira Biscainho

Aprovada por: Prof. Luiz Wagner Pereira Biscainho
Profa. Mariane Rembold Petraglia
Prof. Márcio Nogueira de Souza

RIO DE JANEIRO, RJ – BRASIL
NOVEMBRO DE 2022

Matar De Lello, Ticiana

A Study on the Effects of Hearing-Aid Fitting Procedures on Spatial Perception/Ticiana Matar De Lello. – Rio de Janeiro: UFRJ/COPPE, 2022.

XVI, 155 p.: il.; 29, 7cm.

Orientador: Luiz Wagner Pereira Biscainho

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2022.

Referências Bibliográficas: p. 124 – 131.

1. Hearing aids. 2. Spatial perception. 3. Binaural cues. I. Pereira Biscainho, Luiz Wagner. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Aos meus sobrinhos, Pedro e Beatrice.

Agradecimentos

Aos meus pais, Marcia e Paulo, pelo amor e pelo apoio incondicionais que tive durante todo este processo.

Aos meus irmãos, Ricardo e Guilherme, pela amizade e pelas palavras motivadoras em todos os períodos difíceis.

A todos os meus outros familiares, que sempre acreditaram em mim, e por cujo carinho sou grata em todas as ocasiões.

Ao Werther, por ter sido meu porto seguro durante este processo e me motivado a ser cada dia uma versão melhor de mim mesma.

A Silvia e Isabella, que tornaram tudo isto possível com sua presença, seu cuidado e seu afeto.

A Pedro Garcia, Carolina Silva, Camila Bulcão, Camila Schultz, Mariana Doroteu, Bárbara Fraga, Diana Moreira, Isadora Sampaio, Marina Ruppelt, Carolina Nogueira e a todos os meus amigos, por me encherem de alegria, otimismo e esperança.

Ao meu orientador, Prof. Luiz Wagner Biscainho, que está ao meu lado na realização de mais uma etapa na minha trajetória acadêmica, com sua paciência, sua amizade, sua compreensão e sua excelência.

A todas as pessoas que me auxiliaram neste trabalho como participantes dos testes subjetivos.

Aos meus colegas de laboratório, que tornaram minha rotina muito mais leve durante a realização do trabalho.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

UM ESTUDO DOS EFEITOS DA COMPENSAÇÃO EM APARELHOS AUDITIVOS SOBRE A PERCEPÇÃO ESPACIAL

Ticiana Matar De Lello

Novembro/2022

Orientador: Luiz Wagner Pereira Biscainho

Programa: Engenharia Elétrica

Este trabalho visa a estudar a influência da compressão e da amplificação por faixa de frequência, duas etapas comuns de processamento de sinais em aparelhos auditivos, nas pistas biauriculares e na percepção espacial. Experimentos foram conduzidos com sinais de fala convoluídos com respostas ao impulso relacionadas à cabeça gravadas em uma câmara anecoica e em dois ambientes reverberantes. Para cada tipo de ambiente, três análises teóricas foram feitas para verificar, respectivamente, a influência do efeito isolado da compressão de canal único, do efeito da amplificação linear, e do efeito da amplificação não-linear na ITD, na ILD e na IC. Para as duas últimas análises, perdas auditivas unilaterais e bilaterais foram simuladas com o uso de dois audiogramas de referência. Dois testes subjetivos foram aplicados em um grupo de participantes para análisar suas percepções referentes ao azimute e ao espalhamento do som utilizando-se sinais anecoicos e reverberantes submetidos à compressão de canal único e à amplificação não-linear para algumas configurações de perda auditiva. As três análises teóricas evidenciaram mudanças nas pistas biauriculares, especialmente na ILD, alterações de timbre e alguma degradação na percepção espacial, especialmente quando as modificações são feitas na orelha mais próxima da fonte sonora. Os testes subjetivos mostraram que a percepção espacial é afetada pela posição da fonte, pelo tipo de ambiente, pelo limiar de compressão e pela configuração de perda auditiva. De maneira geral, os sinais modificados não alteraram de forma drástica a percepção do azimute, mas causaram uma sensação de maior espalhamento do som. As perdas auditivas desbalanceadas foram as que mais prejudicaram a percepção espacial.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

A STUDY ON THE EFFECTS OF HEARING-AID FITTING PROCEDURES ON SPATIAL PERCEPTION

Ticiana Matar De Lello

November/2022

Advisor: Luiz Wagner Pereira Biscainho

Department: Electrical Engineering

This work is a study about the influence of compression and frequency-selective amplification, two common signal processing steps used in hearing aids, on the binaural cues and the spatial perception. Experiments were conducted with speech signals convolved with HRIRs recorded in an anechoic chamber and in two reverberant environments. For each environment type, three theoretical analyses were performed: the influence of the isolated effect of single-channel compression, of the effect of linear amplification, and of the effect of non-linear amplification on the ILD, the ITD and the IC. For the former two analyses, unilateral and bilateral hearing loss configurations were simulated using two different reference audiograms. Two listening tests were applied to analyze the listeners' perception of azimuth and of the spreading of sound for anechoic and reverberant signals submitted to single-channel compression and to non-linear amplification with several hearing loss configurations. The three theoretical analyses showed that there were some changes in the binaural cues, especially in the ILD, alterations in sound quality and some degradation of spatial perception, especially if the modification of the original signal is applied to the ear which is closer to the sound source. The results of the listening tests showed that the spatial perception was affected by the source position, the type of environment, the compression threshold and the hearing loss configuration. In general, the modified signals did not contribute to a drastic change in azimuth perception, but caused the sensation of a more spread-out sound. The unbalanced hearing loss configurations were the most harmful to localization perception.

Contents

List of Figures

List of Tables

List of Abbreviations

BTE	Behind-the-Ear, p. 14
CAPD	Central Auditory Processing Disorder, p. 11
CIC	Completely-in-the-Canal, p. 14
HATS	Head-and-Torso Simulator, p. 38
HRIR	Head-Related Impulse Response, p. 30
HRTF	Head-Related Transfer Function, p. 30
IC	Interaural Coherence, p. 35
IIC	Invisible-in-the-Canal, p. 14
ILD	Interaural Level Difference, p. 31
ITC	In-the-Canal, p. 14
ITD	Interaural Time Difference, p. 31
ITE	In-the-Ear, p. 14
ITF	Interaural Transfer Function, p. 31
MWF	Multichannel Wiener Filter, p. 36
RIC	Receiver-in-the-Canal, p. 14
RITE	Receiver-in-the-Ear, p. 14
RMS	Root-Mean-Square, p. 40

Chapter 1

Introduction

The sense of hearing is a complex and fascinating topic within the human fisiology. It played an essential role in the evolution of the human species, making verbal communication, awareness of predators, hunting and spatial orientation possible, and continues to be an essential factor that makes most peoples' lifestyles possible. Nevertheless, hearing loss is a condition that affects more people than one can usually imagine. According to the World Health Organization, over 1.5 billion individuals (nearly 20% of the world population) live with some degree of hearing loss nowadays [1]. Out of this amount, around 430 million are considered to have a disabling hearing ability – that is, a degree of hearing loss that has a very significant negative impact on everyday life. By 2050, the number of individuals with hearing loss can increase up to 2.5 billion, out of which 700 million people will represent a disabling hearing ability – that is approximately one out of ten people [1]. These numbers require a very special attention, since it has been proven that hearing impairment has considerably negative impacts on language development and communication, cognition, school and academic performance, risk of unenployment or underemployment, and also contribute to lower quality of life, social isolation, mood disorders, poor general health and even an increased mortality risk [2, 3].

With these numbers and this scenario in mind, one can easily see the importance and urgency of hearing rehabilitation. There are several procedures than can be used to improve the lifestyle of hearing-impaired listeners, and they range from external auditory prostheses to surgical interventions, depending on the characteristics of each subject's hearing loss. A common type of auditory prostheses are hearing aids, which are small electronic circuits that amplify and modify sounds before they enter the hearing-impaired person's ears in order to make those sounds more audible. There are a lot of hearing aid models for different purposes and types of hearing loss, and the hearing aid technology has impressively improved in the last decades. However, it is known that, currently, only 17% of the population that requires hearing aids actually use one [2]. Out of the percentage of people that indeed own

one or two hearing aids, it is well known that many of them simply do not use them as they should, as observed in a research run by Dillon et al. with the hearing-impaired Welsh population [4], which indicated that 20% of the observed hearing aid users did not use their hearing aids at all, while 30% used them only some of the time.

This apparent difficulty in wearing hearing aids has been reported to have several reasons, such as poor effectiveness of the aids in noisy situations, poor benefit, poor sound quality, hearing aid not suitable for the user's hearing loss, comfort, and difficulty in putting the hearing aids in and taking them out [3]. It is highly likely that this lack of effectiveness provided by the hearing aids also have to do with a distorted spatial hearing perception. Several works regarding sound source localization ability of hearing aid users have been published, and it was seen that aided localization is poorer than that of normal-hearing listeners, and in some cases may be even poorer than unaided localization when tested at the same sensation level [5]. Reasons for that were associated to some characteristics present in modern hearing aids, such as compression, noise reduction strategies and directional microphones [6, 7]. The mere presence of the hearing aid itself is also associated to the distortion of some spectral cues which help localization [5]. When the hearing aid is placed behind the ear, for example, the signal captured by the microphones do not suffer the influence of the pinna cues, which have an important role in sound localization. Besides that, the use of occlusion of the ear canal caused by some hearing aid models reduces the magnitude of high-frequency sounds that could signal a change in elevation of the input signal. Furthermore, some compensatory strategies developed by hearing-impaired people might be damaged with the use of hearing aids. Since hearing loss deprives the brain of the necessary acoustic inputs, this leads the listener to develop cognitive or functional compensatory strategies (e.g. head and body movements, visual searching) to help them localize that do not require the correct binaural or monaural cues. With the hearing aids, the newly amplified signals may not be "recognizable" to the brain, which might lead to continued localization errors, unless the brain is retrained to interpret the new input [5].

In this work, it was intended to focus on the influence of compression and amplification strategies on the binaural cues and on the localization ability, since these steps are always present in the signal processing chain of modern hearing aids. This was also motivated by the numerous attempts in creating noise reduction techniques that present a satisfactory performance in reducing noise while also creating the smallest possible distortion on the binaural cues [8–11]. Since the noise reduction step is often performed before the alterations in the signal due to the amplification strategies and compression, this work raises the question whether the binaural cues which are attempted to be preserved during the noise reduction procedure are

once again distorted by these processing strategies that come next. Several works have evaluated the effect of noise reduction on the binaural cues, but not many have focused on the effect on the binaural cues and the localization ability caused by compression and amplification strategies themselves. Keidser et al. [6] aimed to verify the individual effects of multichannel wide dynamic range compression, noise reduction and directional microphones on the binaural cues and localization perception on twelve participants fitted with hearing aids. Hassager et al. [7], on the other hand, attempted to analyze the effect of independent and synchronized wide dynamic range compression between the left and right channels on spatial perception in a reverberant environment, by presenting the signals to normal-hearing and hearing-impaired listeners through headphones. While Keidser et al. found that wide dynamic range compression did not significantly affect localization performance, in spite of altering the interaural level differences, Hassager et al. stated that both independent and synchronized fast-acting compression resulted in more diffuse and broader sound images, internalization, and image splits relative to linear processing. With this in mind, this work provides another attempt at verifying the isolated effects of compression and linear amplification on sound localization in anechoic and reverberant environments, both through an objective perspective regarding the preservation of the binaural cues and a subjective perspective regarding the perceptions of the listeners in terms of source location and the precision of the sound images.

It is hoped that the phenomena analyzed in this dissertation enables a better understanding of the effects of linear and non-linear amplification strategies used in hearing aids in localization performance, in order to verify if state-of-the-art noise reduction strategies and other processing techniques with binaural cue preservation are indeed helping localization performance or if further adjustments later in processing chain of the hearing aid should still be studied in order to minimize possible detrimental effects on spatial perception caused by frequency-selective amplification and compression.

1.1 Objectives and methods

The general objective of this work is to analyze the influence of compression and amplification strategies in hearing aids on sound localization. It can be segmented into the following specific objectives:

- Analyze the isolated influence of a single-channel compression strategy on the binaural cues and on the spatial perception;
- Analyze the effect of a linear frequency-selective amplification strategy (that is,

with no compression applied, as it was done for the older hearing aid models) on the binaural cues and on the spatial perception;

- Analyze the effect of a non-linear frequency selective amplification strategy (that is, an amplification procedure which applies compression, as is used in modern hearing aids) on the binaural cues and on the spatial perception;

- Compare the subjective perceptions regarding sound localization to the binaural cue measurements, in order to see if a change in spatial perception implies a correspondent change on the binaural cues, or if a change in the binaural cues is detected from a subjective perspective.

In order to achieve these goals, the chosen signals for analysis were several spoken sentences from a speech database developed in the Signals, Multimedia and Telecommunications Laboratory from the Federal University of Rio de Janeiro, which is described in more detail in Chapter 4 and in Appendix B. The signals were converted from mono to stereo and resampled from 48 kHz to 16 kHz. They where then convolved with several head-related impulse responses from the Kayser et al. HRIR databse [12] recorded at several positions in an anechoic chamber and in two reverberant environments. Since there was no feasible possibility to reproduce the signals in actual hearing aids, all the alterations in the original signals were done via software with the use of MATLAB ® and the two-channel signals were always reproduced via headphones.

The first theoretical experiment was conducted with the signals spatialized with the HRIRs from the anechoic chamber. Firstly, a single-channel compression strategy was implemented and applied to the signals, and the binaural cues for both the original and the output signals were calculated and analyzed. Secondly, the same unprocessed spatialized signals were altered by a linear amplification procedure implemented with the use of a cosine-modulated filter bank, and the binaural cues for the new output signals were also obtained and compared to the original ones. Lastly, the spatialized sentences were altered instead by a non-linear amplification procedure — an amplification procedure that applies compression — implemented with the same filter bank and the binaural cues for the output signals were calculated and compared to the original ones. The two latter procedures received as input simulated bilateral and unilateral hearing loss configurations by using two audiograms with different degrees of hearing loss as reference.

The same three analyses were repeated in the same way for the second theoretical experiment, which used the signals spatialized with the HRIRs from two reverberant environments (an office and a cafeteria). The choice of using HRIRs from three different environments (anechoic chamber, office and cafeteria) allowed an overview of the effect of compression and amplification strategies on the binaural cues in

situations with and without reverberation and an analysis of the role of reverberation on the binaural cues and spatial perception.

After the theoretical experiments were conducted, the subjective perceptions regarding sound localization were analyzed. Two sentences of the same speech database used for the theoretical experiments were convolved with HRIRs of the Kayser database at several positions in the anechoic chamber and in the cafeteria and then used for the creation of two listening tests. The listening tests were performed with 23 normal-hearing participants, which listened to several reference and altered spatialized signals through headphones, and must answer questions regarding the sound source position and the sensation of how "diffuse" was the sound image. With the first listening test, it was aimed to evaluate the isolated effect of compression on sound localization — the subjective counterpart of the first analyses performed in the two theoretical experiments. Therefore, the spatialized signals were presented without compression and with two different compression configurations. The second test, on the other hand, was intended to verify the combined effect of compression and frequency-selective amplification on sound localization with a non-linear amplification strategy. For that, reference and altered versions of the spatialized signals were also used. The altered signals simulated a non-linear amplification procedure applied to bilateral, unilateral and unbalanced hearing loss configurations by using two audiograms with different degrees of hearing loss.

Lastly, statistical analyses with the listening tests' results were conducted. The influence of factors such as environment, compression threshold, hearing loss configuration, listener experience, talker and position on the test answers regarding the source position and the perceived "diffusiveness" of sound were analysed. The test answers for the source position were compared to what the binaural cues were indicating for the same signals, in order to verify if there was a coherence between the measured binaural cues and the perceived source positions, as well as the subjective sensitivity to changes in the binaural cues caused by the alterations applied to the original spatialized signals. The conclusions were then formed based on the most important findings regarding the theoretical experiences and the listening test results.

1.2 Chapter organization

This work is divided into six chapters and two appendices. After the overview presented in this chapter, considerations regarding the main aspects of hearing loss, hearing aid models, compression and amplification strategies are addressed in Chapter 2. Chapter 3, on the other hand, explains some aspects of spatial hearing and introduces the mathematical descriptions of the binaural cues (interaural level dif-

ference and the interaural time difference) and the interaural coherence. It also briefly explains the importance of the monaural cues for sound localization and the existence of cones of confusion. Chapter 4 describes in detail how the two theoretical experiments were performed and presents all the analyses and results regarding them. Chapter 5 focuses on describing the preparation and application of the two listening tests and discusses its results, while also presenting a comparison between the test results and the binaural cues calculated for the signals used in the tests. Lastly, Chapter 6 presents all the conclusions and addresses possible perspectives for future works.

Appendix A presents the floors plans of the environments where the HRIRs from the Kayser database were recorded, with indications of the positions of the loudspeakers and the head-and-torso simulator. Also in Appendix A are additional figures containing the binaural cue values across frequencies for some signals evaluated in Chapter 4. Appendix B, on the other hand, presents a compilation of descriptions of several Brazilian Portuguese speech databases, including the one used in this work. Each database has particular characteristics regarding the type of environment where the signals were recorded, the variety of regional accents of the speakers, and the audio quality. This compilation was originally done for another work that was being developed in the Signals, Multimedia and Telecommunications Laboratory concerning dereverberation of speech signals. The work was paused, but the databases' descriptions were nevertheless considered worthy of documenting. Therefore, they were inserted in this dissertation in Appendix B.

Chapter 2

Hearing loss and hearing aids

Human hearing is a fascinating topic that can be better understood by describing the physiology of the human auditory system. Within the auditory system, there is a flow of information that comprises the arrival of sound at the ear, its conversion into electrical stimuli, the transmission of those electrical stimuli to the brain by the auditory nerves, and, finally, the processing and interpretation of those stimuli by the brain. It is a complex mechanism that involves multiple aspects, which are here described in more detail.

Firstly, it is necessary to understand the anatomy of the ear. The human ear can be divided into three main parts, each with its own specific structure and function: the outer ear, the middle ear, and the inner ear. As can be seen in a frontal cut of the human ear in Figure 2.1, the outer ear consists of the pinna, the ear canal and the eardrum; the middle ear, of three ossicles connected to the eardrum — hammer (malleus), anvil (incus) and stirrup (stapes); and the inner ear, of the cochlea, which is a spiral-shaped organ filled with a watery fluid called endolymph.

Air-borne sound enters the outer ear and passes through the pinna and the ear canal, causing the eardrum to oscillate by absorbing the incoming sound. The ear canal has two important advantages: protecting the eardrum and the middle ear from damage, and enabling the inner ear to be positioned very close to the brain, which reduces the length of the nerves and results in a short travel time for the action potentials in the nerve [14]. The eardrum, on the other hand, is a thin cone-shaped membrane that separates the outer ear from the middle ear. An important aspect to be considered is that the eardrum does not seal the ear canal completely reflection-free, and this causes the canal to act as a $\lambda/4$-resonator with a resonance frequency of ca. 4000 Hz. Due to this, human hearing is particularly sensitive at the region around that frequency [14].

When the eardrum absorbs the incoming sound and starts vibrating, this vibration is then transmitted to the hammer, the anvil and the stirrup in the middle ear. The three ossicles act as a mechanical lever by providing a large amplification of

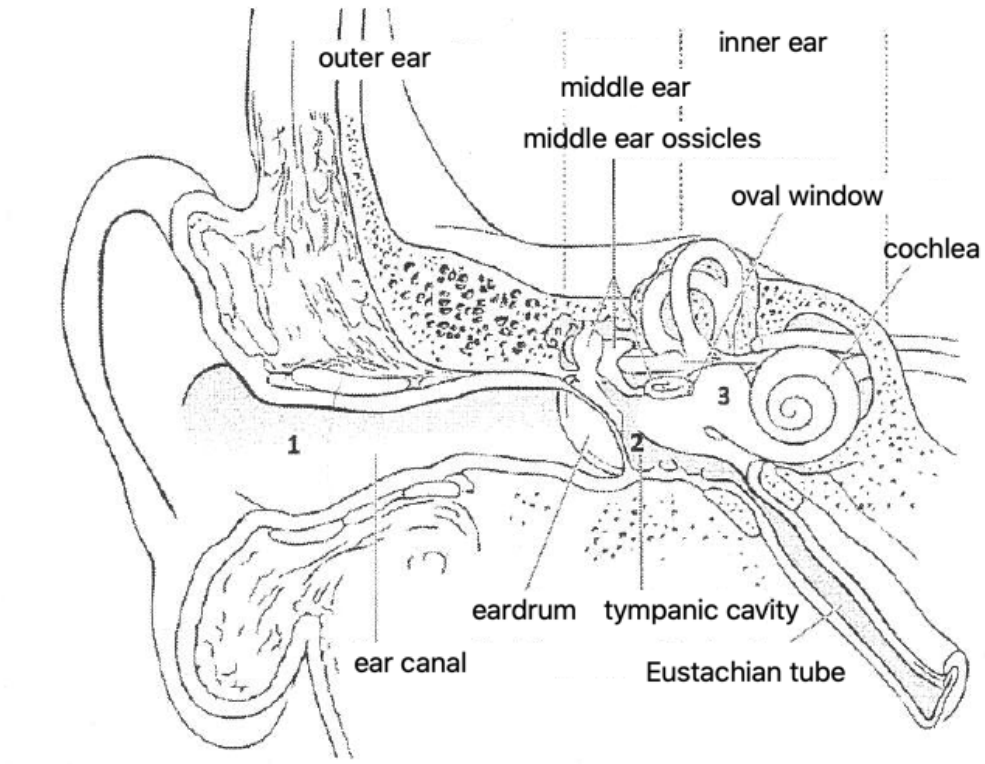


Figure 2.1: Structure of the ear. Adapted from Fels [13].

pressure due to the differences between the lengths of the hammer and the anvil, as well as the difference in area between the surface of the eardrum and the footplate of the stirrup, which are shown in Figure 2.2. With this mechanism, the middle ear provides the matching of acoustic impedance between the outer ear, which is filled with air particles, and the inner ear, which is filled with fluid. The impedance matching is necessary in other to avoid energy losses due to the difference between the acoustic impedances of the air and the fluid [14]. Figure 2.2 illustrates the three ossicles of the middle ear.

Finally, the footplate of the stirrup transmits the oscillation to the fluid in the cochlea via a small entrance called oval window. The cochlea and the semicircular canals are the organs of the inner ear. The semicircular canals are the structure responsible for balance, while the cochlea has the role of detecting the individual frequencies that form the sound wave and of converting the physical sound into electrochemical impulses. The cochlea has a length of about 30-35 mm when uncoiled, and three distinct parallel running canals: (*scala vestibuli*, *scala tympani* and *ductus cochlearis*) [13]. They are all bordered by the basilar membrane, where the organ of Corti, also called spiral organ, is localized. The cross section of the cochlea and the organ of Corti can be seen in Figure 2.3.

When the vibrations from the stapes enter through the oval window, they are

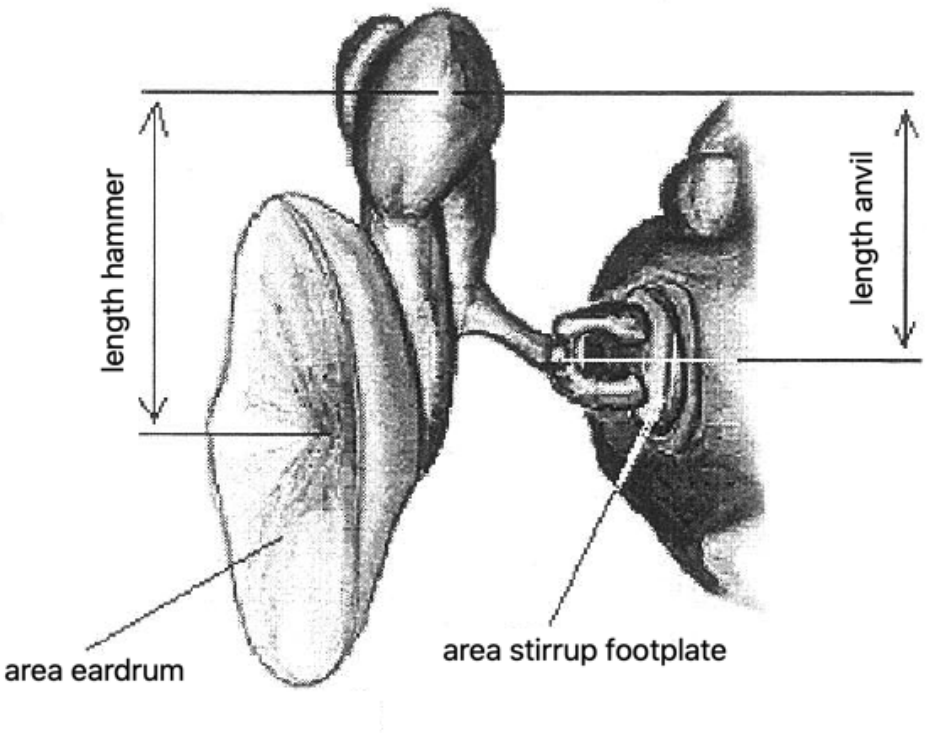


Figure 2.2: Length and surface relations in the middle ear. Adapted from Fels [13].

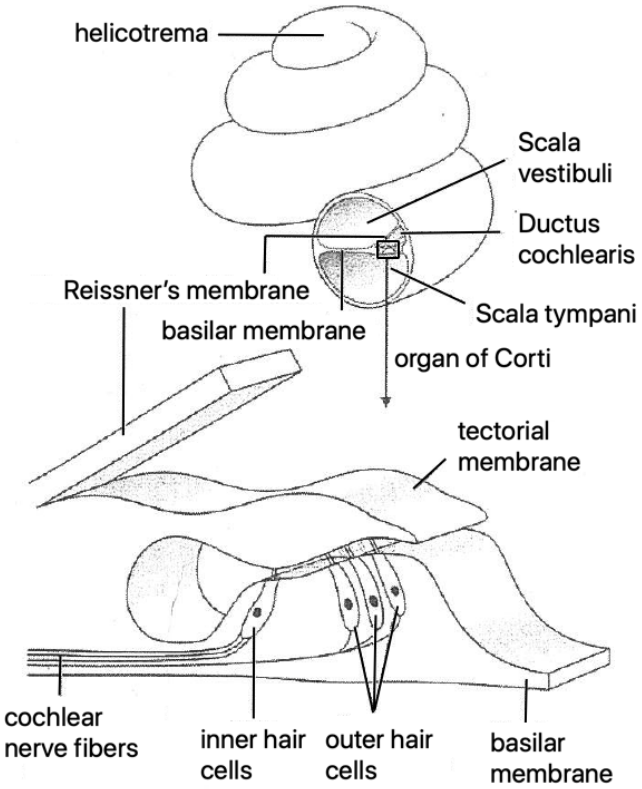


Figure 2.3: Cross section of the cochlea and enlarged section of the organ of Corti. Adapted from Fels [13].

propagated inside the endolymph, causing the basilar membrane to vibrate due to a difference in pressure between the *scala vestibuli* and the *scala tympani*. This vibration is propagated all the way from the oval window to the apex of the cochlea (called helicotrema), and it shows a maximum deflection at a certain position which varies with the sound frequency. Due to this reason, different frequencies are detected at different regions of the basilar membrane, with high frequencies being detected closer to the oval window and low frequencies generating a maximum closer to the helicotrema. The convertion of the vibrations into electrochemical impulses is done by the hair cells located at the organ of Corti, which are triggered by a relative motion of the basilar membrane and the tectorial membrane at the region of the maximum [13]. The electrochemical pulses are then finally transmitted to the auditory nerve, which forwards the information all the way to the brain, where sound is finally processed and interpreted.

When a subject has a hearing loss, one or more parts of this flow of information from the outer ear to the brain is impaired in some way, as will be seen in the following section.

2.1 Hearing loss and its implications

Hearing loss or hearing impairment is a multifaceted partial or total loss of hearing ability. It may happen due to a multitude of causes, which can be congenital (arising before or shortly after birth) or acquired at any moment of one's life. Depending on which area of the ear is damaged, the hearing loss can be classified as conductive, sensorineural, combined, or central.

Conductive hearing loss occurs when air-borne sound is damped prior to entering the inner ear. It can be caused by a blocking of the outer ear canal with earwax or a foreign particle, ventilation problems due to infection or inflammation of the middle ear, otosclerosis (a restriction of the mobility of the stapes caused by age-related porosity of the bones), and some anatomical deformities of the outer or middle ear [15]. Besides the use of several types of hearing aids as a therapeutic solution, this type of hearing loss may be cured in some cases by surgical measures or appropriate medication.

Sensorineural hearing loss, on the other hand, occurs when either the transformation of mechanical acoustic signals into electrical neural signals in the cochlea or the transportation of the electric signals to the auditory nerve is disrupted. It can have congenital causes, such as genetic factors, maternal infections like syphilis and rubella during pregnancy, use of particular medications or drugs that can be damaging to the fetus, and birth asphyxia; or it can also be acquired at a later moment in life due to infectious diseases, injury to the head or ear, particular drugs, expo-

sure to noise, and aging [16]. This is the most common type of permanent hearing loss [15, 17]. Since it consists of damage of the inner or outer hair cells inside the cochlea, and these are not able to regenerate, sensorineural hearing loss is usually irreversible, although individuals with it can largely benefit from the use of hearing aids or cochlear implants. If a conductive as well as a sensorineural hearing loss is diagnosed, the subject is said to have a combined hearing loss.

Lastly, central hearing loss is caused by a dysfunction at the central auditory pathway and or at the auditory cortex [18]. It provides complex disease patterns involving the understanding of speech, noise suppression and timbre recognition. It may be caused by an infarction, hemorrhage, tumor, and multiple sclerosis in adults, among other factors, and by "central auditory processing disorder" (CAPD) in children [18]. Therapeutical treatments have not yet been verified satisfactorily.

Contrary to popular belief, hearing loss usually does not simply imply listening to all sounds at a "lower volume" than normal-hearing people do, or not listening at all. Particularly for the case of sensorineural hearing loss, except when noted, all of these characteristics apply, according to Dillon [15]:

1. Decreased audibility: it is very common for hearing-impaired people to hear some sounds better than others. When it comes to speech, for example, a person with mild or moderate hearing loss may say that "speech is loud enough, but not clear enough" — that is, they can hear the voice itself, but have trouble understanding what is being said, and often make some confusions (e.g. "pick the black harp" could be heard as "kick the cat hard"). That occurs because essential parts of some phonemes are not audible to them. Phonemes have harmonics in several frequency ranges that are necessary for them to be distinguished from other phonemes and, if the person has decreased audibility in a particular frequency range, that is no longer possible. Most people have the greatest loss at the high frequencies, where these harmonics are usually located. Thus, hearing aids must provide a different amount of amplification at different frequency regions.

2. Decreased dynamic range: While soft sounds can be made audible merely by amplifying them, it is not appropriate to amplify all sounds by the same amount. A sensorineural hearing loss increases the threshold of hearing much more than it increases the threshold of loudness discomfort. That means the dynamic range of an ear with sensorineural impairment is smaller than that of a normal-hearing one. Therefore, the same increase in sound level will produce a bigger loudness increase for a hearing-impaired person than for a normal-hearing person. This phenomenon is known as recruitment.

3. Decreased frequency resolution: as stated earlier in this chapter, each region

of the cochlea specializes in detecting sound in particular frequency ranges. Sensorineural hearing loss makes this frequency selectivity less precise. This implies that two distinct nearby regions in the cochlea which would normally detect two different sounds of nearby frequencies may now be damaged and function as a single region, which will be simultaneously stimulated by both sounds. As a consequence, the brain is unable to separate these two frequencies. This is a reason why hearing-impaired people have greater trouble separating speech from noise. If frequency resolution is sufficiently decreased, it can adversely affect speech intelligibility even without the noise, given that relatively intense low-frequency parts of speech may mask the weaker higher frequency components, which are essential to the distinction of consonants. That is, neural fibers which would normally respond in synchrony with these higher frequency components are instead captured by the waveforms of other lower harmonics. This phenomenon is known as upward spread of masking [19, 20]. Lastly, it is worth noting that frequency selectivity is worse at higher sound pressure levels. Thus, people with severe and profound hearing loss may have an even poorer frequency resolution due to their already damaged cochlea as well as their need to listen at higher levels.

4. Increased temporal masking: Temporal masking describes the phenomenon of intense sounds masking weaker sounds that immediately precede or follow them. It happens to a greater extent for people with sensorineural hearing impairment than for normal-hearing people, and negatively affects speech intelligibility. While listening to speech in a fluctuating background noise scenario, normal-hearing people make use of the weaker moments of the background noise to extract useful snippets of information that help them understand speech. This ability is referred to as "listening in the gaps". Hearing impaired people partially lose this ability to hear weak sounds during the gaps in a masking noise, and it decreases as hearing loss gets worse [21].

5. Deficit in spatial release from masking: The binaural auditory processing system in normal-hearing people enables them to focus attention to sounds coming from one direction and suppress sounds from other directions. This is particularly important for many daily situations, such as separating the target speech that needs to be heard from competing speech or noise coming from a different direction. This ability to separate the target signal from a competing signal based on the direction of arrival is called spatial release from masking, and is adversely affected by the cochlear distortions occuring in sensorineural hearing loss. The increased difficulty in separating sounds coming from different directions can be a major disadvantage for hearing-impaired people and has not

yet been completely diminished by the use of hearing aids, as will be discussed in later parts of this text. It is worth noting that reduced spatial listening ability is also present in people with conductive hearing loss, but for different reasons. A conductive hearing loss can be interpreted as mere attenuation of sound; therefore, a hearing aid that can provide adequate amplification allows the normal cochlea of people with conductive loss to resolve sounds as well as the cochlea of someone with normal hearing. However, as conductive loss increases, the proportion of sound reaching the cochlea by bone conduction also increases. The sounds coming from bone conduction in each ear are more similar to each other, and the ability of the brain to combine these two signals to selectively attend to sounds coming from one direction decreases. Hearing aids increase the proportion of sound received by air conduction, but some mixing with sounds coming from bone conduction for large conductive losses is likely to remain, thus reducing spatial listening abilities in people with conductive loss.

It can be seen that hearing loss is in many cases a much more complex and multidimensional phenomenon than it may be perceived by common sense. Therefore, hearing aids must do significantly more than merely amplifying sounds. Firstly, they must be able to provide different amounts of gain in different frequency bands, so that it can better fit to the hearing loss pattern of each user, and deliver more amplification in the frequency regions where the loss is greater. However, the decreased dynamic range of the individual with sensorineural impairment cannot be simply solved by amplifying all frequency bands by the same amount for all input levels.

The hearing aids must therefore function in a way that weak sounds are given more amplification than intense sounds. This squashing of a large dynamic range of levels in the environment into a smaller range of levels at the output of the hearing aid is called compression. It is used in practically all hearing aids nowadays and is one important feature that will be better explained in the next section.

Besides gain and compression, hearing aids must also increase the signal-to-noise ratio (SNR) of sounds reaching the ear, due to all increased difficulties faced by hearing-impaired people to spatially separate speech from noise, understand weaker higher harmonics of speech, and listen to speech in gaps during fluctuating background noise. This can be done by noise-reducing strategies inside the hearing aid and by the use of directional microphones, as will also be seen later in this chapter.

2.2 Introduction to hearing aids

A hearing aid is basically a system that captures an acoustic signal, amplifies it (and makes other necessary processing), and delivers the modified acoustic signal to the user's ear. Its essential components are [15]:

- one or more microphones to convert sound into an electrical signal;
- an amplifier to increase the strength of the electrical signal, providing different amounts of gain to different frequency regions and different input levels;
- a miniature loudspeaker, called a receiver, to convert the electrical signal back into sound;
- a means of coupling the amplified sound into the ear canal;
- a battery to provide the power needed by the amplifier.

Some common types of hearing aids are BTE (behind-the-ear), ITE (in-the-ear), ITC (in-the-canal), CIC (completely-in-the-canal) and IIC (invisible-in-the-canal) hearing aids.

In BTE hearing aids, which are the most common, the microphones and electronics are mounted in the characteristic banana-shaped case, or in some artistic variations of it, and the receiver is often also mounted in the case. The case usually sits behind the ear of the user, hence the name BTE. The sound from the receiver is then conveyed via an acoustic tube to an earmold that is placed in the ear canal entrance. A variation of the BTE hearing aid is the RITE (receiver-in-the-ear) or RIC (receiver-in-the-canal) BTE, in which the receiver is located within the ear canal rather than in the BTE case, and an electrical cable rather than an acoustic tube runs from the case to the ear canal.

The next type of hearing aid is the ITE. These are not positioned behind the ear, but rather have cases that fill half or the entire concha. Following this model, there are the ITC hearing aids, which are much smaller. They occupy a sufficiently small portion of the cavum concha and their outer face is parallel to the ear canal opening.

Lastly, the CIC and the IIC hearing aid are the smallest types and fit entirely within the ear canal. These hearing aids use components small enough that none of the hearing aid need protrude into the concha, and may even be invisible from the outside, save sometimes for a small nylon finishing line with a small knob on the end that helps remove the aid.

Figure 2.4 shows all these hearing aid models and Figure 2.5 presents the typical locations of the components in an ITC and in a BTE hearing aid.

Figure 2.4: Main hearing aid models. Source: I Denture & Hearing [22].

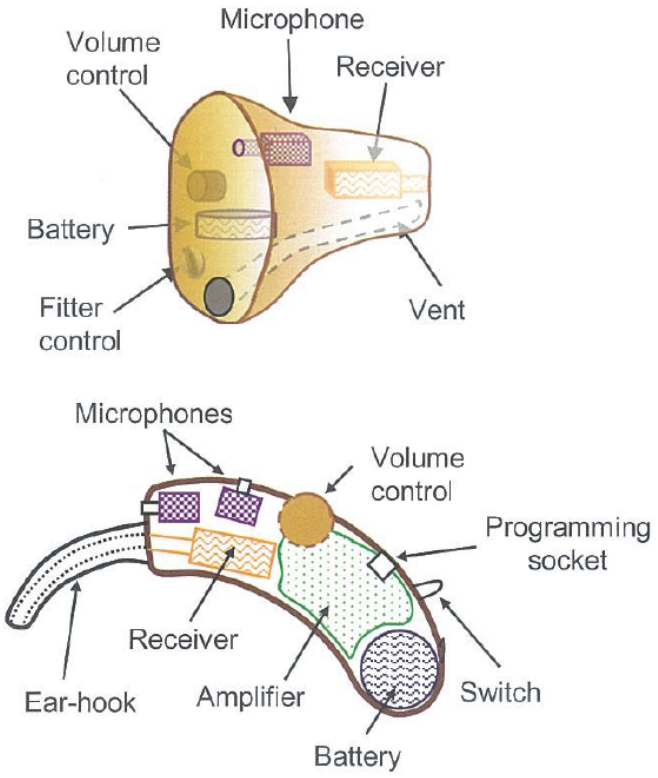


Figure 2.5: The typical location of components in an ITC and in a BTE hearing aid. Source: Dillon [15].

Modern hearing aids have digital integrated circuits and components that allow sophisticated processing strategies, as well as tools that enable wireless communication between one hearing aid and the other. The incoming signal is first captured by the microphones, which in BTEs are usually directional, and can then be modified in several ways. Two of the main steps in the processing of the signal delivered by the microphones are noise reduction and fitting, respectively. A block diagram illustrating these steps that occur between the incoming signal and the output signal delivered to the hearing aid user's ears is shown in Figure 2.6.

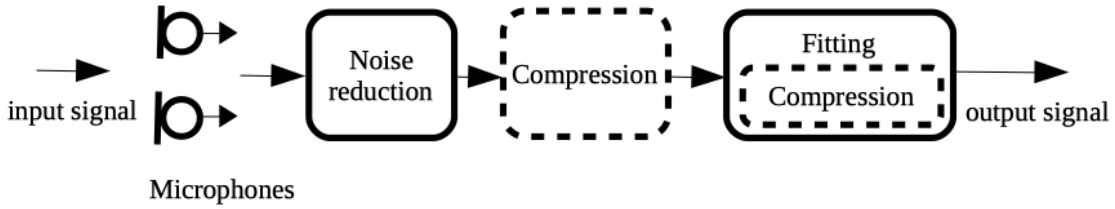


Figure 2.6: Block diagram of the main steps of signal processing in hearing aids.

Firstly, sound is captured from the environment by the hearing aid microphones. BTE hearing aids usually have at least two microphones, each in a different position within the case. These also have a fixed or adjustable polar pattern. Their polar patterns, the particular position at which they are situated and the distance from one to the other are essential for the subsequent processing of the input sound, particularly for spatial perception. It is common for modern BTE hearing aid microphones to have an adjustable directional configuration; that is, they can work as omnidirectional microphones for some situations (e.g. when there is homogeneous background noise), or can change their polar pattern to a specific directional response when a speech signal which is more intense than the background noise is detected as coming from a specific direction. The directional response can then enhance sounds coming from the direction of speech and minimize the capture of sounds coming from other directions. With that in mind, directional microphones can be considered a powerful tool for noise reduction and SNR increase, which is a desirable feature for hearing-impaired people, as described in Section 2.1.

After the important role played by the microphones, modern hearing aids commonly have an additional block for digital noise reduction, in order to further increase the SNR of the incoming sound. A common digital noise reduction technique is Wiener filtering [8–11]. Though Wiener filtering-based noise reduction algorithms can have very good results at reducing noise, they must also be applied with care, in order to minimize a potential distortion of the user's spatial hearing ability [9–11]. This is further discussed in Chapter 3.

Finally, sound is then adjusted for the particular audiometric needs of the user. Given the user's hearing loss pattern in each ear, different amounts of gain are applied at different frequency bands in order to maximize audibility. This step in the block diagram is referred to as *fitting*. Fitting procedures can be divided as linear and non-linear. Linear fitting strategies prescribe the same frequency-dependent gains for any input signal level. With non-linear fitting procedures, on the other hand, the amount of gain for each frequency band varies with the input level, leading to a phenomenon called *compression*. For non-linear fitting prescriptions, compression can be seen as a feature of the fitting step itself and therefore can be placed inside

the fitting block. Nevertheless, it can also be treated as a separate entity, since the hearing aid may or may not operate with it, and it can also be applied at another stage of the processing chain besides the fitting procedure itself, as illustrated in Figure 2.6, where compression is applied both at the fitting procedure itself and before it, right after noise reduction. A hearing aid might as well have several compressors working in combination, and each of them can act at a different stage of the processing chain. The next section describes the particular characteristics of compression in more detail.

2.3 Compression

As mentioned earlier in this chapter, when the cochlea is damaged, as in a sensorineural hearing loss, the impairment not only consists of a loss in audibility, but also of a shrinking of the subject's dynamic range between the threshold of hearing and the threshold of pain. Therefore, it is important that hearing aids decrease the dynamic range of signals in the environment so that all signals of interest can fit within the restricted dynamic range of the hearing-impaired person [15]. This mechanism is called *compression*, and it is used in virtually all hearing aids available nowadays. A compression is an amplifier that automatically reduces its gain once the input sound pressure level rises above a certain threshold. There are, however, many ways in which compression can be implemented in order to better fit the hearing aid user's needs.

Figure 2.7 illustrates three possible ways in which the dynamic range of signals can be reduced. In the left panel, gain starts reducing as soon as the input level rises above *weak*. By the time a moderate input level has been reached, the gain has been already sufficiently reduced, and linear amplification can be applied to all higher input levels. On the right panel, the opposite occurs: the input signal is amplified with no compression until it reaches a moderate SPL, and then moderate and intense sounds are severely squashed by compression into a small range of output levels. Finally, the middle panel shows an example where compression is applied less aggressively in the same amount along a wide dynamic range of input signals. These three ways in which compression can be applied are called low-level compression, high-level compression and wide dynamic range compression, respectively. Special attention will be given to the latter two later in the text.

2.3.1 Compressor parameters

Although a hearing aid can have multiple compressors that work in a combined and complex way, the operation of each compressor within the aid can be described in a

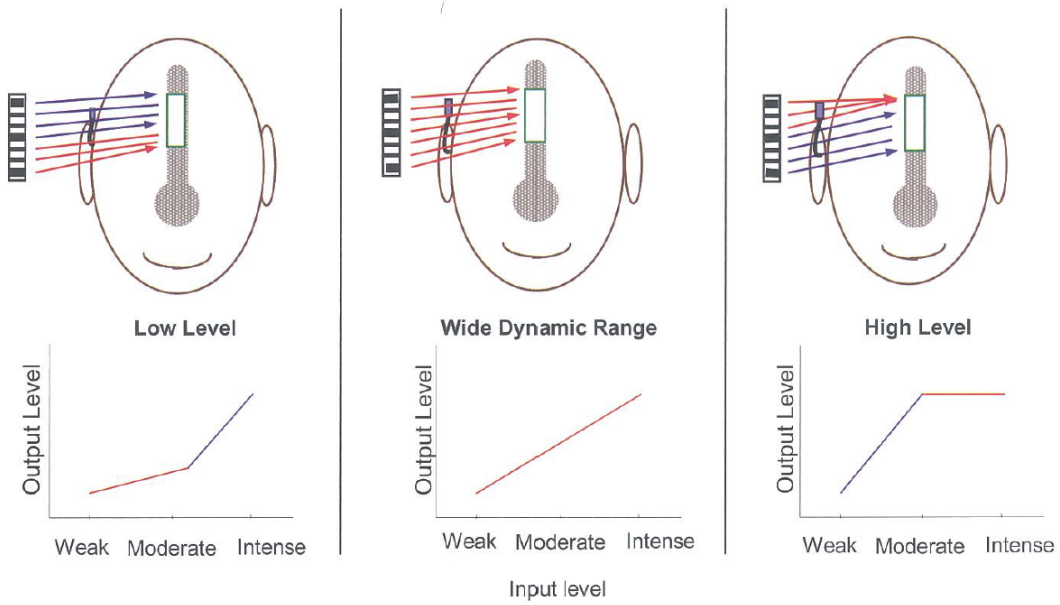


Figure 2.7: Three ways in which the dynamic range of signals can be reduced. In each case, the upper figure shows the spacing of different signal levels before amplification (the left end of the lines) and after amplification (the right end of the lines). The lower figure shows the same relationships, but as an input-output function. In each case, compression is occurring in the red region. Source: Dillon [15].

relatively simple way by a few dynamic and static parameters. Firstly, the output signal SPL is continuosly measured by a detector, which converts the waveform to a smooth control signal by means of rectification and smoothing. Once the output level rises, the measured signal in the detector increases gradually to its new value and is gradually passed on to the compression control circuit. The compressor then starts reacting to this change in level by turning the gain down until it is sufficiently decreased. This reaction time taken by the compressor is called *attack time*, and is formally defined as the time taken for the output to stabilize to within 3 dB (ANSI S3.22) of its final level after the input of the hearing aid increases from 55 to 90 dB SPL (ANSI S3.22 [23]) [15]. Similarly, the compressor also takes some time to react to a gradual change in level measured by the detector once the input signal level decreases. This time interval is called *release time*, and is the time taken for the compressor to react to a decrease in input level by gradually diminishing the gain reduction. It is formally defined as the time taken for the output signal to increase to within 4 dB of its final value following a decrease in input level from 90 to 55 dB SPL, as in ANSI S3.22 [23]) [15].

The attack time (AT) and the release time (RT) are the *dynamic* parameters of a compressor. ATs can be configured to be as small as less than 10 ms, but might also be much longer, greater than 100 ms, depending on what is supposed to be achieved with the compressor. This flexibility also works for the RTs, which are rarely less

than 20 ms and can be much longer, up to the order of magnitude of 1 s. Different types of compression imply different ranges of attack and release times, and this is a crucial decision, since their values can have a direct impact on the signal's envelope, and might even lead to signal distortion, as will be better described later in this chapter.

The static parameters of a compressor, on the other hand, are the *compression threshold* and the *compression ratio*. The compression threshold (CT) is defined as the input SPL above which the hearing aid begins compressing [15][1]. After this threshold, there is a compression region, whose gain reduction proportion is defined by the compression ratio. A compression ratio of 3:1, for example, says that, for each 3 dB increase in input level, there is only a 1 dB increase at the output level. Higher compression ratios indicate more compression. With that in mind, hearing aids sometimes have several compression regions with different compression ratios. Besides possible compression regions with moderate compression ratios already at moderately low or moderate input levels, it also is common for hearing aids to have a compression region starting at a very high input level with a very large compression ratio, usually greater than about 8:1 [15]. This mechanism is often called limiting is used to greatly reduce the gain for very high input levels in order to avoid peak clipping, which is undesirable in hearing aids, since the clipping will lead to a perceivable signal distortion.

Before the first compression threshold, however, most hearing aids amplify linearly. Some continue with linear amplification up to infinitesimally small input levels, and some even have a region of *expansion*. Expansion is the opposite of compression: an increase in the input level will produce a proportionally *greater* increase in the output level, so that it expands the dynamic range of that region, instead of squashing it. Consequently, the compression ratio of an expansion region is less than 1:1. Expansion is particularly useful to minimize the perception of the inner noise of the hearing aid itself when external noises are at a very low SPL.

A simple and very straightforward way to visualize the behavior of the hearing aid for different input levels is the I-O curve. The I-O curve has an axis, usually horizontal, with input level values in dB SPL, and another axis, vertical, showing the corresponding output level delivered by the hearing aid for each input level. One can notice that regions in which compression is applied have a smaller slope compared to those with linear amplification, and the line tends to get more and more horizontal as the compression ratio increases. An example of a common type

[1] In some circumstances, it is the *output* SPL at which compression commences. Since modern hearing aids have multiple compressors working in combination, it has become simplistic to describe hearing aids as either having input-controlled compression or output-controlled compression [15]. For the sake of simplification, compression threshold will always be referred to in this dissertation as the *input* SPL at which compression begins.

of I-O curve for a hearing aid with regions of compression and linear amplification is shown in Figure 2.8.

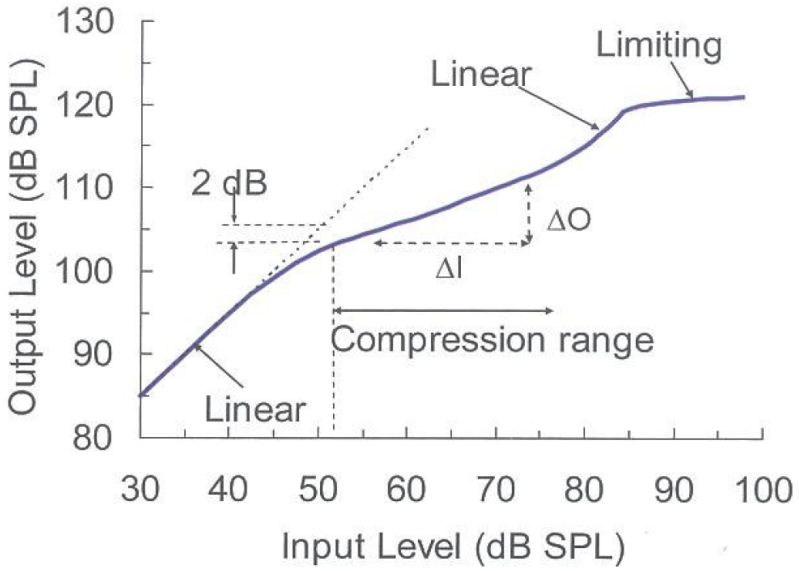


Figure 2.8: Example of I-O curve with regions of compression (including limiting) and linear amplification. Source: Dillon [15].

It is important to notice, as previously said, that hearing aids nowadays might have several different compression regions with different compression ratios, and digital technology has allowed the hearing aid to basically sculpt the I-O curve to best match the audiometric needs of the user. Therefore, for better understanding, the rest of this text presents the basic concepts in a compact, succinct manner and describes several types of compression in a didactic way, with simple examples.

2.3.2 Output limiting compression *versus* WDRC

As briefly seen at the beginning of this section, there are several ways in which compression can be applied. It can be applied to low levels and be replaced by linear amplification once the input level is considered "moderate"; it can begin at high input levels in order to strongly reduce the level of intense sounds; or it can even exist along a large dynamic range with a moderate compression ratio, this case referred to as wide dynamic range compression. With that in mind, there are two main types of compression in terms of its static parameters (compression threshold and compression ratio) that will receive special attention here: wide dynamic range compression (WDRC) and output-limiting compression.

Output-limiting compression is basically a high-level compression. The I-O curve presents linear amplification up until a high compression threshold, and then suddenly a strong compression region with a large compression ratio begins. A high

CT means that the hearing aid begins to compress at a relatively high input SPL (i.e., 60 dB SPL or more) [24]. A high CR is roughly defined by most practitioners as being greater than 3:1 or 4:1 [24]. This type of compression might be the better choice for subjects with severe to profound hearing loss, given that a maximum gain can be used in most of the subject's dynamic range for better audibility [24].

Wide dynamic range, on the other hand, is roughly the opposite of output-limiting compression. It is associated with low CTs (less than 60 dB SPL) and low CRs (less than 4:1); that is, instead of limiting the ceiling, it "lifts the floor". WDRC hearing aids are almost always in compression, because virtually all kinds of inputs, from very soft speech to a scream, will cause it to go into compression [24]. Once it goes into compression, however, it does so with a moderate compression ratio. One can say, therefore, that WDRC basically provides a weak degree of compression over a wide range of inputs. WDRC is an interesting choice for people with mild-to-moderate hearing loss, given that the "floor" of their dynamic range is raised, but the "ceiling" not so much; WDRC will then do its job of amplifying soft input sounds more than moderate and loud input sounds. Figure 2.9 shows a brief comparison of both output-limiting compression and WDRC I-O curves.

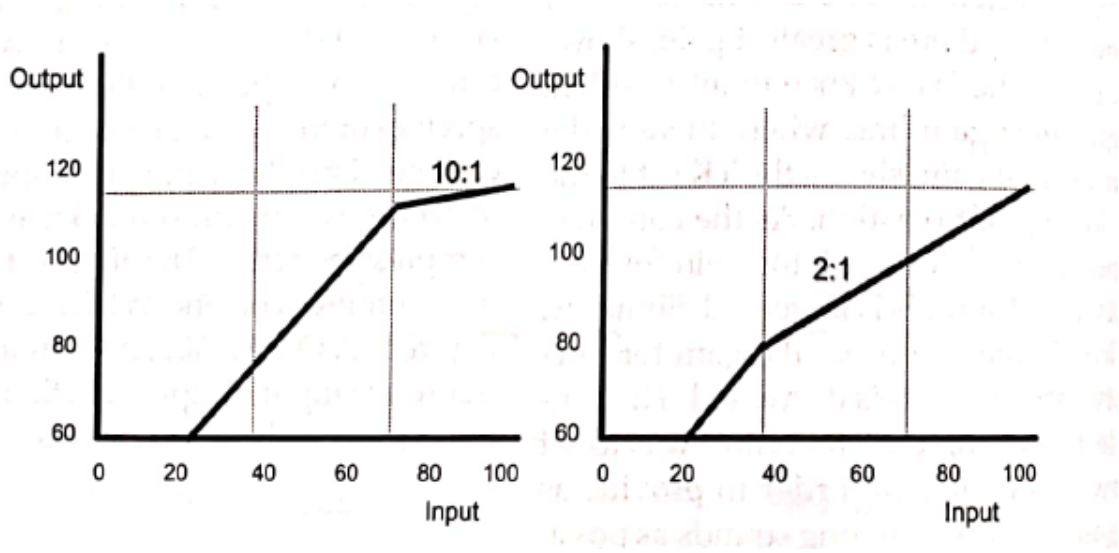


Figure 2.9: Output-limiting compression (left) *versus* WDRC (right), with compression ratios of 10:1 and 2:1, respectively. Source: Metz[24].

It must be noted that WDRC, which greatly amplifies soft sounds, and amplifies loud sounds less, might distort the waveform of signals such as speech, because the more intense peaks would be given less gain than the less intense valleys. This might decrease important contrasts in the waveform of speech and make it less intelligible. This effect is even increased when very short attack and release times are used, as will be seen in Section 2.3.4.

2.3.3 Single-channel versus multichannel compression

So far, some types of compression and the main compression parameters have been introduced, and it was implicitly considered that a single type of compression, with a single combination of compression parameters, would be applied equally to all frequency bands of the signal. When this occurs, it is called *single-channel compression*, because all frequency bands will be treated equally and be treated as only one channel in the hearing aid's processing chain.

Multichannel hearing aids, in contrast, split the incoming signal into different frequency bands, and each band of signal passes through a different amplification channel [15]. Each amplification channel then contains its own compressor. There are two basic reasons why multichannel compression might be desirable [15]:

- the amount of compression varies with hearing loss, but hearing loss usually varies with frequency;
- the amount of compression varies with signal level, but signals and noises in the environment have more energy in some frequency regions than in others.

That being said, multichannel compression can be adjusted so that the static and dynamic compression parameters are chosen separately for each frequency band. Although there are many ways in which compression can vary from one channel to the next, the degree of compression often either increases or decreases with frequency [15]. When the degree of compression is greater in the high-frequency channels than in the low frequency channels, the high-frequency components will have more gain for low input levels than for high input levels; this type of multichannel compression is called TILL (treble increase at low levels). Conversely, when more compression is applied to low-frequency channels than to high-frequency channels, low-frequency components will be more amplified at low input levels than at high input levels. This type of compression is labeled as BILL (bass increase at low levels). Nowadays, the use of these terms has decreased, given that multichannel compression in modern hearing aids has become more complex.

2.3.4 Adjusting compression parameters

Now that a description of how the compressor works has been given, and that the main static and dynamic compression parameters have been introduced, some choices for parameter tuning aiming at different purposes are described here.

Firstly, one of the main goals of compression is to avoid discomfort, distortion and damage. This can be done by using an efficient way of controlling the maximum output level. It must be noted that, for very high input levels, a hearing aid without compression would saturate and this would cause the amplified signal to clip. In

order to minimize the distortion in the output signal caused by clipping and the user's exposure to very loud SPLs, output-limiting compression is a very good choice.

This can be obtained by generating a compression region that starts at a high compression threshold–given that it would work on loud input levels — and has a high compression ratio, in order to sufficiently decrease the gain for those input levels. This technique is known as *compression limiting*, as opposed to the hard clipping that would occur otherwise. The attack time should be short, so that the gain is decreased rapidly enough to prevent loudness discomfort. On the other hand, softer sounds following an intense sound should not be overly attenuated; therefore, the release time should be also short or adaptive (that is, adjusted by the hearing aid depending on the context). If a hearing aid contains wide dynamic range compression and therefore does not include a compression limiter, the output signal will eventually clip, but, for good quality hearing aids, the input level at which peak clipping will occur will normally be high enough not to matter. Table 2.1 shows the parameter choices for this type of compression.

Compression to control maximum output
Compression ratio > 8:1
Attack time < 15 ms and release time between 20 and 100 ms or adaptive
Compression threshold low enough to avoid discomfort
Single or multichannel

Table 2.1: Compression rationales to control maximum output [15].

Secondly, it might also be interesting for a compressor to help reduce the dynamic range of speech signals. The most intense speech sounds (some vowels) are about 30 dB louder than the weakest sounds (some unvoiced consonants) [15]. For hearing-impaired people with a very reduced dynamic range, it might be difficult to make weak speech sounds audible enough without causing more intense speech sounds to be excessively loud. Apart from this problem, weaker phonemes can also be masked by stronger adjacent ones, or the weak formants [25] might be masked by the stronger formants. Therefore, compression can be a powerful tool to help decrease those intensity differences, by having a compressor that increases its gain during weak syllables (or phonemes) and decreases its gain during intense syllables (or formants). This type of compression is called syllabic or phonemic compression.

While syllabic/phonemic compression might be helpful to reduce the dynamic range of speech to better suit the user's dynamic range and to avoid masking, the altering of the intensity relationships between phonemes and syllables might be a potential problem for intelligibility. If the hearing aid wearer uses the relative intensities of sounds to help identify them, altering relative intensities can make the speech signals less intelligible, even if it makes them more audible. Therefore, the

choice of compression parameters must be done carefully.

For syllabic compression to be applied, the compression threshold must be low enough for compression to be active during typical speech levels. Compression ratio must be high enough to sufficiently decrease the dynamic range of the speech signal, but low enough to leave some intensity differences intact. Attack and release times should be short enough so that it can vary sufficiently from one syllable or phoneme to the next, but not so short that they create significant amounts of distortion to the waveform. These characteristics are summarized in Table 2.2.

Compression to reduce inter-syllabic level differences
Compression ratio $>$ 1.5:1, but $<$ 3:1
Attack time from 1 to 10 ms and release time from 10 to 50 ms
Compression threshold $<$ 50 dB SPL
Single or multichannel

Table 2.2: Compression rationales to reduce inter-syllabic level differences [15].

Lastly, an alternative use of compression is to decrease the longer-term dynamic range, but without changing the intensity relationships between syllables that follow each other. This type of compression is called *automatic volume control* (AVC), because the compressor varies the gain in a similar way a person would manually adjust the volume control to partially compensate for differences in the incoming level of sounds. AVC is a slow, natural adjustment to a change in intensity of the surrounding sounds, i.e. when a person moves from one environment to another, or when a talker raises his/her voice.

For such a compression strategy to be applied, attack times and release times must be long enough not to distort inter-syllabic differences, but rather to respond more slowly to input level changes. A big problem, though, is that long attack times cannot account for abrupt variations in input level. If the input level rises very rapidly, the compressor will not be fast enough to sufficiently reduce the gain and, therefore, a compression limiter is needed for this job. The opposite problem, a sudden decrease in level, might lead to more difficulty to hear softer sounds right after intense sounds. This can be minimized by having release times only long enough to avoid rapid increases in gain during brief pauses in an environment with otherwise loud sounds. Table 2.3 summarizes the configurations for AVC-compression.

Modern hearing aids can have all these types of compression working in combination in a way that is better suited for each type of environment and situation.

Compression to decrease long-term level differences (automatic volume control)
Compression ratio > 1.5:1, but < 4:1
Attack time > 100 ms and release time > 400 ms
Compression threshold < 50 dB SPL
Single or multichannel

Table 2.3: Compression rationales to decrease long-term level differences [15].

2.4 Fitting procedures

After a description of directional microphones, noise reduction, and compression, now comes the part where the audiometric data of the subject is explicitly used to adjust the hearing aids for the amplification needs of the user. This information is given by the audiogram, which indicates how the subject's hearing loss is configured across frequencies.

An audiogram is a diagnosis given by a hearing test. In a hearing test, several exams are performed, and the main one to evaluate hearing loss for air-borne sound is the pure-tone audiometry. During this exam, pure tones are played to the subject usually at eight standard frequencies ranging from 250 Hz to 8000 Hz. The audiologist running the exam plays each pure tone at an increasing sound level, and the subject must indicate when the sound first becomes audible. The sound levels where this occurs are then registered to form the subject's audiogram.

The audiogram is a graph whose horizontal axis measures frequency, and whose vertical axis measures sound level. The horizontal axis usually ranges from 250 Hz to 8000 Hz, and the vertical axis has 0 dB HL (HL for "hearing loss") as a reference for the normal hearing threshold. Going down the vertical axis, the sound levels indicate the amount of hearing loss in dB HL for each frequency. That is, the higher the sound level for which sound at a specific frequency first becomes audible, the greater the hearing loss at that frequency. If, for example, a person has 40 dB HL for the frequency of 1 kHz, it means that, for this person, sounds at that frequency must be at least 40 dB above 0 dB HL to be audible. Table 2.4 shows some criteria to define the degree of hearing loss, based on the hearing loss values averaged across 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz for each ear.

Grades of hearing impairment	**Mean hearing loss**
0 – No impairment	25 dB or better
1 – Slight impairment	26–40 dB
2 – Moderate impairment	41–60 dB
3 – Severe impairment	61–80 dB
4 – Profound impairment including deafness	81 dB or higher

Table 2.4: Degrees of hearing loss. Adapted from Zahnert [18].

Figure 2.10 shows an example of an audiogram for a person who has a hearing impairment for sounds from 2 kHz onwards at both ears. The audiometric curve for the left ear can be colored blue and has, as a defined rule, the thresholds indicated by the symbol "X", and the curve for the right ear can be colored red and has the thresholds indicated by "O".

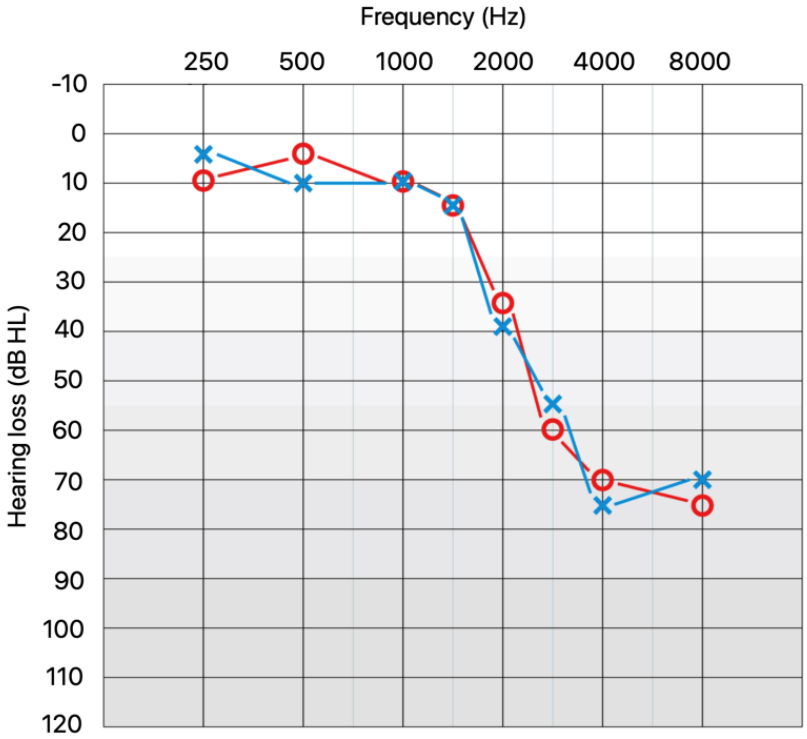


Figure 2.10: Example of audiogram for left and right ear (airborne sound). Adapted from Botella [26].

With a diagnosed hearing loss which can be read in the audiogram, the subject must have the amplification characteristics of the hearing aids adjusted. This is done by a *prescription procedure*, also known as fitting procedure.

One could assume that, for restoring normal hearing, the amplification given for each frequency would simply be the same as the amount of hearing loss at that frequency. In sensorineural hearing loss, however, the gain needed to restore normal loudness perception is equal to threshold only when the person is listening at threshold [15]. For all higher levels, this amount of gain would be excessive, given the narrowing of the dynamic range for people with sensorineural impairment.

Therefore, prescription procedures should take a more careful approach. One of the first findings was the one by Lybarger [27], where it was observed that the amount of gain chosen by the examined hearing-impaired people was approximately half of the threshold loss. This originated the *half-gain rule*, which consists of prescribing a gain equal to half of the subject's hearing threshold for that frequency. A bit later

it was observed that this approach was still too simplistic and not entirely correct, and the prescription procedures had to be further refined.

After decades of development, fitting procedures can be divided into two main groups: linear and non-linear. Linear fitting procedures consist of calculating a specific gain prescription that remains the same for any input signal sound level. Non-linear fitting procedures, on the other hand, have an input-level dependent method of calculation, which means that compression will be applied.

Some examples of linear prescription procedures are POGO [28] and POGO-II [29], NAL [30], NAL-R [31] and NAL-RP [32], and DSL [33]. Of all of them, POGO is one of the simplest, consisting only of the *half-gain rule* plus a fixed frequency-dependent constant. A later modification, called POGO-II, added another term to provide additional gain for people with severe and profound hearing loss. The NAL-R (National Acoustics Laboratory, revised) [30] formula, on the other hand, was a revised version of the original NAL formula, which aimed to maximize speech intelligibility at the listening level preferred by the hearing aid wearer. While the original NAL prescribes a minor variation of the half-gain rule (0.46 times the 1 kHz threshold plus adjustable frequency-dependent constants), the NAL-R formula retains the 0.46 factor for the three-frequency average gain (that is, the average hearing loss between the 500 Hz, 1 kHz and 2 kHz hearing losses), but adds another factor of 0.31, which reflects the amount of variation of the gain-frequency curve shape with respect to the shape of the audiogram. The NAL-RP formula makes further adjustments for severe-to-profound hearing losses.

The NAL-R formula is calculated as:

$$H_{3FA} = (H_{500} + H_{1k} + H_{2k})/3 \tag{2.1}$$

$$X = 0.15 H_{3FA} \tag{2.2}$$

$$IG_i = X + 0.31 H_i + k_i, \tag{2.3}$$

where H_i is the hearing loss at frequency i, H_{3FA} (three-frequency average) is the average between the 500 Hz, 1 kHz and 2 kHz losses, and IG_i is the calculated prescription in the form of an insertion gain (the difference between the sound pressure level measured near the eardrum with a hearing aid in place, and the sound pressure level measured in the unaided ear) for frequency i. The constants k_i are shown in Table 2.5.

Freq(Hz)	250	500	750	1k	1.5k	2k	3k	4k	6k
k	-17	-8	-3	1	1	-1	-2	-2	-2

Table 2.5: Value of k constants for NAL-R formula.

As for non-linear fitting procedures, some of the most famous are NAL-NL1 [34], NAL-NL2 [35], FIG6 [36], CAMEQ [37] and CAMREST [38]. Nowadays, hearing aid manufacturers also often have their own dedicated non-linear fitting procedures, which are becoming increasingly more complex and refined, so that one is virtually able to sculpt an input/output curve based on the subject's individual needs. Both universal and manufacturer-owned non-linear procedures have in common the fact that gain prescription rules change depending on the input sound level. That means that compression is applied in a specific way for different input level ranges.

One of the most straightforward non-linear procedures is the FIG6, which specifies how much gain is required to normalize loudness for medium and high-level input signals [15]. It has prescription rules for 40 dB SPL, 65 dB SPL and 95 dB SPL input signals. The rules are interpolated for input levels in between these values.

For low level (40 dB) input signals, the gain is prescribed on the basis that people with mild or moderate hearing loss should have aided thresholds 20 dB above the normal hearing thresholds; it is not worth providing more gain than this, since background noise would prevent very soft sounds from being perceived regardless of how much gain is prescribed. After the first 20 dB of hearing loss, every additional decibel of loss is compensated by an extra decibel of gain. This rule is then relaxed to a half-gain rule once the unaided threshold exceeds 60 dB HL, because otherwise the high gains would most likely cause feedback oscillation in the hearing aid device [15].

For typical (65 dB) input signals, gain is prescribed so that narrowband sounds perceived at a comfortable level by a normal-hearing person will also be perceived as comfortable by the hearing-impaired person.

For high-level (95 dB) input signals, the gain is prescribed so that the sound is perceived as loud for the hearing aid wearer as they would be for a normal-hearing subject [15].

The FIG6 equations are:

- for 40 dB SPL input levels:

$$IG_i = \begin{cases} 0 & \text{for } H_i < 20 \text{ dB HL}, \\ H_i - 20 & \text{for } 20 \leq H_i \leq 60 \text{ dB HL}, \\ 0.5H_i + 10 & \text{for } H_i > 60\text{dB HL}; \end{cases} \tag{2.4}$$

- for 65 dB SPL input levels:

$$IG_i = \begin{cases} 0 & \text{for } H_i < 20 \text{ dB HL}, \\ 0.6(H_i - 20) & \text{for } 20 \leq H_i \leq 60 \text{ dB HL}, \\ 0.8H_i - 23 & \text{for } H_i > 60\text{dB HL}; \end{cases} \tag{2.5}$$

- for 95 dB SPL input levels:

$$IG_i = \begin{cases} 0 & \text{for } H_i \leq 40 \text{ dB HL,} \\ 0.1(H_i - 40)^{1.4} & \text{for } H_i > 40\text{dB HL.} \end{cases} \tag{2.6}$$

In all cases, H_i is the hearing loss at frequency i and IG_i is the prescribed insertion gain for the respective frequency band.

Given that fitting procedures heavily distort the input signal, especially when compression is involved, it is expected that the hearing aid user may perceive an increase in audibility, but also have greater difficulty to tell where the sound is coming from. Investigating the subject's sense of sound localization when dealing with signals processed by fitting and compression is one of the main goals of this work, and measurements to estimate sound localization are described in the following chapter.

Chapter 3

Sound localization and binaural cue measurement

One of the many attributes of the sense of hearing is the ability to localize sound in space. Spatial hearing is a powerful tool for mammals to perceive what is happening in their environment, hunt, detect the location of warning signs (a predator, a car horn, a siren), respond to another subject's calling, and to increase awareness of their own position in an environment, even when sight is unavailable.

The ability to detect the localization of sound sources is made possible thanks to the existence of two ears, to the ear's shape and to the shape of one's own head and torso. When sound is propagated in space, it reflects and diffracts in a very complex way when it meets our body, so that the signal the reaches our ear canal is already an altered version of the one emitted by the sound source. One can think of it as a system whose input is the sound generated at the source, and whose output is the sound that reaches the ear canal. The transfer function for this system is known as Head-Related Transfer Function (HRTF), which is a product of all reflections and diffractions that occur at our pinna, head, shoulders and torso. In time domain, this is the equivalent of the Head-Related Impulse Response (HRIR). Figure 3.1 shows a block diagram that better illustrates how this system works. It is worth noting that one can also include the presence of the ear canal in the HRTF and consider the output signal to be the one that reaches the eardrum, rather than the entrance of the ear canal. This approach, though, is not followed in this work.

There is one specific HRTF for every direction of incidence of sound. To better set up an orientation system, one can imagine three main planes that intercept a subject's head. These are the horizontal plane, the vertical plane, and the median plane, which are shown in Figure 3.2. With that in mind, one can use an azimuth angle and an elevation angle to determine any position inside this three-dimensional space. Therefore, there is an HRTF for each combination of azimuth and elevation angles.

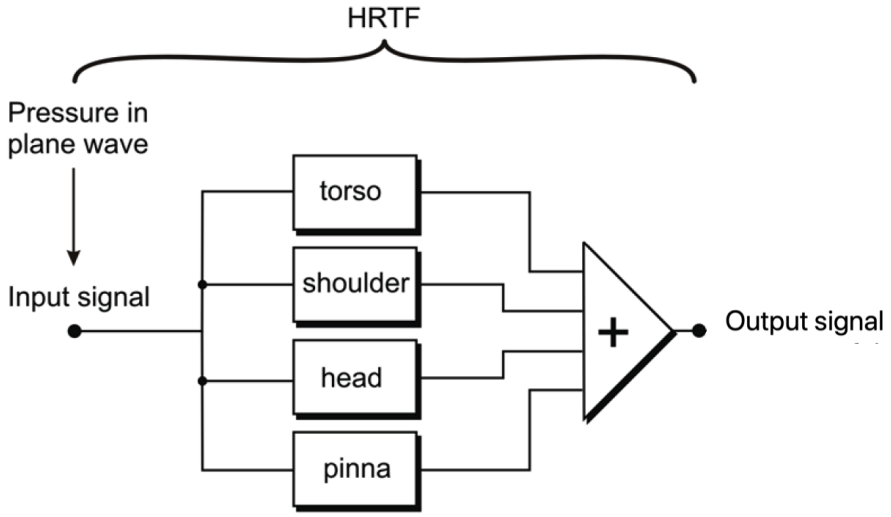


Figure 3.1: Block diagram illustrating the path of sound from the source to the entrance of the ear canal. Adapted from Vorländer [39].

In order to deduce the position of a sound source in space, the subject uses some *cues*. These are called binaural cues and monaural cues. Monaural cues are those that are related to the spectral shape of the HRTF at a specific position, and are identical for both ears [40]. Binaural cues, on the other hand, are related to the differences between the incoming signals at both ears. The binaural cues are called Interaural Time Difference (ITD) and Interaural Level Difference (ILD) and are in most cases the main cues to determine the location of a source.

Interaural level difference is the difference between the sound pressure levels of the incoming signals at each ear. Due to reflections and diffractions suffered by the sound waves that must travel through and around the head to reach the ear that is furthest from the sound source (a phenomenon known as head-shadow effect [41]), the sound level at the ear that is furthest from the source is lower than at the closer ear. Therefore, the ILD can be mathematically expressed by the logarithmic squared magnitude of the interaural transfer function (ITF) between the signals arriving at both ears. The instantaneous ITF for a generic source is expressed as [42]:

$$\mathrm{ITF}(\omega) = \frac{X_\mathrm{l}(\omega)}{X_\mathrm{r}(\omega)}\,, \tag{3.1}$$

where $X_\mathrm{l}(\omega)$ and $X_\mathrm{r}(\omega)$ are the Fourier transforms of the incoming signals at the left ear and at the right ear, respectively, and ω is the domain of angular frequencies. For a single point source, the incoming signals are a product between the Fourier transform of the signal emitted by the sound source and the acoustic transfer functions of the path from the source to each ear, which are equivalent to the HRTFs.

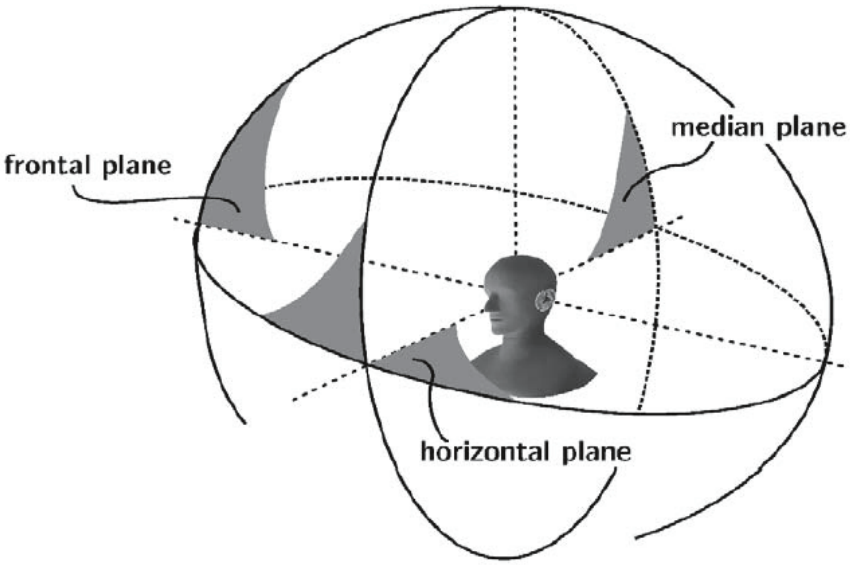


Figure 3.2: Head-related coordinate system. Source: Vorländer [40].

Therefore, the ITF, which, in this case, can also be called RTF (relative transfer function), can be rewritten as [10, 11]:

$$\mathrm{ITF}(\omega) = \mathrm{RTF}(\omega) = \frac{H_\mathrm{l}(\omega)}{H_\mathrm{r}(\omega)}\,, \tag{3.2}$$

where $H_\mathrm{l}(\omega)$ and $H_\mathrm{r}(\omega)$ are the HRTFs for the left and the right ear, respectively. Consequently, the instantaneous ILD can finally be expressed as [10, 11, 42]:

$$\mathrm{ILD}(\omega) = 10 * \log_{10} |\mathrm{ITF}(\omega)|^2. \tag{3.3}$$

The interaural time difference is the difference between the times of arrival of the incoming signals at each ear. Sound will always reach the ear closest to the sound source and then travel a longer path to reach the opposite ear. Therefore, the instantaneous ITD is related to the phase of the ITF (which, in case of a point source, can be replaced by the RTF) by [10, 11, 42]:

$$\mathrm{ITD}(\omega) = \frac{\angle ITF(\omega)}{\omega}. \tag{3.4}$$

It must be noted that the average binaural cues of a point source are equal to its instantaneous binaural cues. Since all signals in this work are speech signals, which are considered to be approximately point sources, Equations (3.1)–(3.4) are applicable to them.

The importance of each binaural cue is not the same for the entire range of audible frequencies. It has been stated in the Duplex Theory [43] that the ITD

is the most important cue for lower frequencies, whereas the ILD is the prevalent cue for higher frequencies. This can be explained due to the fact that it is easier to detect a difference in phase when the wavelength has an order of magnitude which is greater than the order of magnitude of the head size. Besides that, given that low frequencies diffract easily around the head, there is no significant loss in sound power when sound reaches the side of the head furthest from the sound source. When the wavelength is very small, however, the phase difference is much harder to detect, but the head-shadow effect is much more prominent, since the wavelengths at higher frequencies are already too small to diffract at the head's surface; thus, the signal that reaches the side of the head closest to the sound source is almost completely reflected, and only a much smaller sound power arrives at the far side. There has been much debate about which frequency should be considered the boundary between the ITD-predominant low frequency region and the ILD-predominant high frequency region [44], but this boundary is usually considered to be around 1500 Hz [45, 46].

Figures 3.3 and 3.4 show, respectively, estimated ITDs and ILDs associated with multiple possible azimuths for some frequencies. The azimuth of $0°$ is located exactly in front of the head, while the azimuth of $-90°$ represents the right side of the head. Head symmetry is assumed, that is, symmetrical azimuths in relation to the median plane have the same magnitude, but switched signals. Both figures assume positive ILDs and ITDs for azimuths at the right side of the head.

The ITD and ILD, however, also have some limitations. There are some regions in space called "cones of confusion", which are cone-shaped surfaces around the interaural axis and centered at the entrance of the ear, where all positions have the same ITD and the same ILD. Positions on the surface of a cone of confusion are therefore more difficult to differentiate [13], causing a phenomenon called front-back confusion. A special case of cone of confusion is the median plane, on which all sound sources have ITDs and ILDs equal to zero. When it is necessary to distinguish two sources located at different positions on a cone of confusion, the listener has no choice but to slightly rotate the head or to use the monaural cues. The monaural cues are identical at both ears and can be seen in the HRTFs as slight spectral discolorations, peaks and notches at specific frequencies due to reflections and refractions at the pinna, the head and the torso. Two different positions that have the same ITD and ILD usually have different monaural cues. From experience, some monaural cues can be assigned to certain directions, thus contributing for sound localization in those cases. Aside from that, slight head movements, which are often done automatically, cause an alteration of the ILDs and ITDs, thus making a better localization once again possible.

A schematic depiction of a cone of confusion can be seen in Figure 3.5. Figure 3.6,

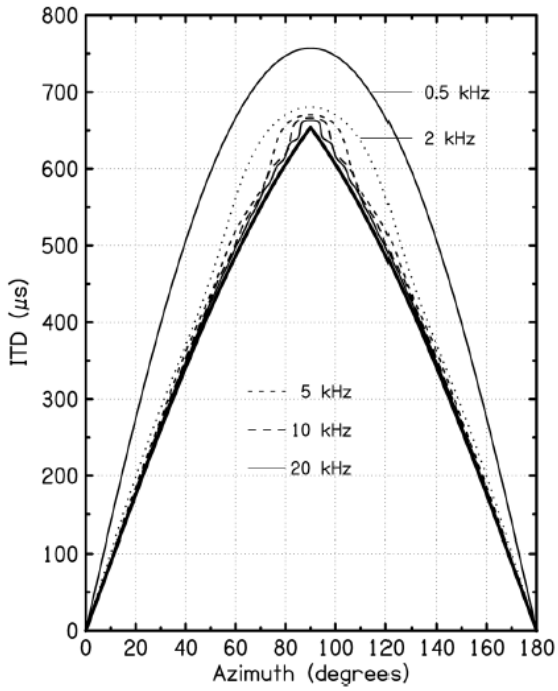


Figure 3.3: ITD values for some frequencies as a function of azimuth. Source: Aaronson and Hartmann [47].

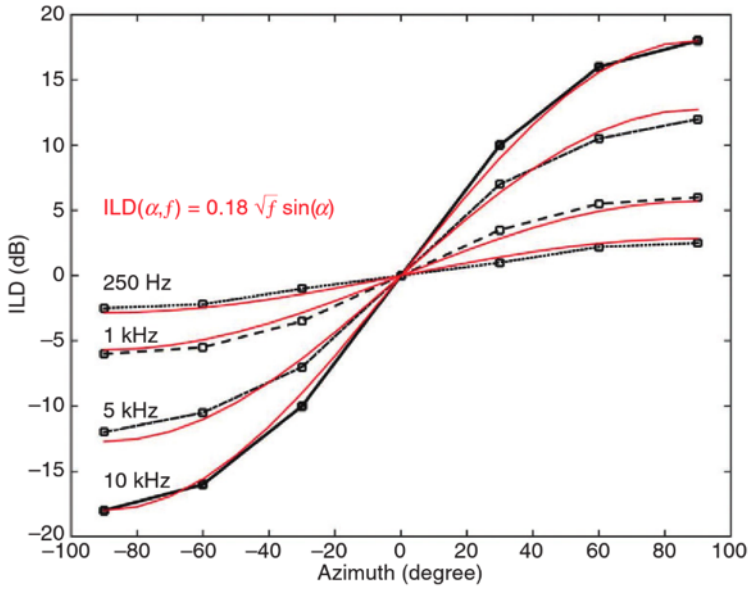


Figure 3.4: ILD values for some frequencies as a function of azimuth. Source: Van Opstal [47].

on the other hand, shows a sound source positioned in an approximate southwest direction in relation to an artificial head, and the HRTFs associated to that position are pictured in Figure 3.7, both in time and frequency domain. In Figure 3.7(a), the delayed sound incidence on the right ear can be noted clearly. In the frequency domain, it is seen that the HRTF of the ear closest to the source has significantly higher levels. The sink between 1 and 2 kHz is generated by the shoulder reflection, whereas the peaks around 4 kHz are caused by the pinna. When analyzing Figures 3.7(a) and 3.7(b), it is evident that the HRTF contains information regarding both the binaural and the monaural cues.

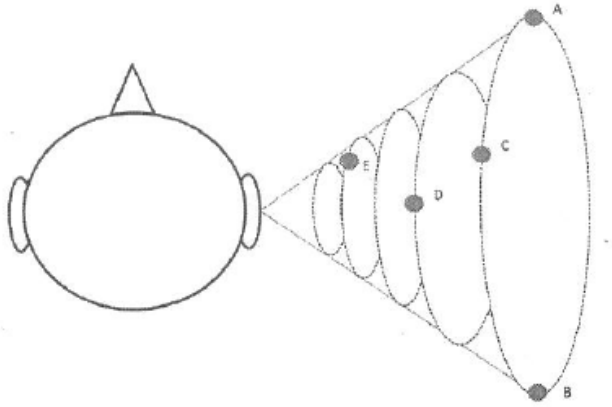


Figure 3.5: Schematic depiction of a cone of confusion. The five random positions on the cone's surface have identical ITDs and ILDs. Source: Fels [13].

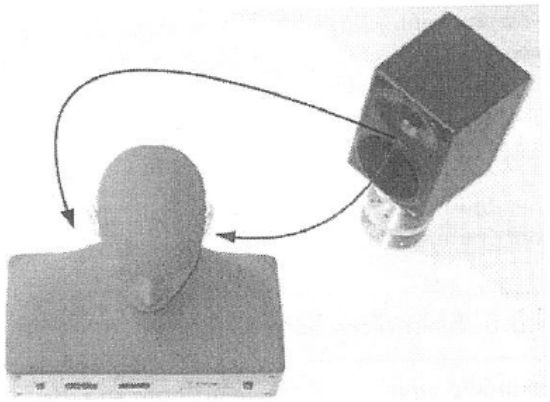
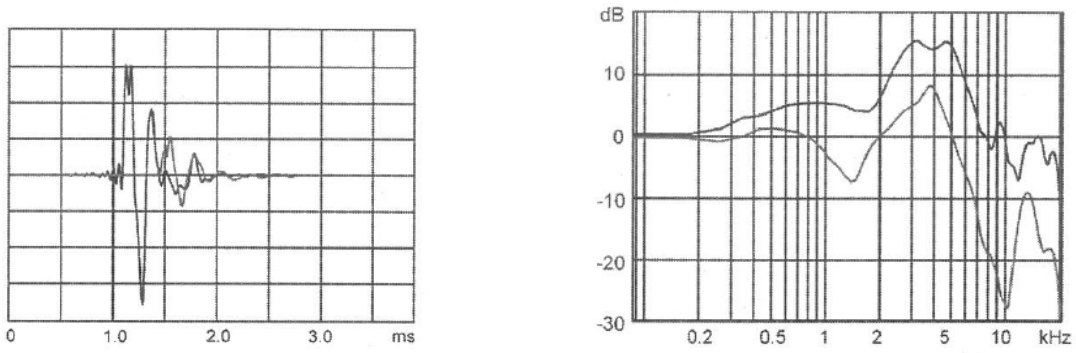


Figure 3.6: Schematic depiction of the sound pathways from the loudspeaker to the ear. Source: Fels [13].

Lastly, an important statistical parameter that can be used to verify the reliability of measured ITDs and ILDs, which here will also be referred to as a binaural cue, is the interaural coherence (IC).The average interaural coherence of a generic signal is the normalized cross-correlation between the expected values of the incoming



(a) Time domain

(b) Frequency domain

Figure 3.7: A pair of HRTFs in time and frequency domain. Source: Fels [13].

signals at each ear [11]:

$$\mathrm{IC} = \frac{\mathbb{E}[X_\mathrm{l}(\omega)X_\mathrm{r}^H(\omega)]}{\sqrt{\mathbb{E}[|X_\mathrm{l}(\omega)|^2]\mathbb{E}[|X_\mathrm{r}(\omega)|^2]}}. \tag{3.5}$$

From this expression, one can see that the IC has a magnitude between 0 and 1. For the instantaneous values of the IC, the expected values can be taken out of the expression. For a single point source, the signals $X_\mathrm{l}(\omega)$ and $X_\mathrm{r}(\omega)$ can be replaced by $H_\mathrm{l}(\omega)S(\omega)$ and $H_\mathrm{r}(\omega)S(\omega)$, respectively, where $S(\omega)$ is the original signal at the source. This results in:

$$\mathrm{IC} = \frac{\mathbb{E}[H_\mathrm{l}(\omega)S(\omega) * H_\mathrm{r}^H(\omega)S^H(\omega)]}{\sqrt{\mathbb{E}[|H_\mathrm{l}(\omega)S(\omega)|^2]\mathbb{E}[|H_\mathrm{r}(\omega)S(\omega)|^2]}} \tag{3.6}$$

$$= \frac{H_\mathrm{l}(\omega) * H_\mathrm{r}^H(\omega)}{\sqrt{|H_\mathrm{l}(\omega)|^2 * |H_\mathrm{r}(\omega)|^2}} \tag{3.7}$$

$$= \mathrm{e}^{\mathrm{j}\angle H_\mathrm{l}(\omega) * H_\mathrm{r}^H(\omega)} \tag{3.8}$$

$$= \mathrm{e}^{\mathrm{j}\angle ITF}. \tag{3.9}$$

It can be seen that, for a point source, the IC magnitude is equal to 1. This value will be used as a reference throughout this work to always verify if the ITDs and ILDs found for the evaluated signals are reliable.

For a while, noise reduction techniques in hearing aids have been facing the big challenge of achieving a successful performance in noise reduction while also preserving the binaural cues of the sound sources as much as possible. Multichannel Wiener Filters (MWF) which aim to preserve the ITF, the RTF, the ITD, the ILD or the IC are common choices for this task [8–11]. Nevertheless, even though some MWF implementations might have had motivating results, there is still the problem concerning what the hearing aids do after applying noise reduction. Even if the noise reduction strategies do manage to preserve the binaural cues as much as possible, one

can naturally expect that the subsequent processing involving frequency-selective amplification and compression inevitably distorts the binaural cues anyway. For this reason, this work aims to evaluate the isolated effect that fitting and compression might have on the binaural cues, with the hope that the possible distortions caused by them can be better understood, while also raises a question whether the binaural cue preservation in noise reduction techniques really makes a difference at the of the processing chain for the hearing aid user. A better comprehension of the distortions on the binaural cues created by fitting and compression can also make room for future studies aiming to minimize them, so that a binaural cue preservation strategy can be created by considering the processing chain in the hearing aid as a whole, rather then assuring cue preservation only for specific steps of the processing.

Chapter 4

Experimental measurements

The theoretical analyses were performed by testing, for several compression and fitting conditions, three different signals of a database generated by the Signals, Multimedia and Telecommunications Laboratory (SMT) at the Federal University of Rio de Janeiro. This database contains 20 phonetically balanced sentence lists spoken by two female and two male speakers. The original signals were recorded at an acoustically treated environment with a high quality microphone, and were stored in mono 48 kHz@24-bit PCM Wave format. More details about this database, which is here referred to as the SMT database, can be found in Appendix B.

The three chosen signals of the SMT database were convolved with head-related impulse responses (HRIRs) from the Kayser et al. HRIR database [12] in order to generate spatialized versions of the original signals in an anechoic environment and in two reverberant environments.

The Kayser database was developed at the University of Oldenburg and has HRIRs recorded at an anechoic chamber and at several reverberant environments (e.g. office, cafeteria, courtyard). The recordings were performed with in-ear microphones placed at a head-and-torso simulator (HATS), as well with BTE hearing aids mounted at the same HATS with artificial ears. Therefore, each HRIR in the database was recorded with eight channels (corresponding to left and right in-ear microphones, and left and right front, middle and back BTE hearing aid microphones, respectively). For the sake of a better generalization, due to the fact that the subjective tests would be performed with normal-hearing listeners wearing headphones, only the channels corresponding to the left and right in-ear microphones were used for all HRIRs in this work.

For the anechoic chamber, impulse responses were measured for distances of 0.8 and 3 m between the speaker and the HATS. For each distance, HRIRs were recorded for azimuths of 0° (front) to -180° (left turn) in steps of 5°; and for elevations of -10° to 20° in steps of 10°. A figure with the coordinate system used for the anechoic chamber can be seen in Appendix A. For this work, only HRIRs measured for the

distance of 3 m between the loudspeaker and the HATS were chosen, in order to avoid the near-field effects presented in the HRIRs measured for distances of 0.8 m.

For the reverberant environments, impulse responses were measured for specific positions within the environment, sometimes for two different head orientations. For this work, the environments Office II and Cafeteria were chosen, since an environment with a shorter reverberation time and another with a longer one were desired (they have reverberation times of 300 and 1250 ms, respectively). Drawings with the room measurements, source positions and HATS positions for Office II and Cafeteria are also shown in Appendix A.

The following original signals of the SMT database, based on the phonetically balanced lists of sentences generated by Alcaim et al. [48] were chosen for the theoretical analyses:

- Sentence 1 of List 2, spoken by speaker F1: "Nosso telefone quebrou." (Our telephone broke). This signal will be referred to as F13;
- Sentence 8 of List 5, spoken by speaker F2: "Hoje eu irei precisar de você." (Today I am going to need you). This signal will be referred to as F2;
- Sentence 7 of List 2, spoken by speaker M2: "Hoje dormirei bem." (Today I am going to sleep well). This signal will be referred to as M22.

The theoretical measurements aim to show the potential effects of compression and fitting strategies on the binaural cues ITD, ILD and IC. Experiment 1 has spatialized versions of the three original sentences in an anechoic environment, whereas Experiment 2 has the three original signals placed at several different positions in the reverberant environments Office II and Cafeteria. For each of the experiments, three different cases were evaluated separately:

- Signals with only compression applied, no fitting strategy;
- Signals processed with a linear fitting strategy;
- Signals processed with a non-linear fitting strategy.

The first case is an attempt to evaluate the isolated influence of compression on the binaural cues, while the latter two cases are intended to evaluate how differently a linear and a non-linear prescription procedure affect the cues.

Some changes in the original binaural cues were expected for the three scenarios, especially in the ILD, since the compression changes the dynamic relations between high and low intensity sounds, and prescription procedures apply different amounts of gain to different frequency bands. Nevertheless, Experiments I and II presented their own particular characteristics and results, which are discussed in the following sections.

4.1 Experiment 1

Experiment 1 was conducted by using the three original signals convolved with four HRIRs of the Kayser database recorded in an anechoic chamber. Four anechoic spatialized signals were generated, all of them at a distance of 3 m from the listeners: F13 was placed at position $(0°, -90°)$, where $0°$ is the elevation and $-90°$ is the azimuth; F2 was used to generate signals at positions $(0°, -30°)$ and $(-10°, 180°)$, respectively; and M22 was placed at position $(20°, 60°)$. These four spatialized signals will be named, respectively, A, B, C and D for better readability. Positive azimuth angles are located at the right hemisphere, and positive elevations indicate sound sources above the horizontal plane.

After spatialization, the obtained signals were downsampled from 48 to 16 kHz. Since most traditional fitting strategies for hearing aids prescribe gains up to the 6-kHz or 8-kHz frequency band, it was considered reasonable to limit the bandwidth up to that region, in order to strictly analyze the behavior of the frequency regions manipulated by the fitting procedures adopted in this work and leave out the influence of higher frequencies.

Since it is usual for audio signals to be evaluated in terms of logarithmic sound power rather then their linear amplitudes, and since fitting and compression procedures all use the logarithmic scale, the amplitude of each signal in this work will be shown logarithmically, according to the following equation:

$$A = 20 * \log_{10}(\tilde{p}/p_0), \tag{4.1}$$

where A is the sound pressure level, $\tilde{p}$ is the root-mean-square (RMS) pressure from a specific time window (here chosen as 10 ms) and p_0 is a constant equal to 20 μPa. In an analog environment, an RMS pressure of 20 μPa, which is equivalent to a 0 dB SPL signal, would equal the human auditory threshold — that is, the least intense 1-kHz tone that the human ear can detect. Nevertheless, the signals are manipulated in a digital environment where quantization is envolved, so a 0 dB signal here does not necessarily have the same sound power as an actual 0 dB SPL signal. Therefore, this way of calculating sound pressure level is used only as a reference for the generation of signals, so that each time an "X-dB signal" is mentioned in this work, it only means that X dB is the resulting sound pressure level according to Equation 4.1 applied to a quantized digital signal. Signals that were subsequently evaluated in a listening test have gone through a sound calibration procedure that will be explained in Chapter 5.

Following Equation 4.1, the resulting downsampled spatialized signals have amplitudes that oscillate approximately between 20 and 40 dB. These low levels are expected since they mimic what would be heard in an anechoic chamber at a 3 m

distance from the sound source. The binaural cues ITD, ILD and IC were then calculated for the four resulting signals, in order to be used as a reference for the processed signals generated later for Analyses 1 and 2. For Analysis 3, an artificial gain was applied to all four signals before the binaural cue calculation; this will be better explained in Subsection 4.2.3.

For the calculation of the binaural cues, a Short-Time Fourier Transform was applied to frames of 0.5 s of each signal, with 50% overlap between frames. Since this work always considers the speech source to be approximately a point source, and since stationarity is assumed within each frame, Equations 3.1, 3.2, 3.3, 3.4 and 3.9 were considered valid to be applied here for binaural cue calculation.

The ITF for each frame was therefore calculated according to Equation 3.1 in order to obtain the cues by Equations 3.3, 3.4, and 3.5. Only for calculation of the ILD, a moving average was applied to the frame's left- and right-channel spectra and a smoothed version of the ITF was calculated from the smoothed spectra. This ITF was then used to obtain the ILD for the frame. Lastly, the final value for each binaural cue at a specific frequency is represented by the median value of this cue at that frequency across all frames of the signal.

An important characteristic of the Kayser database and the rigid sphere model that is used as reference in Kayser et al. [12] is that they do not present a maximum ILD at the 90° azimuth — the ILD follows instead a slightly different pattern. That is because, in reality, the human ears are not located exactly at ±90° azimuth, as assumed in Figures 3.3 and 3.4. The ears' position is actually slightly closer to the back of the head, at an azimuth of approximately ±100°. This is the approximate position of the ears of the HATS used for the recordings of the HRIRs of the Kayser database, and also the position considered for the rigid sphere model used as reference in Kayser et al. [12] for the estimation of the ITD and ILD values across azimuth.

Figures 4.1 and 4.2 show the ITD and ILD values across azimuth of the modeled HRTFs following the theoretical rigid sphere model of the head (red dashed lines) and of the measured HRTFs with the HATS (solid lines). While the ITD does present a parabolic-like shape with a maxima at −90° azimuth, the ILD's behavior is a bit more complex. The ILDs across frequency seem to increase from 0° to −60°, decrease again until 90°, increase until 120° and after that decrease continuosly until −180°. This behavior is due to the position of the ears. Therefore, we can assume that the ILD has its highest values at around ±60° and ±120°, and these are the figures that will be used as reference for the ITD and ILD values found for all the signals in this study. In Kayser et al. [12], due to the noticeable ripple that occurs at higher frequencies for the ILDs of the measured HRTFs, they are considered to match the reference ILD values until about 4 kHz. With this information and the

fact that the ITD and the ILD are more relevant for, respectively, the lower and the higher frequency regions, the frequencies which will be taken as reference for comparison between the cues measured in this work and the cues across azimuth according to Figures 4.1 and 4.2 are 500 Hz and 1 kHz for the ITD and 1 kHz and 4 kHz for the ILD.

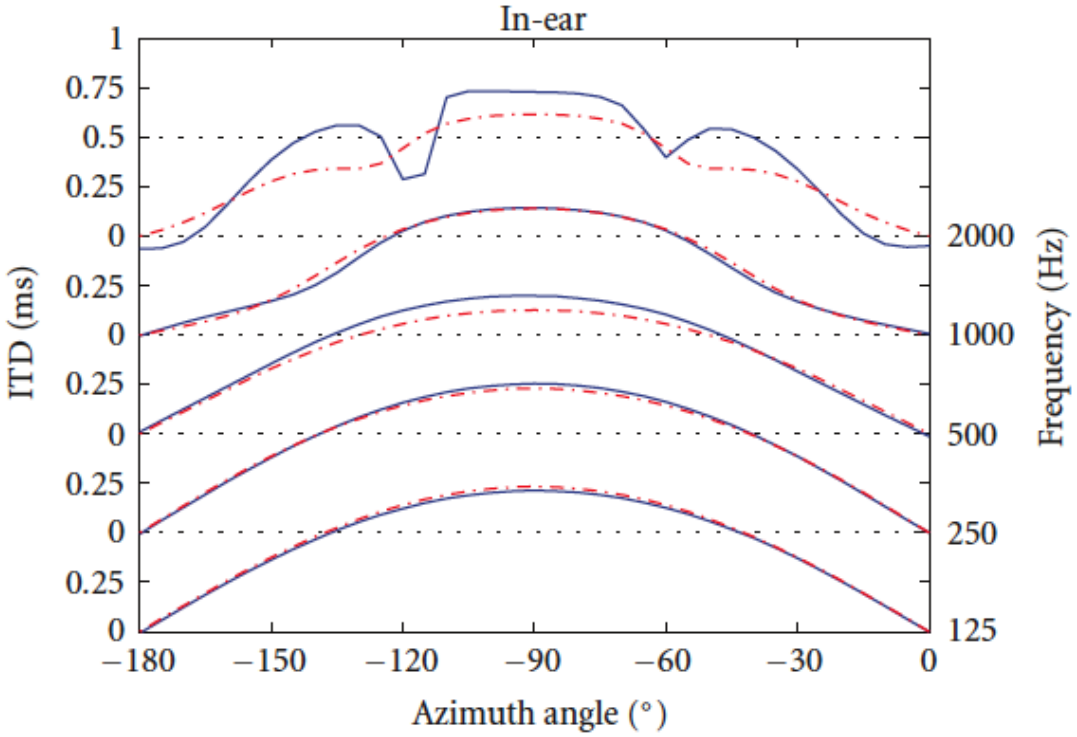


Figure 4.1: ITDs calculated from the measured (solid lines) and the modeled HRTFs (dashed lines) for the in-ear microphones. The ITDs for the mid frequencies in octaves from 125 Hz to 2 kHz are shown as indicated on the right-hand ordinate axis. An offset of 0.5 milliseconds is added to separate the curves from each other for a better overview. Ticks on the left-hand ordinate are 0.25 milliseconds apart. Source: Kayser et al. [12].

Figures 4.3 – 4.6 show the ILD, the ITD, the IC magnitude and the IC phase for signals A, B, C and D, respectively. It was observed, for those and also for every other signal measured in this work, that the calculated ITDs present heavy distortions for frequencies up to 500 Hz. This was unfortunate, since the ITDs are considered to be the most important cue for frequencies below 1.5 kHz. Nevertheless, in order to leave out the influence of distortions in the measurements, the ITD values are considered to be only valid for frequencies starting from 500 Hz. This is the criterion which will be used for all signals in this work, for all the analyses.

The magnitude of the IC, on the other hand, frequently presents several, if not many notches, along the frequency axis. This was interpreted as a lack of spectral content in those specific frequency regions in the signal. Lastly, the IC phase, which

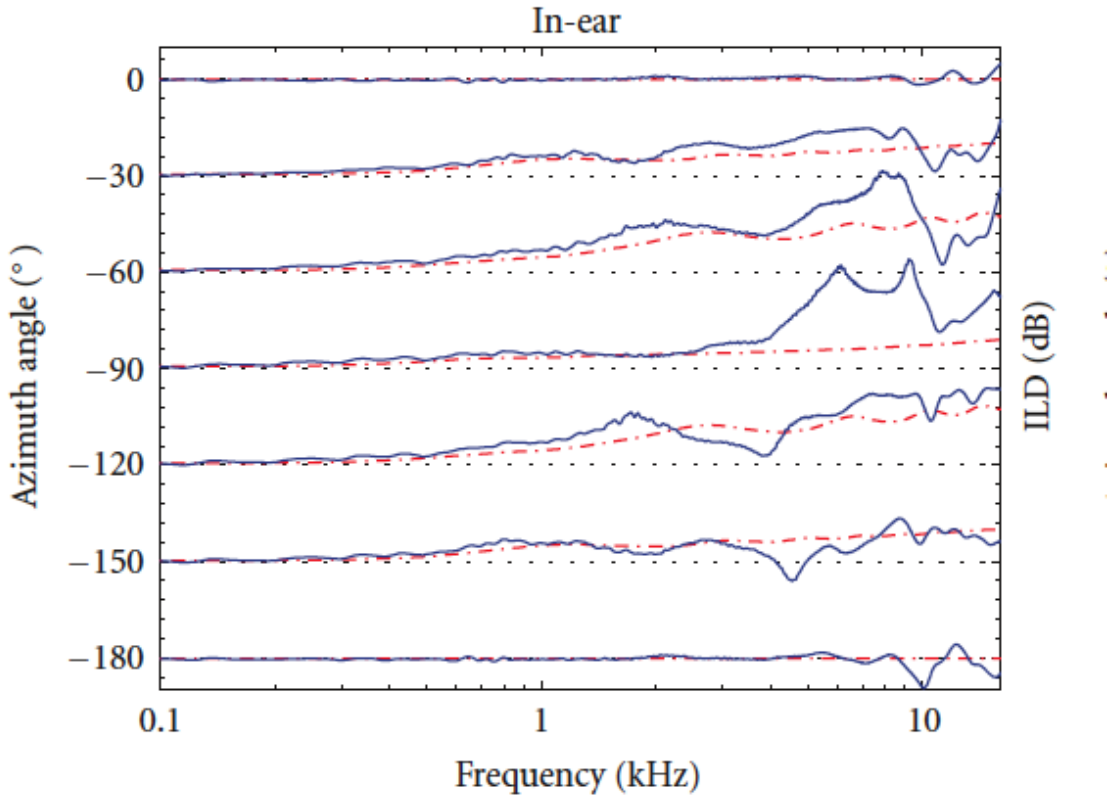


Figure 4.2: ILDs calculated from the measured (solid lines) and the modeled HRTFs (dashed lines) for the in-ear microphones. Ticks on the right ordinate are 6 dB apart, and dashed straight lines correspond to an ILD of 0 dB. Source: Kayser et al. [12].

is always approximately linear if it is seen with a linear frequency scale, shows a positive slope when the signal is coming from the left hemisphere, and a negative one when it comes from the right hemisphere, as expected, given the way it is calculated. Figure 4.7 shows the IC phase from signal A with a linear frequency scale.

For the comparison between the obtained ITDs and ILDs and the reference values shown in Figures 4.1 and 4.2, a symmetry between both hemispheres in relation to the mid-sagittal plane is assumed, so that an ITD or ILD for an azimuth angle at the right hemisphere has the same absolute value as that found at the symmetrical azimuth angle at the left hemisphere.

The ITDs of the modeled HRTFs (solid lines) in Figure 4.1 are used as reference for comparisons with the measured ITDs in this work. Signal A has an ITD of around 750 μs at 500 Hz and of around 700 μs at 1 kHz, which are slightly overestimated in comparison the approximate values of 700 ms and 650 ms in Figure 4.1 for 500 Hz and 1 kHz, respectively. Signal B has an ITD which is a bit less than 400 ms at 500 kHz and a bit less than 300 ms at 1 kHz. These are also a bit overestimated when compared to the reference values of approximately 300 ms and 200 ms for 500 Hz

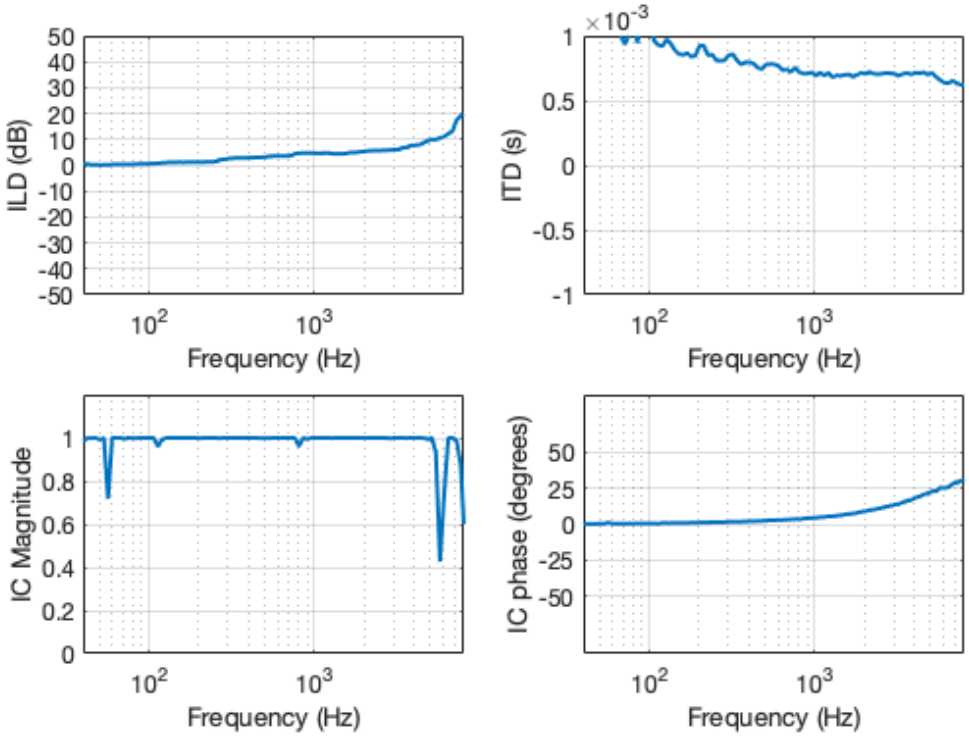


Figure 4.3: Binaural cues for signal A. Source at position $(0°, -90°)$.

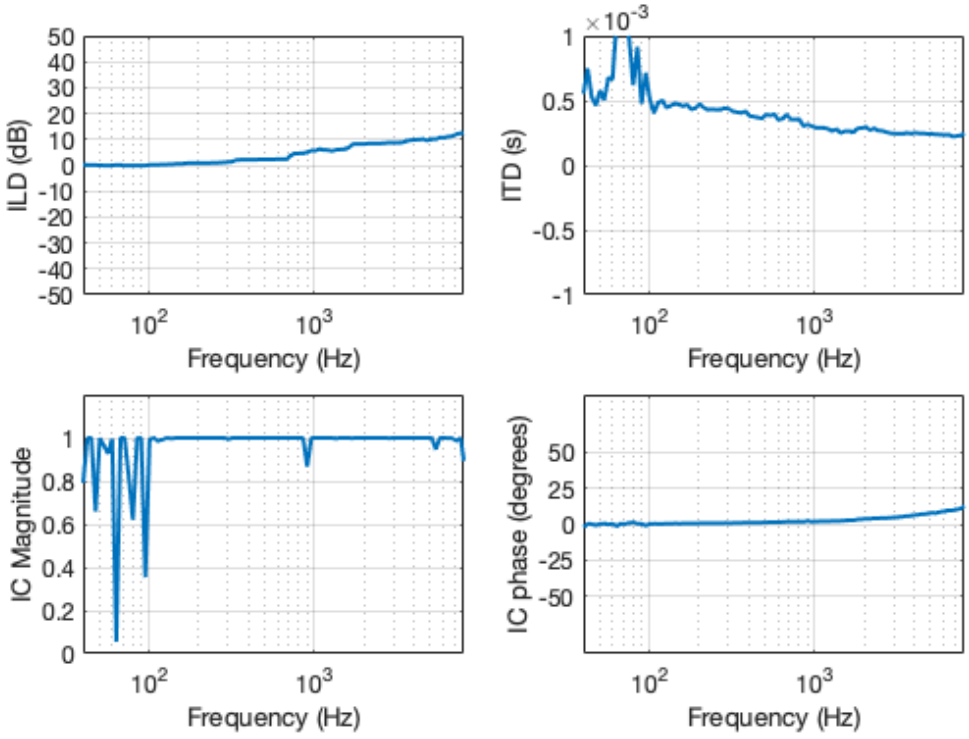


Figure 4.4: Binaural cues for signal B. Source at position $(0°, -30°)$.

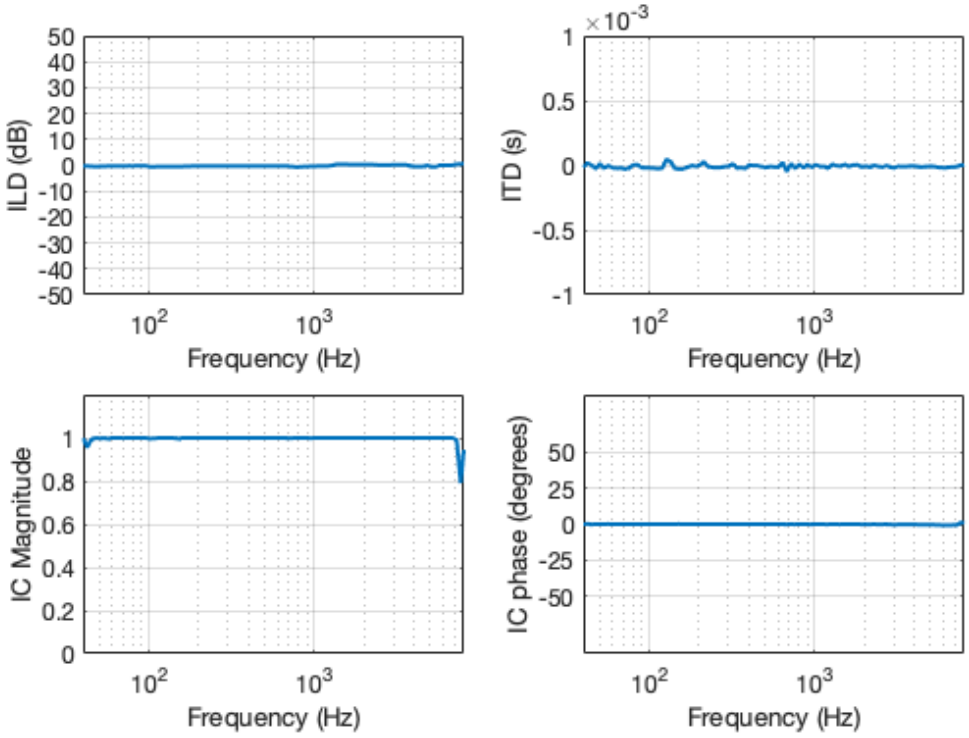


Figure 4.5: Binaural cues for signal C. Source at position $(-10°, -180°)$.

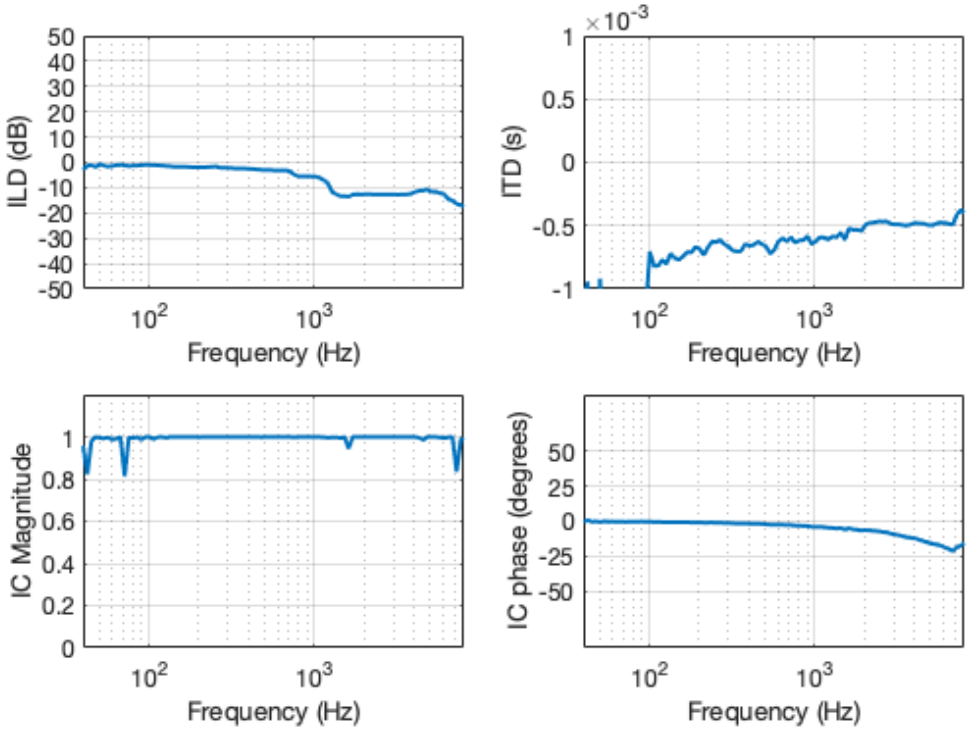


Figure 4.6: Binaural cues for signal D. Source at position $(20°, 60°)$.

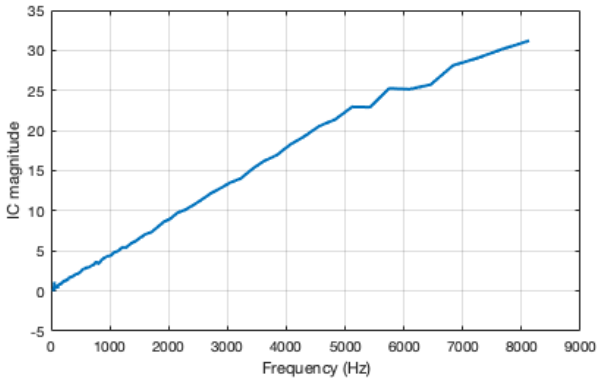


Figure 4.7: Phase of the IC from signal A plotted with a linear frequency scale.

and 1 kHz, respectively. Signal C presents an ITD of approximately zero along the entire frequency axis, which is expected and is in agreement with Figure 4.1, since the signal is located at an azimuth of $\pm 180°$. Lastly, signal D has absolute ITDs of around 660 ms and 620 ms at 500 kHz and 1 kHz, respectively, which are, as in signal B, a bit overestimated in comparison to the expected values of approximately 600 ms and 550 ms in Figure 4.1.

The analysis of the ILD was performed similarly, by comparing the measured ILDs with the ILD estimations illustrated by the solid lines in Figure 4.2. Signal A has ILDs of 5 and 8 dB at 1 kHz and 4 kHz, respectively, which are in accordance to the estimated values for these frequencies. Signal B has ILDs of approximately 6 and 10 dB at 1 and 4 kHz, which also match the reference values for these frequencies at this azimuth. Signal C, whose ILD is approximately zero, is in complete accordance with Figure 4.2. Lastly, signal D has absolute ILDs of around 6 and 12 dB at 1 kHz and 4 kHz, respectively, which is consistent with the estimated ILD for these frequencies at the symmetric position of $-60°$.

The slight overestimation of the ITDs which range from around 50 to 100 ms are possibly due to the characteristics of the signals in this work, since the ITDs here are calculated directly from the signals and not from the HRTFs. Nevertheless, they correctly follow the pattern of increasing as the source moves closer to the side of the head, and also preserve the relative difference in value for the two different frequencies analyzed. With that in mind, the binaural cue measurements in this work are considered to be reliable, and the measured binaural cues for signals A, B, C and D will be considered as references for all the analyses that follow.

For this work, it was originally planned to analyze the effects of compression and frequency-selective amplification on both azimuth and elevation, hence the non-zero elevation angles present in two of the four reference signals. However, since binaural cues are related to azimuth rather than elevation, it was chosen to focus on the evaluation of the effect of these types of processing on azimuths alone.

4.1.1 Analysis 1

Analysis 1 demonstrates the isolated effect of single-channel compression on signals A, B, C and D. The four signals were submitted to a single-channel compression algorithm with adjustable values for compression threshold, compression ratio, attack time and release time for each ear, and the final processed signals were obtained.

This step was done in three different ways: identical compression configuration in both ears; compression applied solely to the left ear; and compression applied solely to the right ear. Lastly, the ILD, the ITD and the IC were calculated for all generated signals and then compared to the original cues of signals A, B, C and D.

It was initially intended to observe the measurements obtained for a signal which is virtually entirely above the compression threshold — that is, a signal in which compression is active all the time. For that, a compression threshold of 0 dB was chosen. A moderate compression ratio of 3:1 was used. A fast enough attack time to act on abrupt level changes was desired, so a value of 10 ms was chosen, and the longer release time of 200 ms was used in order to avoid a "pumping" effect from being perceived. For each signal, this set of configurations was applied firstly to both ears, then exclusively to the left ear, then exclusively to the right ear.

Figure 4.8 shows the binaural cues for signal A and its respective compressed versions, where compression is applied to both ears, solely to the left ear, and solely to the right ear, respectively. The cues of the uncompressed signal are represented with a blue line. It can be seen that for the case in which both ears are submitted to the same compression configuration, the changes in ITD and ILD are very subtle. When compression is applied to only one ear, however, the ILD changes proportionally, since the gain reduction caused by compression increases the difference in level between both ears. When listening to the signal, this can be perceived as a change in source position — that is, when compression is applied to the left ear only, the signal sounds louder at the right side and the source appears to be coming from the right, rather than from the left: and when compression is applied to the right ear only, the sound is perceived as coming from the left side, where the sound source originally is, even though there is less sense of externalization and spatialization when compared to the uncompressed signal.

The ITD, on the other hand, presents some distortions when compression is applied to one ear only. Since the changes in ILD are very explicit, it is hard to evaluate the effect of the distorted ITDs alone in the subjective perception of the sound source's position. The IC phase presents no relevant change, while the IC magnitude has more notches for the compressed signals in all three cases. The increase in notches is observed for all types of processing in all analyses performed for Experiment 1 in a similar way than what is shown in Figure 4.8. For signal A

and its compressed versions, the median values of the IC magnitude are all above 0.98, which shows that our ILD and ITD measurements are reliable. Since the IC shows a similar behavior regarding the median value for the subsequent signals in this work, the ITD and ILD are from now on implicitly considered to be reliable.

Similar phenomena regarding the cues and localization were observed for the other three signals, whose binaural cues are shown in Figures 4.9–4.11. When compression is applied to one ear only, the greater the ILD, the more the perceived source position leans to the uncompressed side. The ITD, on the other hand, presents approximately the same shape for most frequencies, but appears to be "changing offset" slightly. This is interpreted as being an effect of compression, especially if it is used in a more aggressive way (as it is the case here, since a low CT and a moderate CR are used), given that its implementation involves delay and other types of processing that may modify the ITD values. The IC magnitudes have median values above 0.97 for signals B, C and D, and all their compressed versions, which secures reliability of ITDs and ILDs. In the compressed versions of all four signals, one can sometimes perceive, in the channels where compression is applied, an abrupt decrease in level after a sudden peak, which happens before compression has time to act.

A further analysis was done for signal A to examine the isolated effect of compression threshold on the binaural cues, when compression is applied to one ear only. The result are presented in Figure 4.12 for compression thresholds of 0 dB, 20 dB and 40 dB. The other parameters were kept the same as in the previous analyses (CR = 3:1, AT = 10 ms and RT = 200 ms). As expected, the higher the compression threshold, the softer the gain reduction caused by compression. Therefore, the ILD proportionally diminishes with the increase of compression threshold. This phenomenon is also observed even when compression is applied to both ears, given that the reduction is greater at the louder side, consequently diminishing the difference in level between both channels. No changes on the ITD and on the IC were observed, in comparison to signal A's original cues.

Lastly, a study of the effect of attack and release time on the binaural cues was done for signal D, when compression is applied solely to the right ear. In Figure 4.13, results are plotted for attack times of 10, 50 and 100 ms, while the release time is kept at 200 ms. The cues for the uncompressed signal are also shown. The CT was set to 30 dB in order to make the signal go in and out of compression several times. CR is kept at 3:1, as in the previous analyses. Figure 4.14, with the same values of CT and CR, show the results for release times of 10, 50 and 200 ms, while the attack time is kept constant at 10 ms (CT = 30 dB, CR = 3:1), as well as for the uncompressed signal.

It is noticed that for frequencies below 1 kHz in both figures, the ILD, which was originally negative at all frequencies (since the signal is positioned at the right hemi-

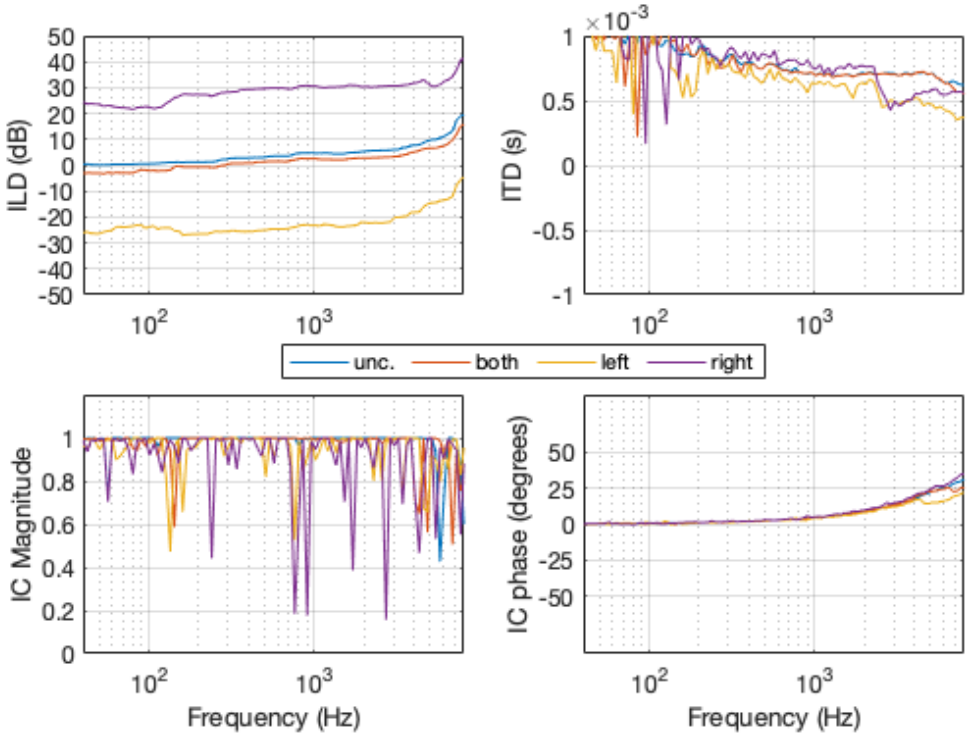


Figure 4.8: Signal A uncompressed (blue line) and with compression applied at both ears, only at the left ear, and only at the right ear.

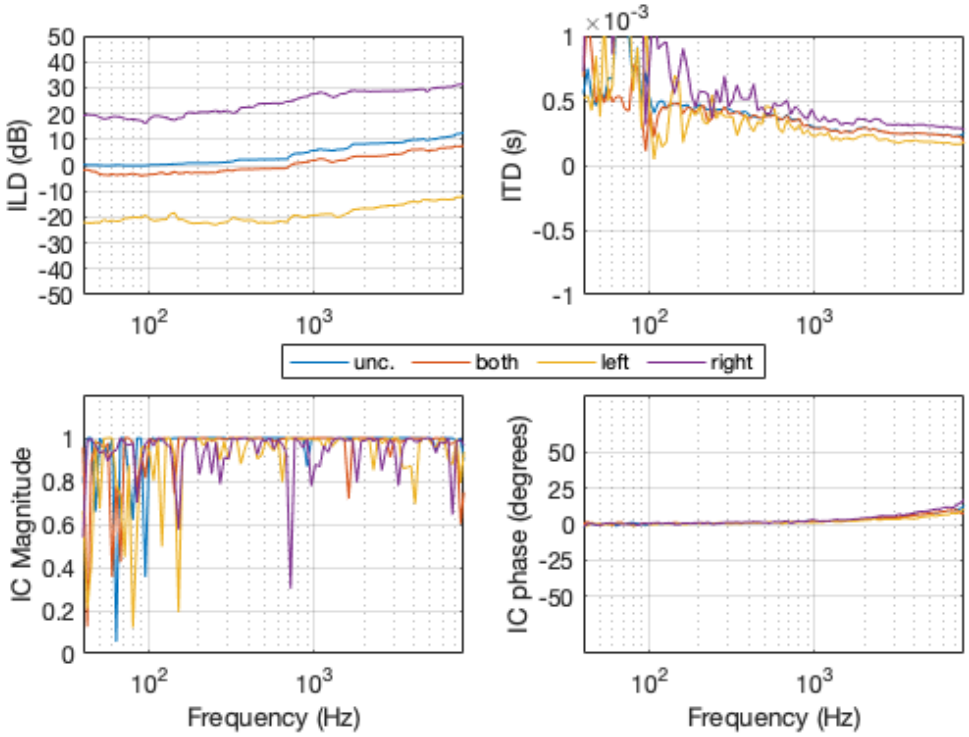


Figure 4.9: Signal B uncompressed (blue line) and with compression applied at both ears, only at the left ear, and only at the right ear.

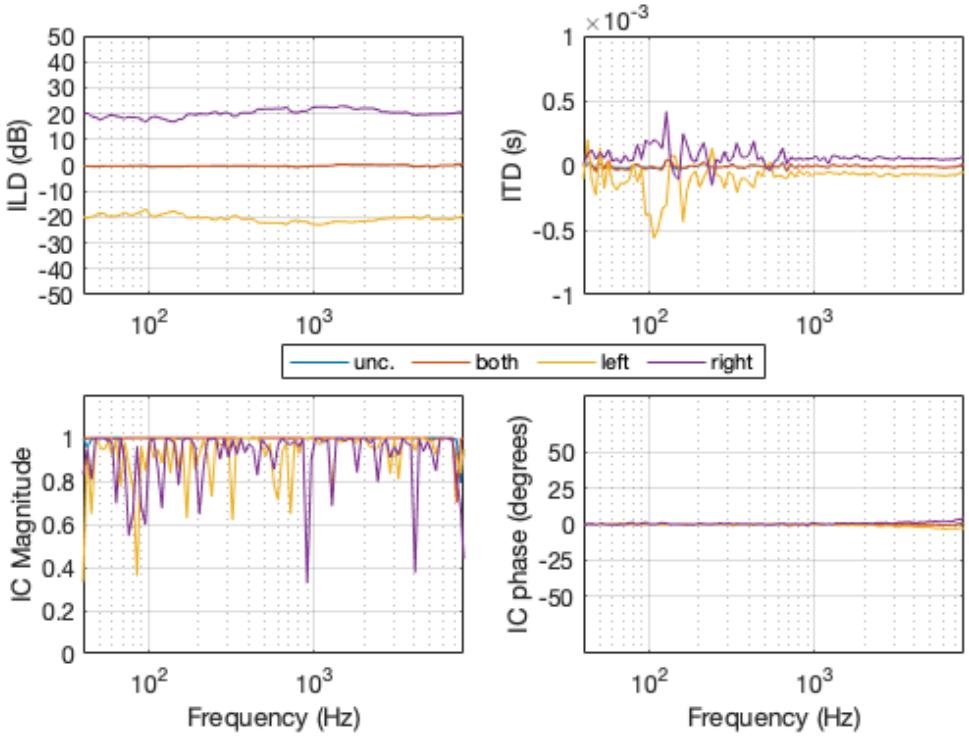


Figure 4.10: Signal C uncompressed (blue line) and with compression applied at both ears, only at the left ear, and only at the right ear.

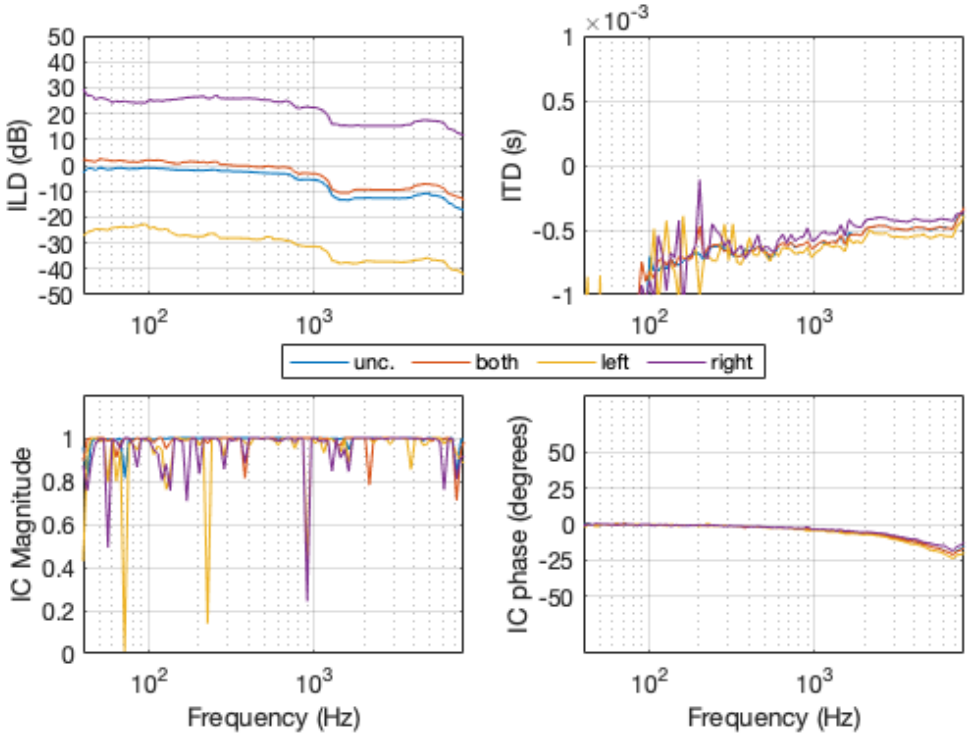


Figure 4.11: Signal D uncompressed (blue line) and with compression applied at both ears, only at the left ear, and only at the right ear.

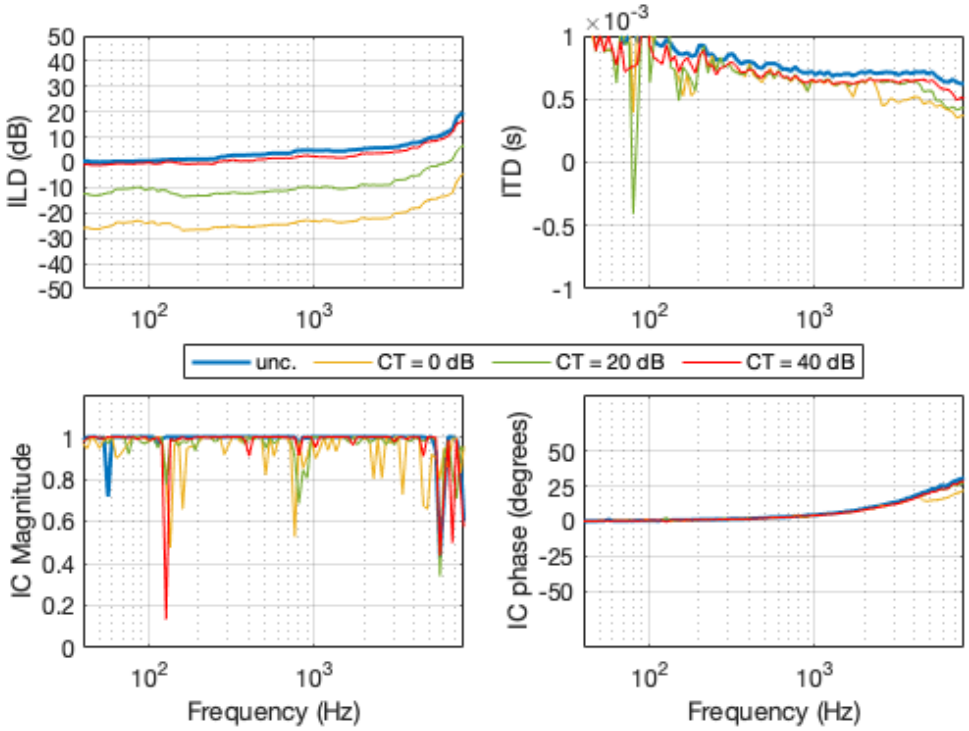


Figure 4.12: Effect of compression threshold on ILD and ITD when compression is applied on the left ear, for signal A. Original uncompressed signal is represented by a thick blue line.

sphere), becomes positive. This happened because the gain reduction at the right channel was so great for the lower frenquencies that the levels at that channel are now smaller than those in the opposite channel for these frequencies. Nevertheless, an effect of the variation of ATs and RTs can still be observed. When the attack time is shorter, compression starts acting faster, thus making the signal suffer the effects of compression (that is, gain reduction) for a slightly longer time. Therefore, shorter ATs produce a slightly greater change in ILD than longer ATs in relation to the ILDs of the uncompressed signal. Conversely, when the release time is shorter, the signal goes faster out of compression, and thus the change in ILD is slightly smaller than that verified when longer RTs are used. The effects can be perceived when listening to the signals, although very subtly, as a slight gain reduction when ATs are shorter (as opposed to longer ATs) and when RTs are longer (as opposed to shorter RTs). The ITDs did not change with the variation of AT and RT, maintaing that same slight decrease in magnitude that was observed in Figure 4.11 for compression on the right channel.

No relevant changes in the IC were observed, so the plots regarding IC were omitted from Figures 4.13 and 4.14 to avoid redundancy.

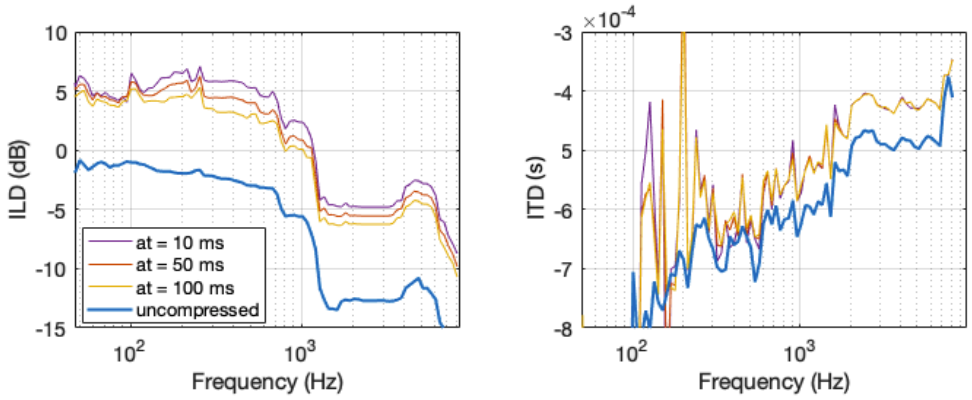


Figure 4.13: Evaluation of the effects of the variation of attack time on ILD and ITD when compression is applied on the right ear, for signal D. The uncompressed signal's cues are represented by a thick blue line.

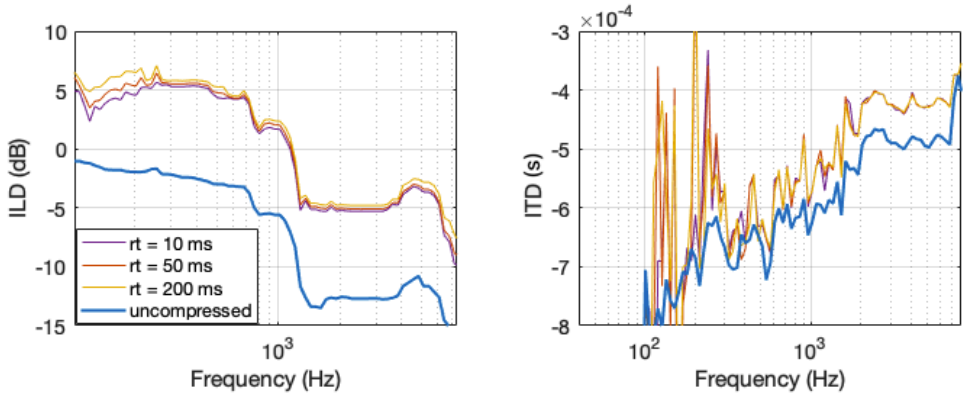


Figure 4.14: Evaluation of the effects of the variation of release time on ILD and ITD when compression is applied on the right ear, for signal D. The uncompressed signal's cues are represented by a thick blue line.

4.1.2 Analysis 2

Analysis 2 aims to simulate a linear fitting in a similar way that would be done with a hearing aid that uses linear amplification. The NAL-R/NAL-RP method, which is described in Chapter 2, was chosen for this purpose due to its replicability and versatility.

Firstly, it was necessary to simulate a hearing loss. The standard audiograms proposed by Bisgaard [49] were used as a reference, and two of them were chosen for this analysis. One is an audiogram for a moderately sloping moderate hearing loss, named N3, and the other represents a steep sloping moderate-to-severe hearing loss, named S3. It must be noted that the criterion for the degree of hearing loss used by Bisgaard is slightly different than that presented at Table 2.4.

For each of two types of loss, three situations were tested: hearing loss at both ears; hearing loss solely at the left ear; and hearing loss solely at the right ear. The

same signals A, B, C and D were used as input.

A 64-channel cosine modulated filter bank was generated according to the formulas described in Vaidhyanathan [50], and several filters were grouped to form larger filters with varying lengths, each of them centered approximately at each frequency used by the NAL-R fitting procedure (250, 500, 750, 1000, 1500, 2000, 3000, 4000 and 6000 Hz). Each of the filters representing the nine bands was multiplied by the gain prescribed by NAL-R for that frequency band and for the amount of hearing loss at that band. The modified filter bank was then convolved with the input signal.

Figure 4.15 shows the magnitude and phase response of the effective filter bank. For bands that have different intervals to the previous and to the next band (e.g. the 1000-Hz band is 250 Hz away from the 750-Hz band and 500 Hz away from the 1500-Hz band), the center of the passband of the filter representing this band does not exactly correspond to the band's nominal frequency; even so, the nominal frequency is always located within the filter's passband.

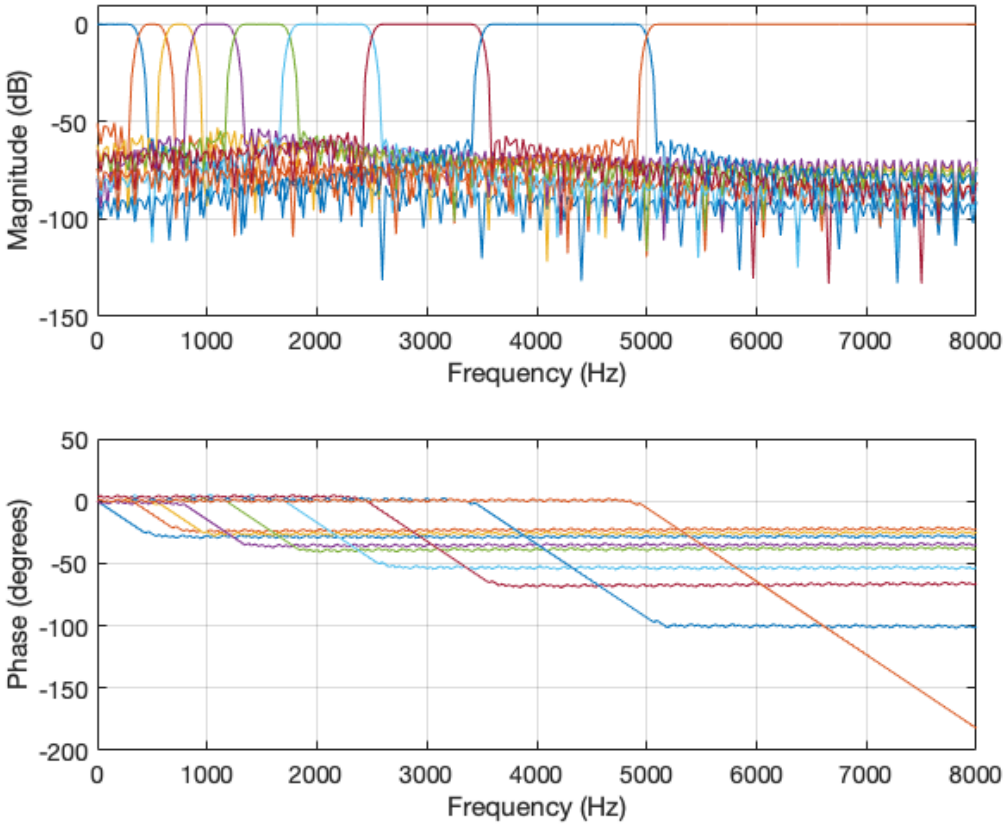


Figure 4.15: Magnitude and phase respond of the filter bank.

After that, a hearing loss compensation had to be performed so that a normal-hearing listener could hear the signal as closely as possible to what a person with that specific hearing loss configuration would hear. In other words, it was necessary

to "transform" the normal-hearing listener into a hearing-impaired listener. The resulting gain for each frequency band in decibels is, therefore, the gain prescribed by NAL-R for the chosen hearing loss pattern minus the actual hearing loss that the listener would have at that frequency.

After this compensating gain is applied, the output signals are obtained. When the hearing loss occurs at both ears, the binaural cues from the new signal are virtually the same or very close to the cues from the input signal. However, a problem occurs when dealing with unilateral hearing losses. When a hearing loss pattern is applied to one ear only and the whole processing chain (filtering, amplification and compensation) is applied at this ear, the final result is a signal with a much smaller sound power on the hearing-impaired side than on the normal hearing side. This happens because, as previously said, the resulting gain, in decibels, on the side of the hearing loss for each frequency band is the gain applied by the fitting procedure minus the hearing loss at that band. When converted to linear form, this resulting gain is always less than 1 for all frequency bands, given that the fitting procedure always prescribes a smaller gain than the hearing loss itself at each frequency band. Therefore, the final scenario for unilateral losses is:

- A channel with no hearing loss at all, which does not suffer any form of processing or has any gain applied;
- A channel that presents a hearing loss pattern and ends up having a smaller sound power than before, given that the resulting linear gain applied to all frequency bands is less than 1.

This difference in sound power between both channels distorts the ILD and makes no room for a subjective evaluation of the changes in the spectrum caused by the fitting procedure itself. Therefore, another adjustment has to be made for unilateral losses. This consists of multiplying the channel with no hearing loss by a constant smaller than 1 that makes the resulting signals on both channels have approximately the same average sound power. This constant k_u is calculated as

$$k_u = \frac{\sum_{n=1}^{N} 10^{(G_v - A)_n/20}}{N}, \tag{4.2}$$

where u and v are the channels without and with hearing loss, respectively; $\mathbf{G}_v$ is the vector with the gains prescribed by the fitting procedure for each frequency band, for the channel with the hearing loss; $\mathbf{A}$, from audiogram, is the vector with the hearing loss pattern (N3 or S3) across frequency bands; and N is the number of frequency bands, indexed by n.

Figure 4.16 shows the original binaural cues for signal C, which is located at (-10°,180°), as well as the binaural cues for the processed versions of signal C using the N3 audiogram. The ITD and the IC are basically the same for all cases. As expected, the changes occur basically in the ILD. One can see that the ILD when the hearing loss is applied to both ears is identical to the original one. Nevertheless, when the loss is unilateral, the aforementioned effect is shown: the lines described as "left" and "right" correspond to ILDs when there is unilateral loss at the left and at the right ear, respectively; they not only have a different shape from the original ILD, but they have an offset which is very far from zero, due to the difference in sound power between the channels with and without the hearing loss. Lines designated as "left k" and "right k", on the other hand, show the same two cases, but now with an equalization generated by constant k_u (from now on referred to as only k, for simplification), which assures that both channels have approximately the same average sound power. One can see that their shape is the same as in the previous case, but now they are located around zero, which is where they are supposed to be.

Some interesting effects can clearly be noticed by listening to all the output signals. Without the equalization provided by constant k on the normal-hearing side, the sound source appears to be at the right hemisphere when the hearing loss is on the left side, and at the left hemisphere when the hearing loss is at the right side. When the normal-hearing side is equalized with constant k, the signal is again perceived at the center, but the sound source appears to be more spread out, possibly due to the fact that the ILD for some frequencies becomes slightly negative, and for others it becomes slightly positive. The difference in sound quality between both channels, the side with the hearing impairment often sounding more muffled than the other side, always generates some confusion regarding the location of the source.

Similarly, Figure 4.17 shows the binaural cues for the same conditions by using audiogram S3 instead. New constants k were calculated for the normal-hearing channel when audiogram S3 is used on the other side. The effects in the binaural cues are identical to those observed for the cases where audiogram N3 was used. Nevertheless, since the S3 curve represents a more aggressive hearing loss, the signal is a bit more modified and thus the ILDs are a bit more distorted in shape than the ones calculated for audiogram N3. This translates in perception as a subtle loss in high frequency definition and as greater ambiguity concerning the sound source's location, although the main part of the signal still sounds as if it is coming from the original location.

A very similar behavior for all binaural cues was observed for signals A, B and D, for bilateral and unilateral losses and the two different reference audiograms. The sensation of a more ambiguous and spread-out sound source position was also

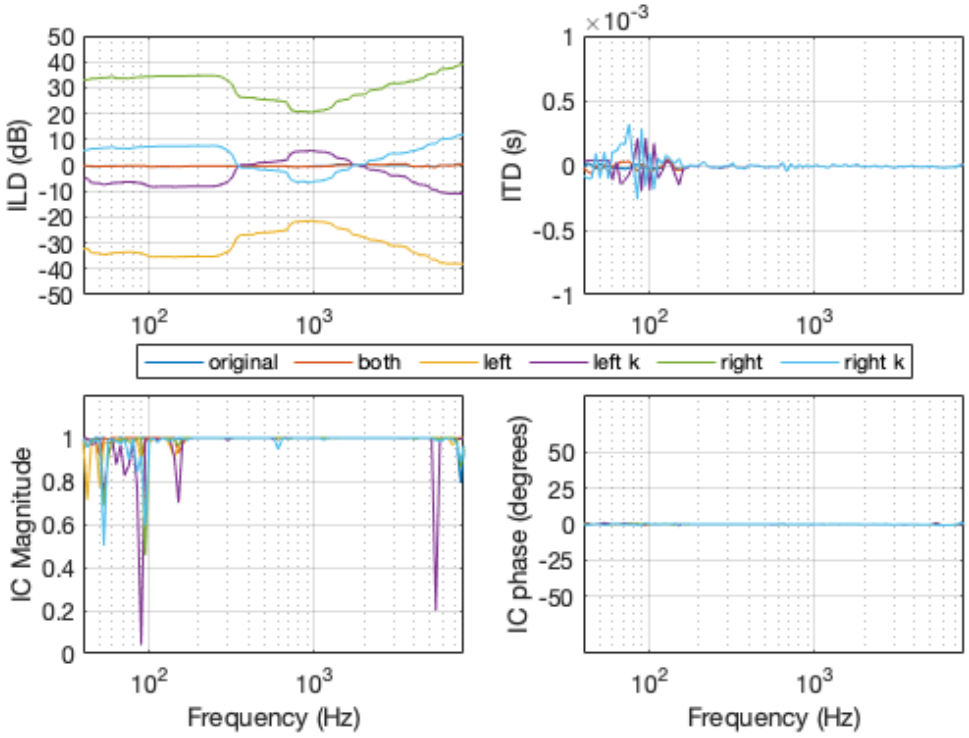


Figure 4.16: Binaural cues from signal C with N3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

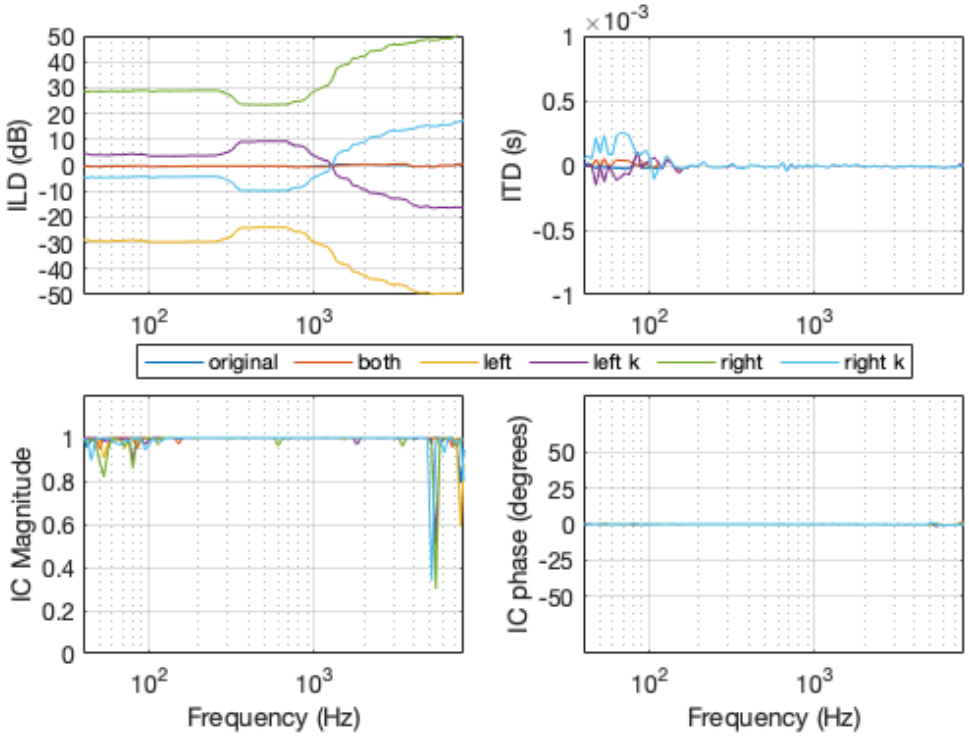


Figure 4.17: Binaural cues from signal C with S3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

perceived for the cases in which unilateral loss with equalization with constant k was applied, in comparison to bilateral loss, even though most of the signal's content still sounds as if it is coming from the correct position, or at least the correct quadrant. It was noted that signals with hearing loss applied at the ear that is closer to the source have more distortions concerning the source position, in comparison to signals with the hearing loss applied to the further ear. In the former case, most part of the signal's content is still perceived at the original position, but the source sounds less externalized and the sensation of spatialization is somewhat damaged. Figures 4.18 and 4.19 show the binaural cues for signal A in all hearing loss configurations for audiograms N3 and S3, respectively. The same plots for signals B and D with the multiple hearing loss configurations are shown in Appendix A.

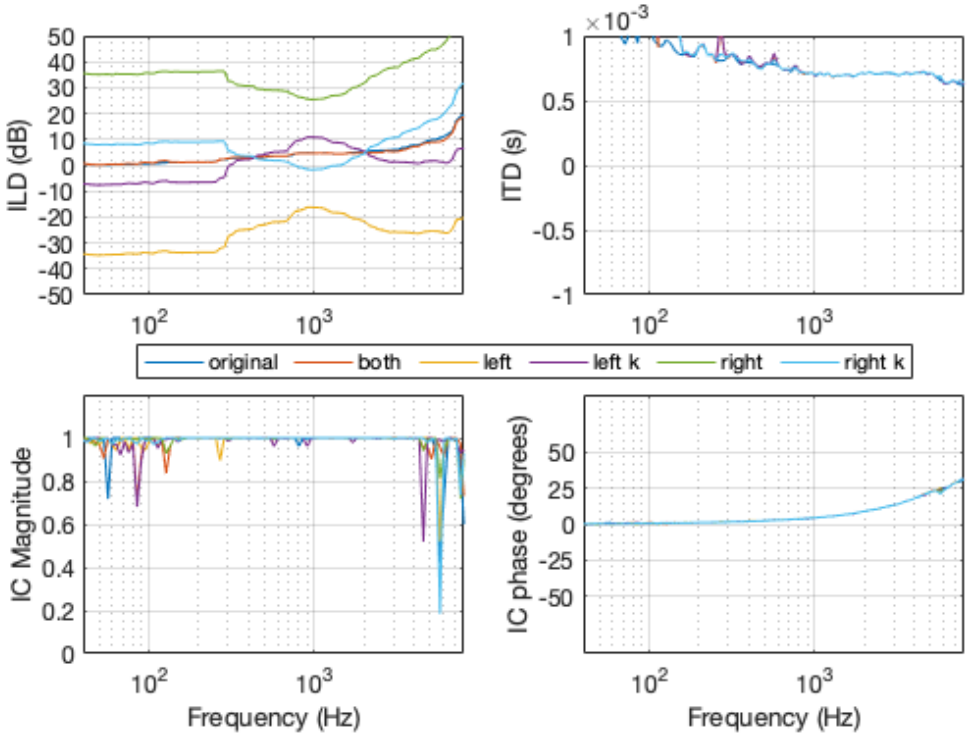


Figure 4.18: Binaural cues from signal A with N3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

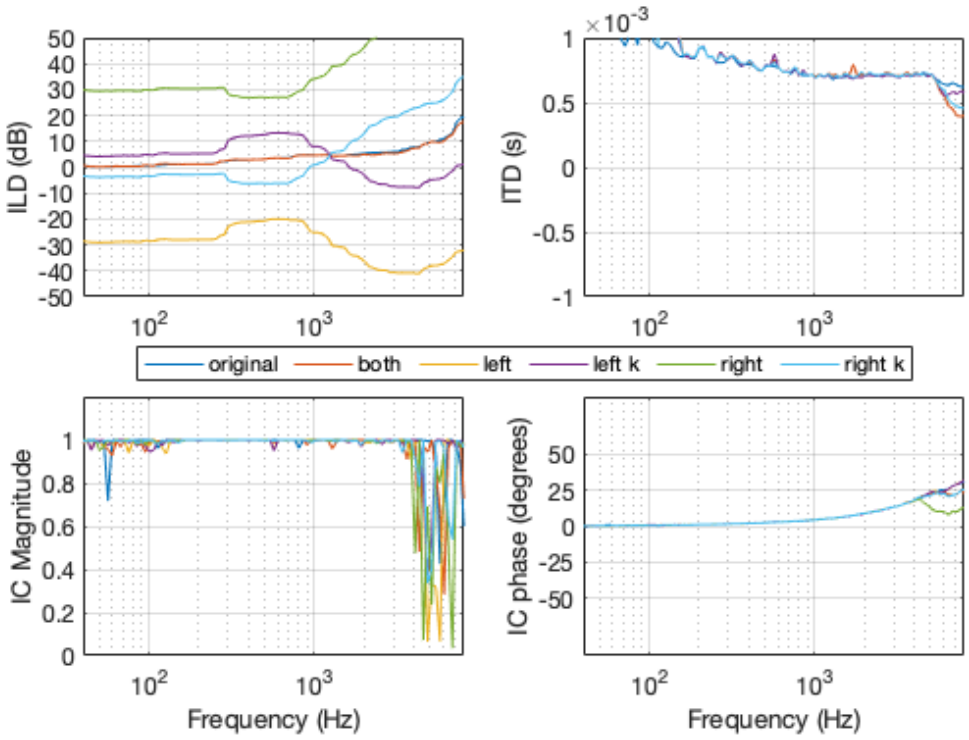


Figure 4.19: Binaural cues from signal A with S3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

4.1.3 Analysis 3

Analysis 3, as the more complex analysis of Experiment 1, aims to simulate the effect of a non-linear fitting procedure. The FIG-6 procedure was chosen due to its reproducibility, since most of the other fitting procedures have fewer information about their implementations and some have proprietary code which can not be reproduced in another platform.

Since FIG-6 has different prescription rules for low (40 dB), typical (65 dB) and high (95 dB) input levels, an artificial gain had to be applied to spatialized signals A, B, C and D, in order to make all four signals have an average sound level above "low" and at least part of the signal to go above "typical". Given that levels of 95 dB and higher would not be possible to be reproduced comfortably in a subjective test, none of the evaluated signals reaches such high input levels. The levels of signals A, B, C and D were adjusted so that their average sound level, according to Equation 4.1, lie between 55 and 70 dB and their peak levels are at about 80 dB.

Similarly to what was done for Analysis 2, two types of hearing loss were simulated, being represented by the same audiograms N3 and S3. For each type of loss, the same three conditions of Analysis 2 were tested: hearing loss in both ears, hearing loss in the left ear only, and hearing loss in the right ear only.

The generated signals were separated by the same filter bank into the same nine frequency bands used for the NAL-R procedure as in Analysis 2. Each signal was analysed in frames of 10 ms. The RMS level of the frame was calculated by using as a full scale reference the instantaneous peak sound level, according to Equation 4.1, calculated for a 2-kHz sinusoid with 16 kHz sampling rate and amplitude 1, which is 94 dB inside the software. This value does not correspond to the real SPL value and is valid only inside MATLAB (where the processing was performed), but it was chosen as a simple and efficient way to provide a reference and preserve the proportional differences in sound power between signals.

Based on the sound level calculated for the frame, the corresponding prescription rule was applied to each frequency band within the frame according to the audiogram of each ear. The prescription rules are given by Equations 2.4–2.6, following the reference input levels 40, 65 and 95 dB. For input levels below 40 dB and between two different reference levels, the prescription functions were linearly interpolated.

In order to avoid abrupt changes in gain from one frame to the other, gain was increased or decreased progressively in each band according to previously chosen parameters attack time and release time. For all signals in Analysis 3, attack time was set as 10 ms and release time as 200 ms. The linear gain for each band b in each frame n was therefore applied in the following way:

```
at = 0.01;
rt = 0.2;

for n=1:nframes
% // input level of the frame is calculated
% // a gain G(n,b) is calculated in decibels for each band
% // according to FIG-6

% conversion to linear
f = 10.^(G(n,:)/20);
    for b=1:nbands
        if f(b) < g(b)
            c = 1-exp(-log(9)/(fs*at));
            % fs: sampling rate
        else
            c = 1-exp(-log(9)/(fs*rt));
        end
        g(b) = (1-c) * g(b) + c * f(b);

        % // g(b) multiplies each band of the frame
```

end

end

The coefficients c based on the attack and release times are calculated, as can be seen in the pseudo-code, in the following way:

$$c_{\mathrm{AT}} = 1 - \mathrm{e}^{-log(9)/(F_{\mathrm{s}} * AT)}\,; \tag{4.3}$$

$$c_{\mathrm{RT}} = 1 - \mathrm{e}^{-log(9)/(F_{\mathrm{s}} * RT)}\,, \tag{4.4}$$

where AT and RT are the chosen attack time and the chosen release time, respectively, and F_{s} is the sampling rate of the signal. These two expressions were taken from Zölzer [51], which contains some examples of compressor structures that helped the implementation of parts of the code for Analysis 3. Further details and derivations for Equations (4.3) and (4.4) can be found in McNally [52] and Zölzer [53].

After the gains were applied and the first output signals were generated, it was necessary, as in Analysis 2, to include a hearing loss compensation for each signal, in order to simulate a hearing loss in the normal-hearing listener. Each frequency band of the filter bank was multiplied by a gain which equals the opposite of the actual hearing loss at that band according to the audiogram used, and the resulting filter bank was convolved with the output signal's channels corresponding to the hearing-impaired ears.

The fitting and the hearing loss compensation generate a difference in sound power between both channels for the cases with unilateral loss, since the resulting gain applied to the hearing-impaired ear, due to the hearing loss compensation, is smaller than 1. In order to adjust this, an equalization between the sound powers of both channels was performed by multiplying the channel with no hearing loss by a constant k, similarly to what was done for Analysis 2. For Analysis 3, however, this procedure is done differently, since the gains in a non-linear fitting procedure are input-level dependent. Therefore, a white-noise signal with 16 kHz sampling rate and a median sound level, calculated in frames of 10 ms with a full-scale reference of 94 dB, as in the implemented FIG-6 procedure, of around 30 dB (which is below the 40-dB "low-level" reference according to the FIG-6 procedure) at position $(0°, 0°)$ in the anechoic chamber was generated as reference, and the FIG-6 gains were calculated, first for audiogram N3, than for audiogram S3, both applied in one ear only. The gains for each frequency band from a frame at around half of the signal duration was extracted, and then the hearing loss in the respective frequency bands were subtracted from them. The resulting gains, when converted to linear, are smaller than 1 for all frequency bands. Finally, the mean gain across frequency bands is calculated, and this is the constant k that multiplies the channel with no hearing loss. There is one specific k for each audiogram (N3 and S3), as in Analysis 2.

It was observed empirically that using a reference white noise signal with a median sound level as described above — which also happens to present instantaneous peak sound levels at around 40 dB — enabled the calculation of constants which proved to be robust in equalizing sound power between the left and right channels for the signals analyzed in this study.

After this procedure, the binaural cues were calculated for all output signals. Figure 4.20 shows the ITD, the ILD, and the magnitude and phase of the IC for signal C and all its processed versions with audiogram N3. The lines described as "left" and "right" represent the binaural cues of the signals with unilateral loss without energy equalization between channels, while the lines named as "left k" and "right k" represent the signals with unilateral loss whose energy between both channels was equalized by multiplying the channel with no hearing loss by constant k. It can be seen that, for the ILDs, there is a shift in offset when k is applied, so that they are now more or less located around 0 dB, where the ILD of the original signal C lies. Without k, the signals with unilateral loss sound louder at the side where there is no loss, which is illustrated by the ILDs. When they are equalized with k, however, sound is once again perceived at the original position in the back of the head, but with a similar spread-out quality that was noticed for this case in Analysis 2, and a noticeable "bouncing" effect — that is, the sensation that some parts of the signal, especially sibilants and other high-frequency sounds, sometimes quickly and subtly jump from one side of the head to the other. This bouncing effect was attributed to compression and the way that it modifies the dynamic range and the level of each frequency band in a different way. It also relates to the phenomenon observed in Analysis 1 where a sudden decrease in level can be perceived after a sudden peak, but this time it happens in a more complex and subtle way, given that each frequency band goes in and out of compression at different moments and with different compression ratios. For the the ITD and the IC, no relevant changes were observed.

The same analysis was repeated for the modified versions of signal C using audiogram S3, and the binaural cues are shown in Figure 4.21. Due to the characteristics of audiogram S3, the shape of the ILDs are more distorted and the audio tracks have a greater low-frequency emphasis or sound a bit more muffled when compared to those generated with audiogram N3. Nevertheless, the same effects concerning the position, bouncing and spreading of sound perceived for signals with audiogram N3 were noticed here.

For the other three signals, similar phenomena concerning the binaural cues, the perception of localization and the sound quality were noticed. In general, the signals generated with audiogram S3 have a more muffled sound quality than those generated with audiogram N3. As in Analysis 2, it was observed that signals with

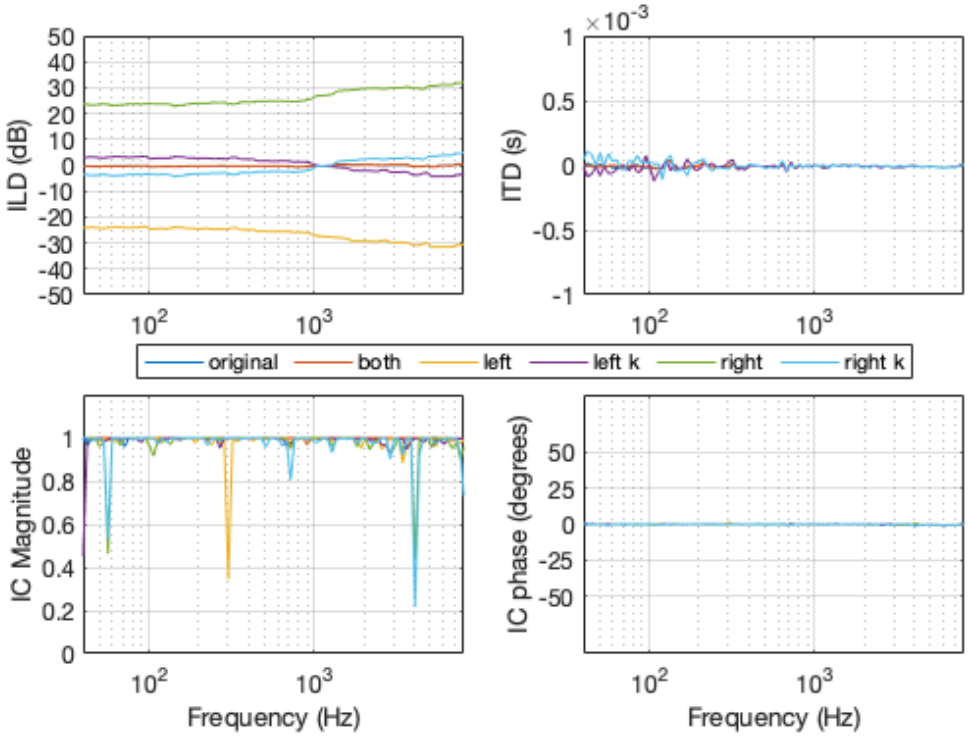


Figure 4.20: Binaural cues from signal C with N3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

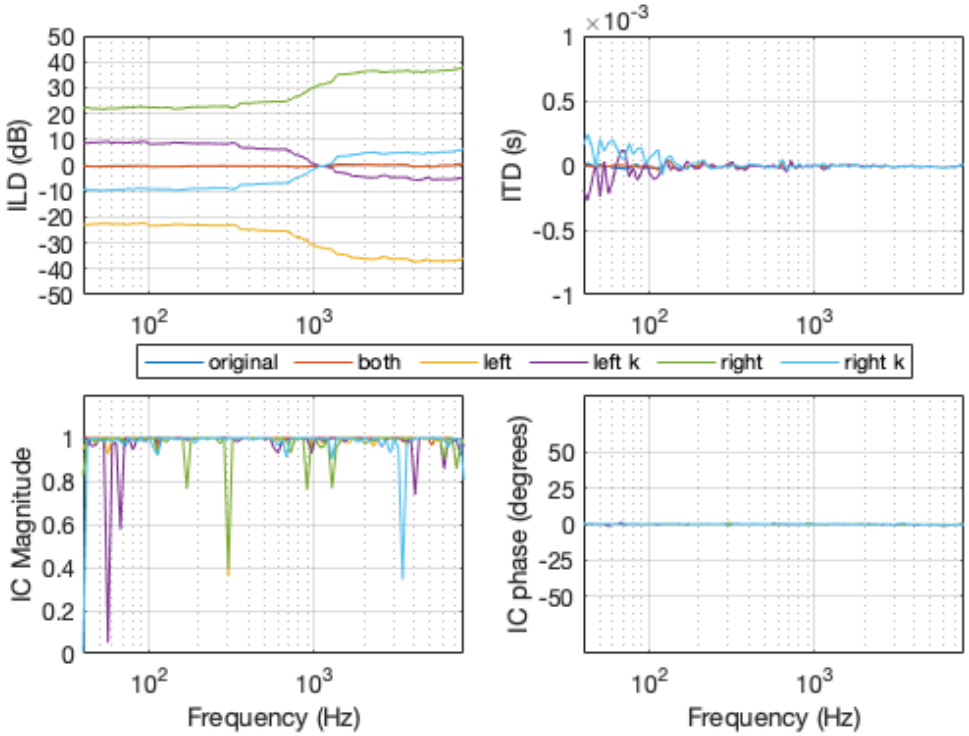


Figure 4.21: Binaural cues from signal C with S3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

unilateral loss are more impaired regarding the sensation of externalization when the loss is present at the ear closest to the source. Considering signals in the same conditions regarding audiogram and laterality of the hearing loss, those submitted to the non-linear fitting procedure suffer less change in sound quality compared to their original counterparts than those submitted to the linear fitting procedure. This is attributed to the fact that, since the linear fitting prescribes a gain that is smaller than the actual hearing loss in each band, there is a tendency to make the high-frequency region sound damped. With the non-linear fitting, however, even if a gain that is smaller than the actual hearing loss is also prescribed, compression makes the signal sound denser and possibly brighter (due to the generation of superior harmonics), thus partially undoing the perceived damping effect of linear amplification. It must also be noted that a later analysis of the gains prescribed by these two different fitting procedures was performed for some of the signals used in this chapter, and it was found that the non-linear fitting procedure does indeed tend to prescribe more gain in the low and high frequencies in comparison to the linear procedure. This is probably also directly related to the difference in sound quality between both cases, although it still cannot be generalized that the non-linear fitting procedure will present this behavior for all types of signals.

Figures 4.22 and 4.23 show the binaural cues for signal A and all its processed versions with audiograms N3 and S3, respectively. The binaural cues regarding signals B and D are shown in Appendix A.

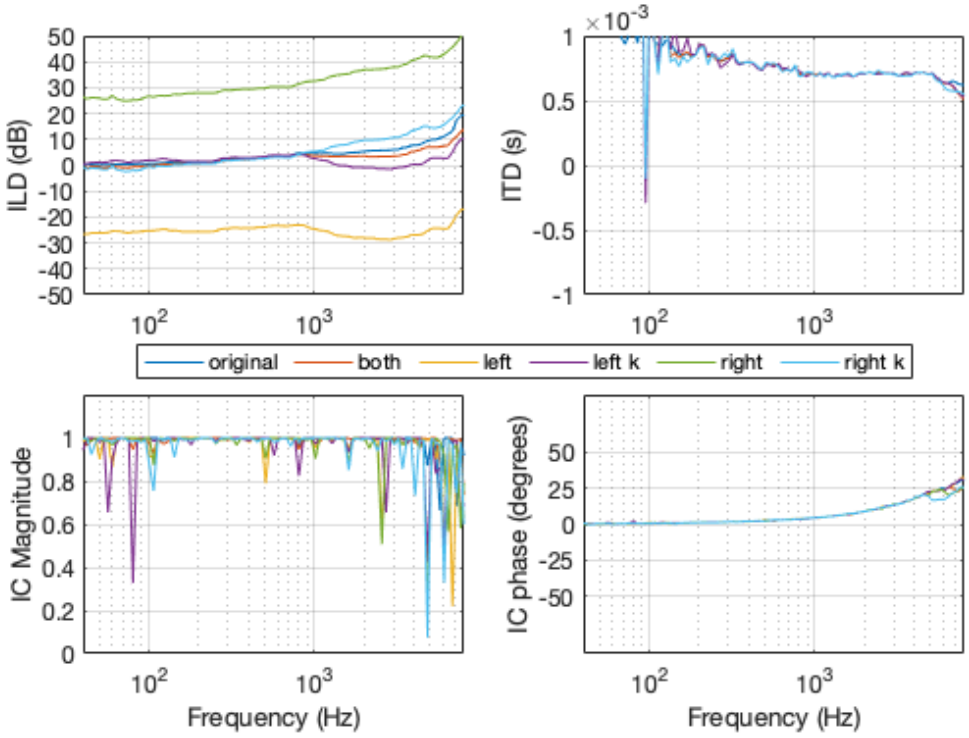


Figure 4.22: Binaural cues from signal A with N3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

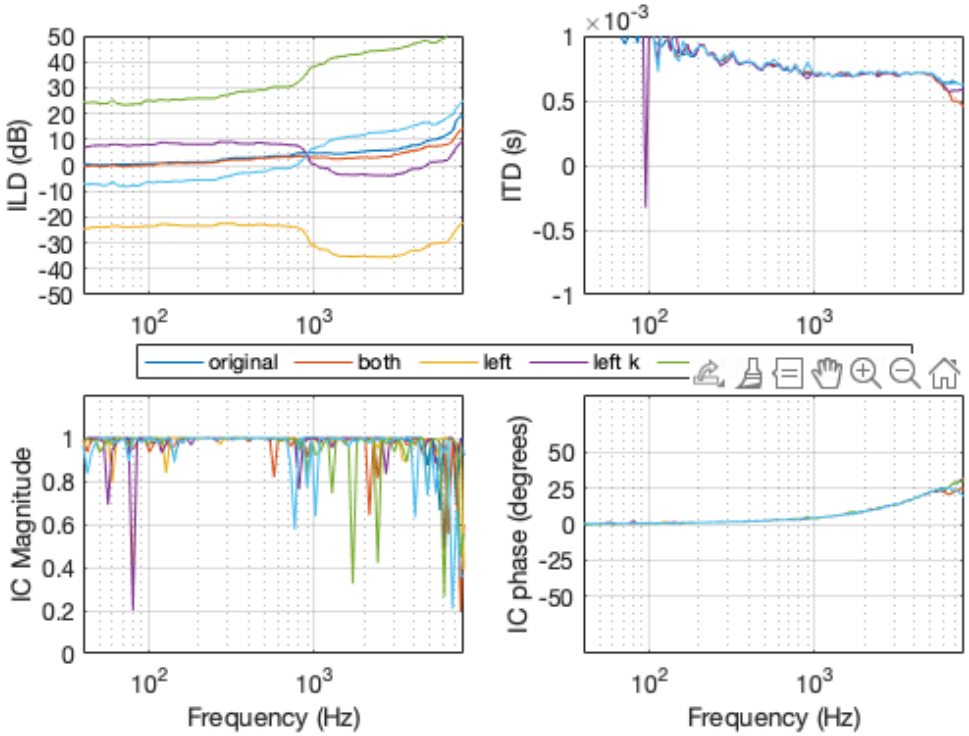


Figure 4.23: Binaural cues from signal A with S3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

4.2 Experiment 2

Experiment 2 was conducted by convolving the three original signals F13, F2 and M22 with several HRIRs from the reverberant environments Office II and Cafeteria. The floor plans of Office II and Cafeteria showing the room measurements, with the position of the HATS, where the listener is supposed to be, and the source positions for which the HRIRs were measured can be seen in Appendix A. For the analyses in this experiment, the chosen source positions for Office II were A, B and D, for head orientation 1, and the chosen source positions for Cafeteria were A, C and D, for head orientation 2.

The original anechoic signals were convolved with the respective HRIRs for each environment, and the spatialized signals were generated. The signals generated for Office II associated to source positions A, B and D and head orientation 1 are named as X-1A, X-1B and X-1D, respectively, with X being replaced by the name of the original anechoic signal. The signals generated for Cafeteria associated to source positions A, C and D and head orientation 2 are named X-2A, X-2C and X-2D, respectively, with X being replaced by the name of the original anechoic signal. The resampling of the signals, the Short-Time Fourier Transform and the calculation of the binaural cues were performed exactly the same way as for Experiment I.

When analyzing the binaural cues measured for the spatialized signals, three phenomena were noticed:

- Heavy distortions in the ITD;
- Magnitude of the IC being exactly 1 for the entire frequency range for signals generated from sentences F2 and F13;
- Magnitude of the IC with a large number of notches for signals generated from sentence M22.

The first effect could be justified by the many reflections occurring in the reverberant environment. The calculation of the ITD based on the ITF does not differentiate the direct sound from the reflections, and this is what leads to the distortions in the ITD. The difference in behavior concerning the magnitude of the IC for sentences F2 and F13 and for sentence M22 is also interesting. We can guess that the interaction of the spectral distributions of the female voices in F2 and F13 (with more high-frequency content) with the respective room frequency response was able to "fill" the notches previously found in the magnitude of the IC, but that effect was not observed for sentence M22. The phase of the IC, on the other hand, due to the effects of reveberation, presents a more distorted shape here than in Experiment 1, for all signals.

Figures 4.24 and 4.25 show as examples the binaural cues for signals F13-1A in Office II and F13-2A in Cafeteria, respectively, where the first two of the aforementioned effects can be seen. For Office II, the ITD has a more chaotic shape, whereas for Cafeteria, there is a very particular valley at around 1 kHz, possibly due to the characteristics of the environment and the measured HRIRs in this position. Figure 4.26, on the other hand, presents the binaural cues of signal M22-2A in the Cafeteria. The difference in behavior of the magnitude of the IC for signals F13-2A and M22-2A can be seen by comparing Figures 4.25 and Figure 4.26.

As far as the ILD goes, it was seen that it is underrated for all signals whose sound source is at position 1A in Office II. For the other positions in Office II, they also presented values near zero. This is understandable for position D, which has an azimuth very near zero, and less so for position B, which also has a small azimuth but should already be presenting greater ILD values. It is suspected that the underestimation of ILDs for positions A and B in this environment is associated to specific acoustic conditions at these positions in the recording environment. In the cafeteria, the ILDs of position A do not stray much from the estimated values for the position, and the ILDs for position D are around zero, as expected. Position C, however, which should be exactly at the back of the head according to the drawing in Appendix A, has an ILD which is slightly above zero, as can be seen in Figure 4.27 for sentence F13. This is an indication that position C, in real life, might not have been exactly at the back of the head, but leaning very slightly towards the left side, or that some reflections in cafeteria contributed to this apparent deviation. This behavior, as well as the behavior of the ITD for this position, which is also slightly above zero, will be discussed again in Chapter 5. For all positions in the office and in the cafeteria, it is important to keep in mind the probable influence of reverberation on the calculation of the binaural cues, which make them most likely present more deviations from the estimated values than the cues measured for the positions in the anechoic chamber.

The same three analyses done for Experiment 1 were repeated here, for the same conditions, and the results are detailed in the respective subsections where they are described. Most figures presented in the following subsections refer to signals generated from sentence F13 at its respective positions in the office and in the cafeteria. The binaural cues from the signals generated from sentence F13 in all the other analyzed positions in the office and in the cafeteria are presented in Appendix A.

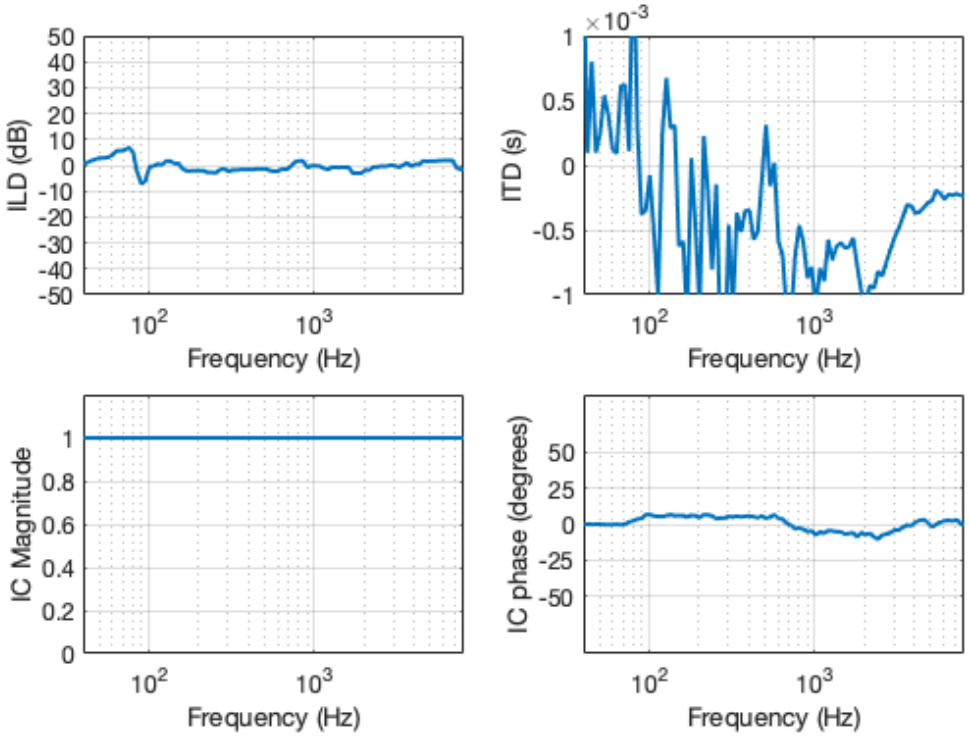


Figure 4.24: Binaural cues of signal F13-1A (Office II).

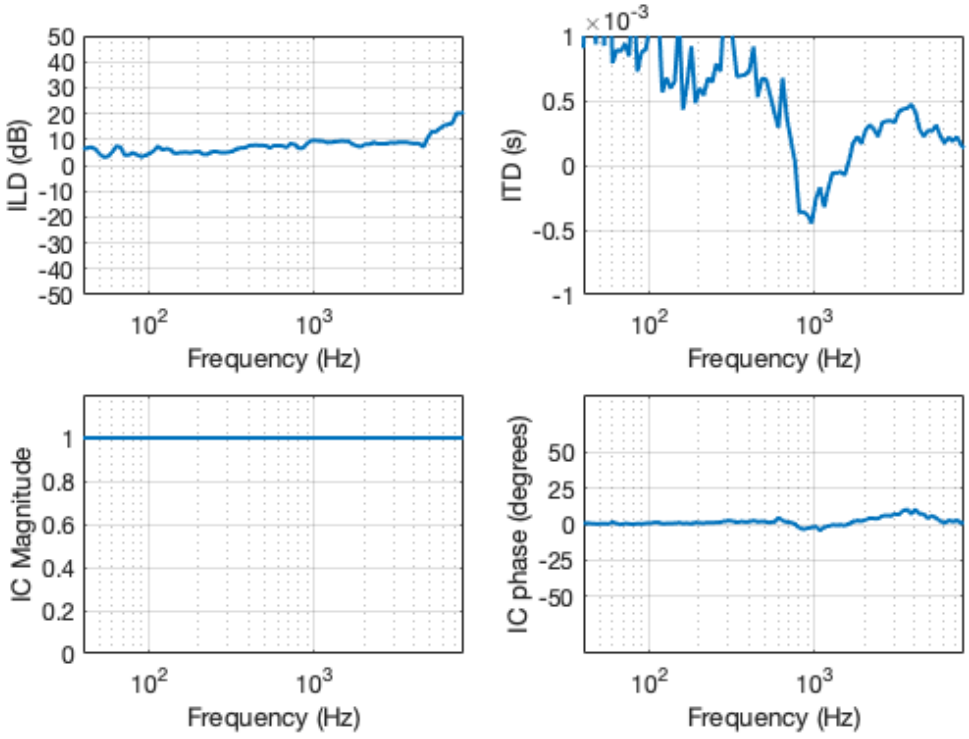


Figure 4.25: Binaural cues of signal F13-2A (Cafeteria).

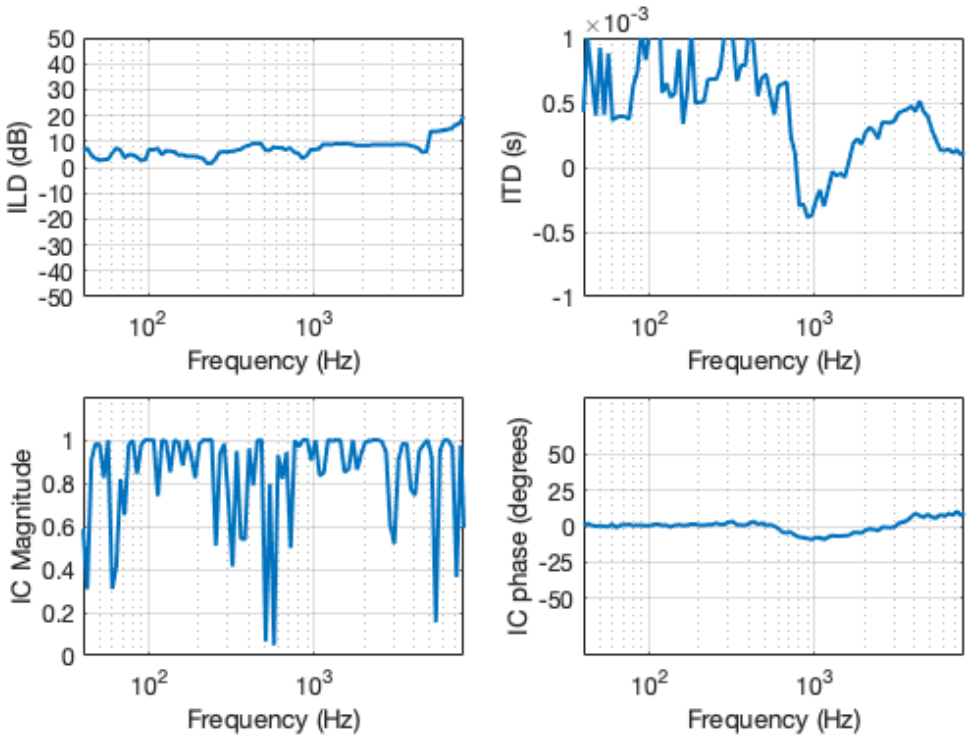


Figure 4.26: Binaural cues of signal M22-2A (Cafeteria).

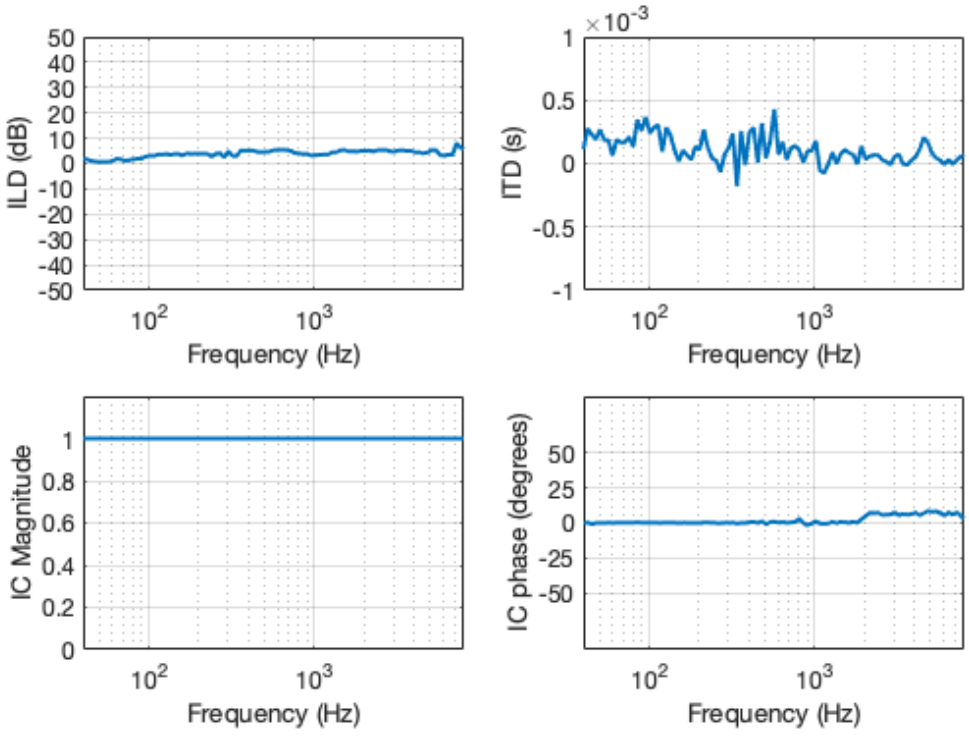


Figure 4.27: Binaural cues of signal F13-2C (Cafeteria).

4.2.1 Analysis 1

Analysis 1 aimed to verify the isolated effect of compression in the same way that it was done in Experiment 1. The spatialized signals were submitted to a single-channel compression system with CT = -100 dB, CR = 3:1, AT = 10 ms and RT = 200 ms.

Figure 4.28 shows the binaural cues for signal F13-1A in Office II and their versions with compression in both ears, compression solely in the left ear, and compression solely in the right ear. The ITDs are very distorted for all cases, more or less mimicking the ITD's shape of the unprocessed signal. It is not possible to establish a relation between the measured ITD and the estimated ITD for this position (around 110°azimuth), which should be around 650 ms and 600 ms for 500 Hz and 1 kHz, respectively, according to Figure 4.1. The ILD, on the other hand, oscillates around zero for all frequencies, which, as seen above, is not what is expected for this source position. Nevertheless, the effect of compression produces similar changes to what was seen for Analysis 1 in Experiment 1: compression in both ears virtually does not modify the ILD, while compression in one ear only makes the ILD change its offset towards the normal-hearing side.

Even though the measured ITD and ILD no longer match the expected values for them, the correct source position can be perceived pretty precisely in the original audio track for signal F13-1A. When compression is applied to both ears, although some changes in sound quality can be noticed, the source is still perceived at around the same position. When compression is in the left ear, the signal sounds as if it is more leaned towards 90°(right side), and when compression is applied in the right ear, most of the signal's content leans to the left hemisphere.

Figure 4.29 shows the binaural cues for signal F13-2A in Cafeteria and their version with compression in both ears, compression solely in the left ear, and compression solely in the right ear. While the ITD here is also heavily distorted for all cases, it can be seen that the ILD has a pattern that resembles the ILD found for signal A (in the position (0,-90°)), in Experiment 1. That is a positive thing, since the sound source here is also located at the left side of the head. When compression is applied to both ears, the ILD changes offset slightly, and when it is applied in one ear only, the offset changes significantly, and the ILD leans to the values corresponding to the side of the normal-hearing ear.

When listening to signal F13-2A with compression in both ears, while the signal appears to come mostly from the left side, a significant bouncing effect can be perceived, due to the effects of compression in both channels independently. Firstly, the direct sound is listened on the left side, and moments after the reflections are listened on the right side, in a less natural and less integrated way than in the

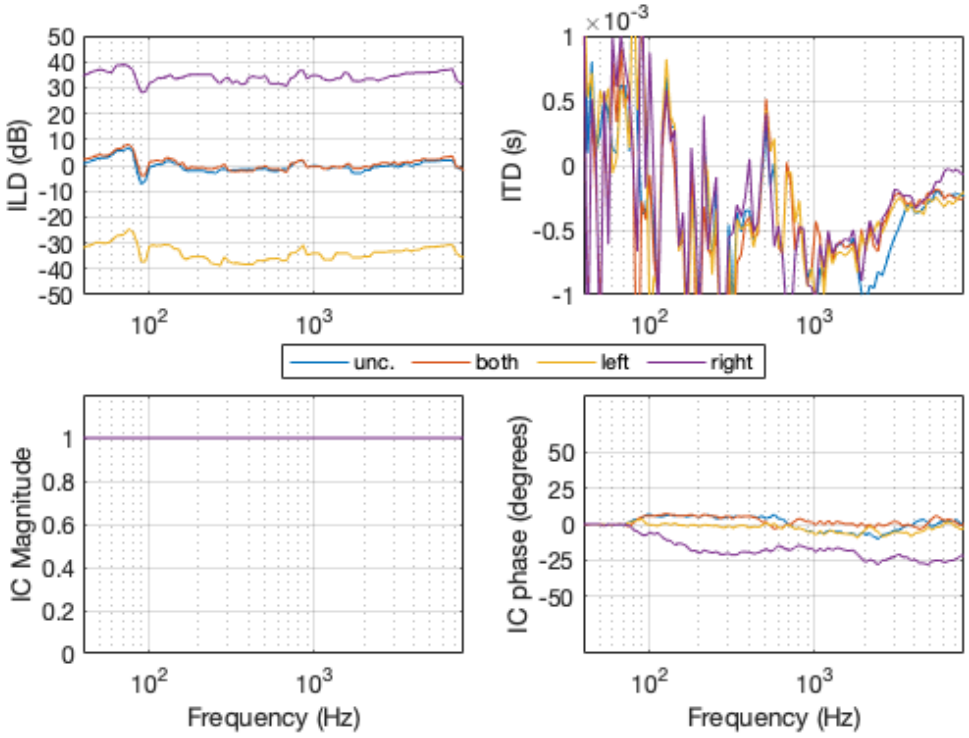


Figure 4.28: Signal F13-1A uncompressed and with compression applied at both ears, only at the left ear, and only at the right ear.

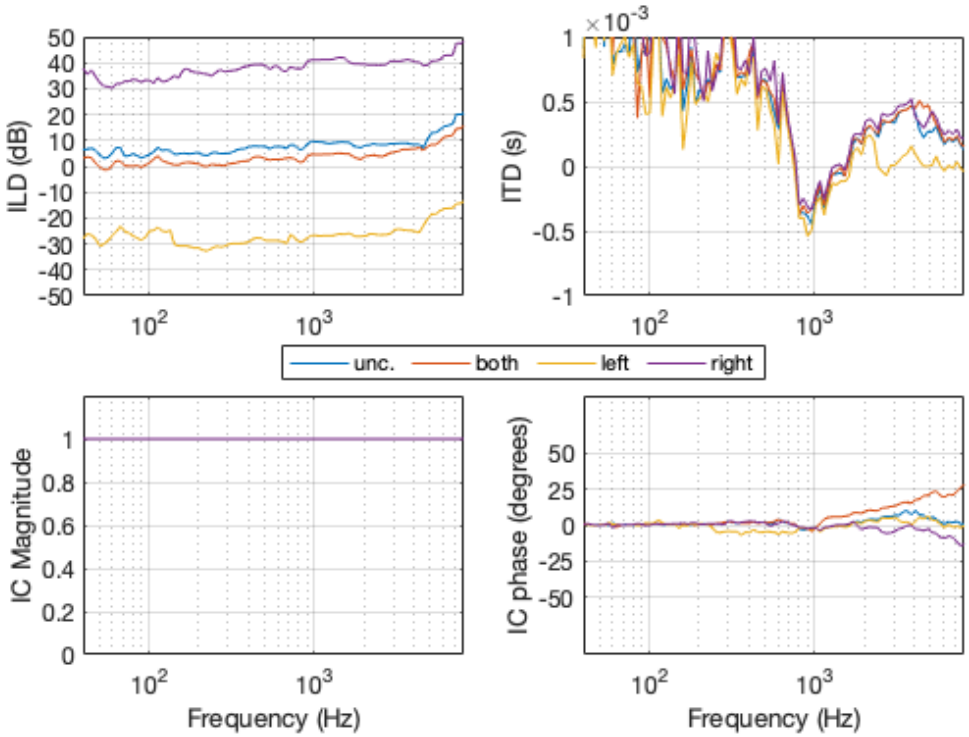


Figure 4.29: Signal F13-2A uncompressed and with compression applied at both ears, only at the left ear, and only at the right ear.

original spatialized signal. When compression is applied in the left ear only, the signal sounds stronger in the right side, and when it is applied in the left ear only, the signal sounds stronger and dryer in the left side, since compression damps a large part of the reflections, which come mostly from the right side.

4.2.2 Analysis 2

Analysis 2 was performed in the same way as in Experiment 1. The procedure NAL-R/NAL-RP was used, as well as Bisgaard's audiograms N3 and S3.

The signals were filtered by the same filter bank. After the prescription rules for the linear fitting were applied, a hearing loss compensation was performed and, lastly, an equalization between both channels with constants k was done for unilateral losses, in the same way as for Analysis 2 of Experiment 1.

Figures 4.30 and 4.31 show the binaural cues for signals F13-1A in Office II using audiograms N3 and S3, respectively. For both cases, the ITD is heavily distorted and cannot be associated to any value in particular. The ILD, which is already underestimated for position 1A, has distortions in shape when the FIG-6 gains are applied, for both audiograms.

However, when listening to the signals with bilateral hearing loss and unilateral hearing loss with equalization by k, they, or at least almost all of their content, sound as if they were still coming from the original source position. In general, all processed signals sound more damped than the original, especially those generated with audiogram S3 (which also have more low-frequency emphasis) and those with bilateral hearing loss.

Figures 4.32 and 4.33 show the binaural cues for signals F13-2A in Cafeteria using audiograms N3 and S3, respectively. The ITDs for both cases maintain the distorted shape of the original ITD with the valley around 1 kHz, and cannot be associated to any particular value. The ILD of the original signal, on the other hand, shows, for both cases, values across frequencies which are a bit overestimated, but are closer to the estimated ILDs for this source position (which is located at an azimuth of approximately -90°) than the measured ILDs of the original signal F13-1A in relation to the estimated ILDs for that position. Nevertheless, the ILDs of the processed signals in the cafeteria have a distorted shape, as those in the office.

In terms of perception, the processed versions of F13-2A with fitting applied in both ears and in the right ear (with equalization by k) still sound as if they are coming from the original source position. For the case of hearing loss in the left ear (with equalization by k), however, the sound is a bit more spread-out and imprecise, given that the source is already on the left side. Modifying the signal that arrives at the closest ear, as seen for other signals of Analysis 2 in Experiment 1, caused

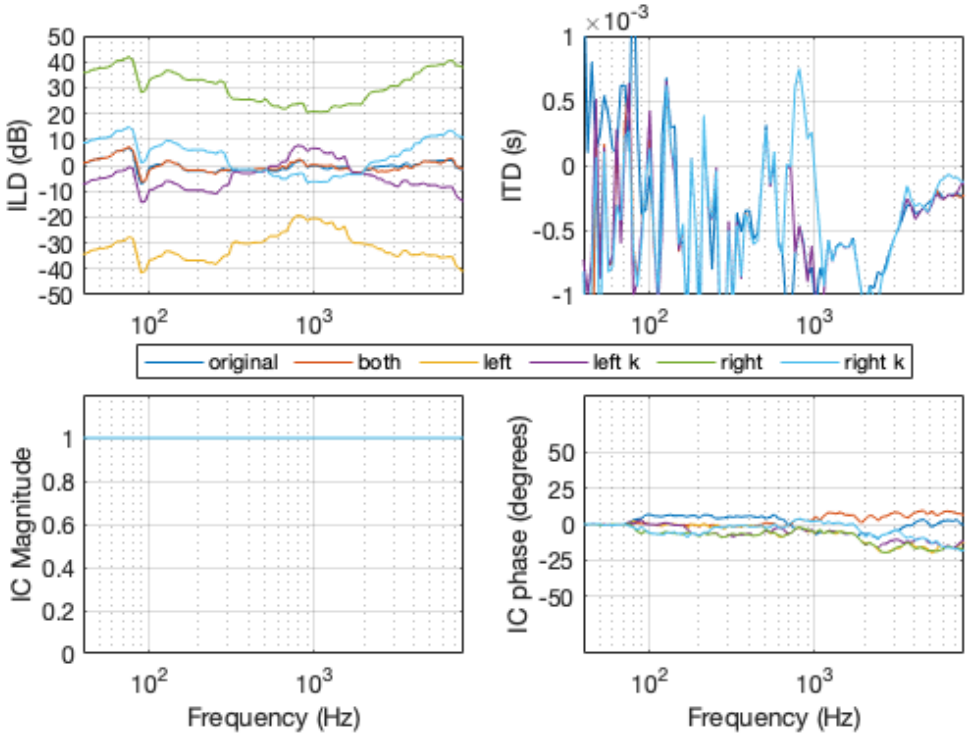


Figure 4.30: Binaural cues from signal F13-1A with N3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

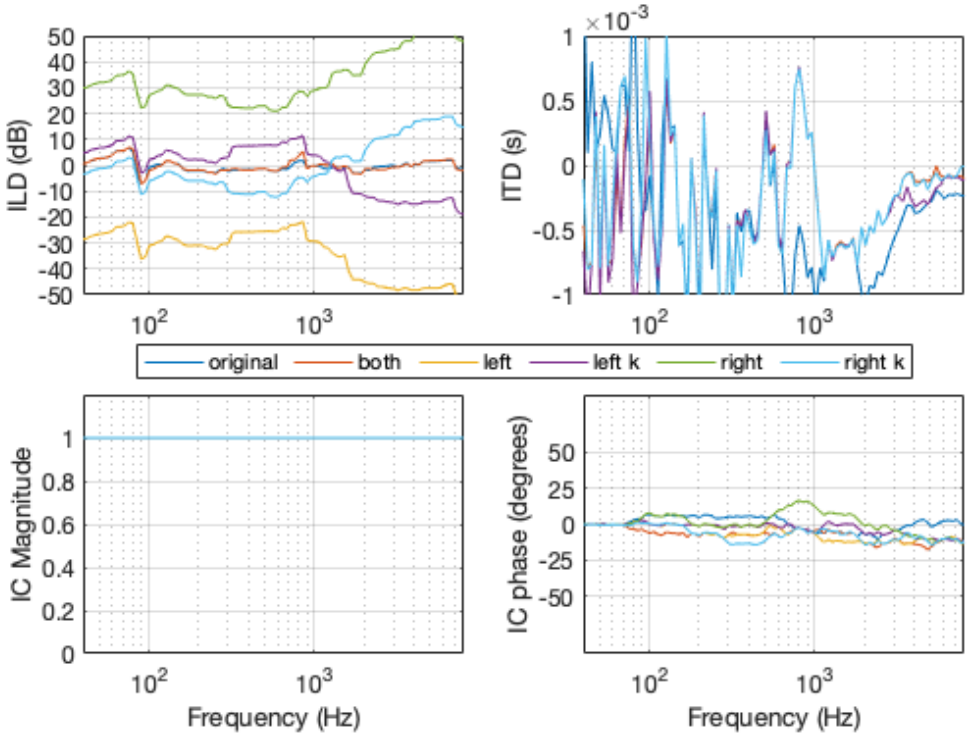


Figure 4.31: Binaural cues from signal F13-1A with S3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

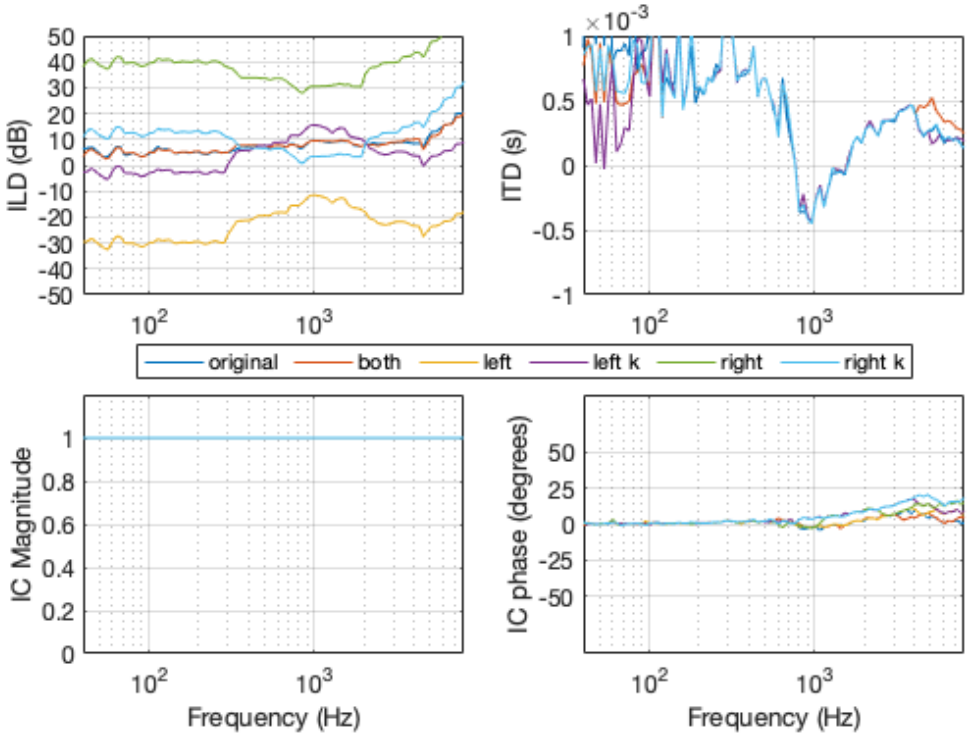


Figure 4.32: Binaural cues from signal F13-2A with N3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

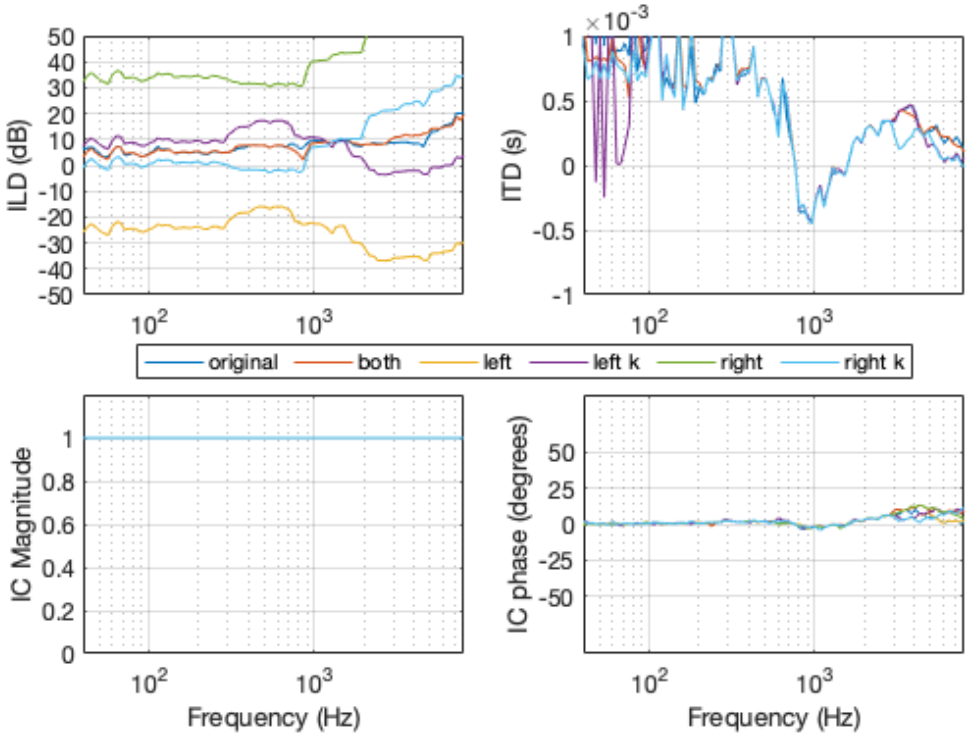


Figure 4.33: Binaural cues from signal F13-2A with S3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

more damage to the localization accuracy.

Similarly as the processed version of signal F13-1A, the processed versions of signal F13-2A present a more damped high-frequency range than the original, especially when there is bilateral hearing loss. The signals generated with audiogram S3 have an even more attenuated high-frequency range, but more emphasis in the low frequencies than those generated with audiogram N3.

4.2.3 Analysis 3

Analysis 3 of Experiment 2, identically to Analysis 3 of Experiment 1, aims to simulate the effect of a non-linear prescription procedure. The FIG-6 procedure was also used here and all steps of the implementation were done in the same way of Analysis 3 of Experiment 1. An artificial gain was also applied to the unprocessed signals, so that their peak levels go up to around 80 dB.

Figures 4.34 and 4.35 show the binaural cues for all signals generated from F13-1A in Office II with audiograms N3 and S3, respectively. It can be seen the ITD remain very distorted for both cases, although the ILDs, particularly audiogram N3, are less distorted than those of their counterparts in Analysis 2, shown in Figure 4.30 and 4.31.

Figures 4.36 and 4.37 show the binaural cues from signal F13-2A and its variations using audiograms N3 and S3, respectively. The ITDs of the processed signals maintain the shape of the original distorted ITD. The ILDs, similarly to what was seen for signal F13-1A, look a bit less distorted in shape and more similar to the ILD of the original signals than those found for the signals generated in Analysis 2, shown in Figure 4.32 and 4.33. This phenomenon is due to the greater gain applied to low and high frequencies by the non-linear fitting procedure, in comparison to linear amplification. When the hearing loss is subtracted after the gain prescription and the resulting gain is calculated, a flatter "resulting" audiogram is obtained when non-linear amplification is used.

By listening to the audio tracks, one can perceive all processed signals to have approximately the original source position. When there is hearing loss at the left ear, however, a small "leakage" of part of the signal to the right side can be noticed. This is consistent with the finding that hearing loss and fitting applied to the closer ear causes more damage to localization perception. In terms of sound quality, the signals generated with audiogram S3 sound more damped, as expected. In comparison to the signals of Analysis 2, the signals here in Analysis 3 sound brighter and also have more low-frequency emphasis, which makes they have a more similar sound quality to the original ones.

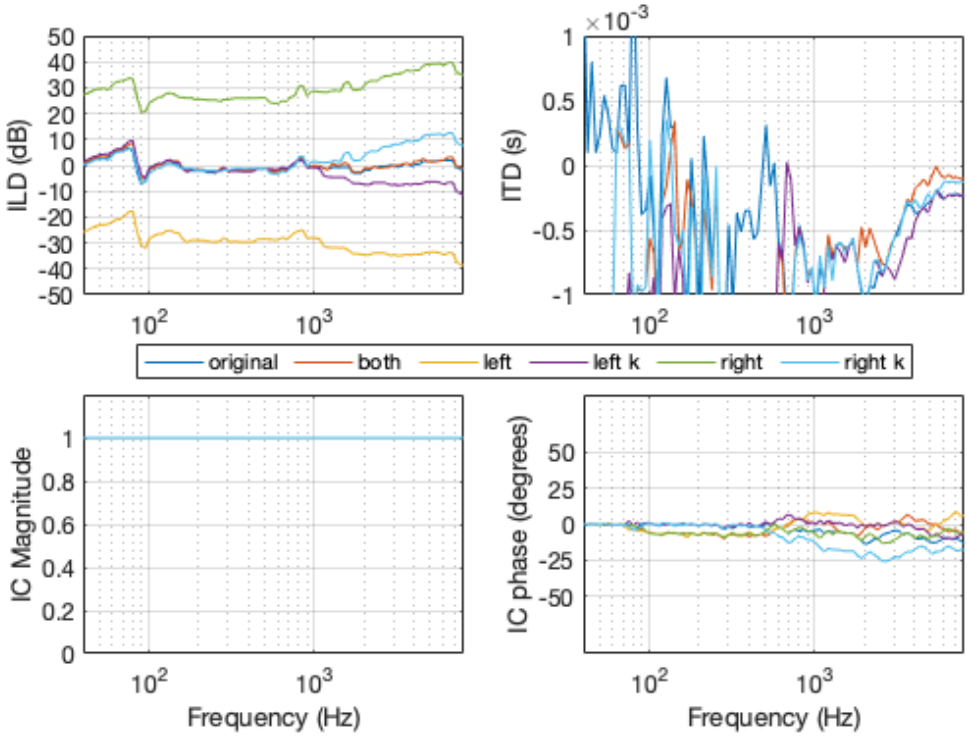


Figure 4.34: Binaural cues from signal F13-1A with N3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

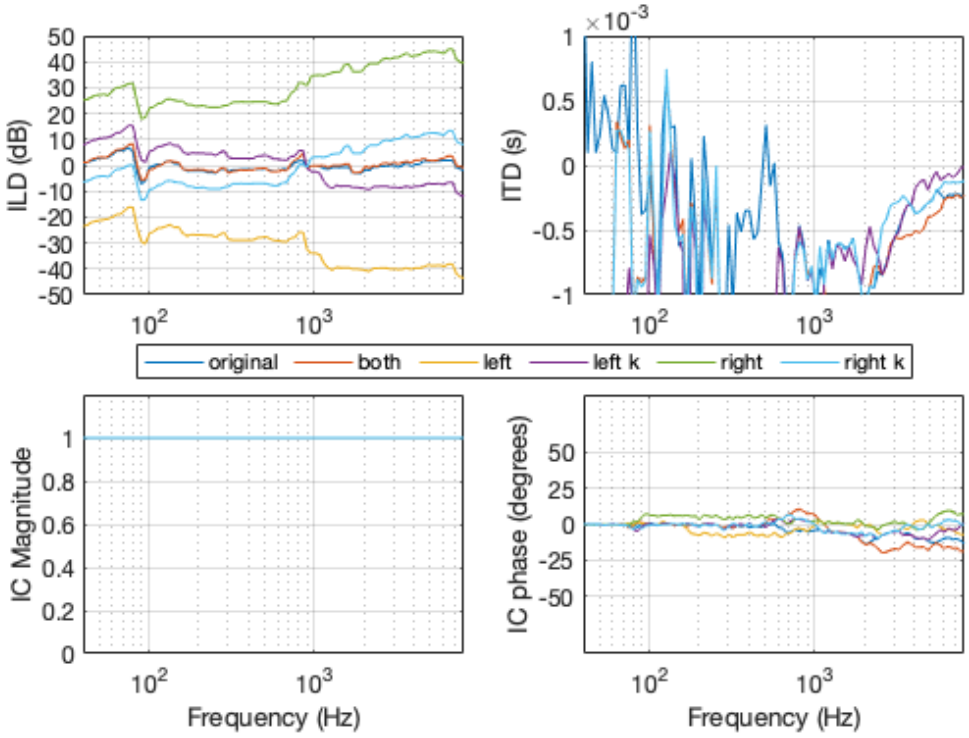


Figure 4.35: Binaural cues from signal F13-1A with S3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

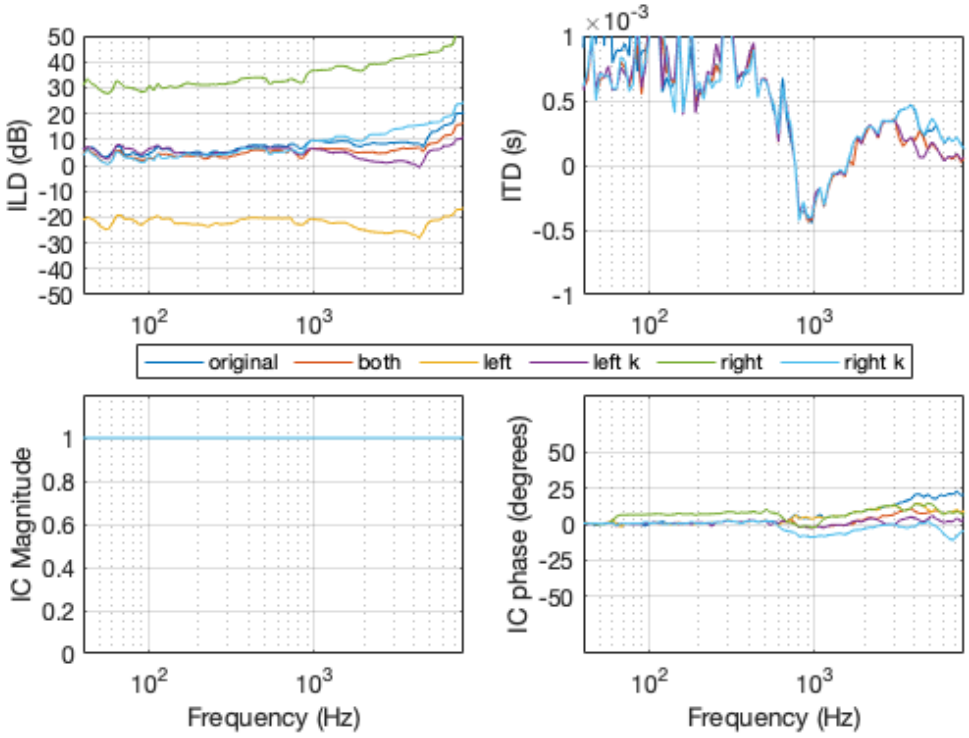


Figure 4.36: Binaural cues from signal F13-2A with N3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

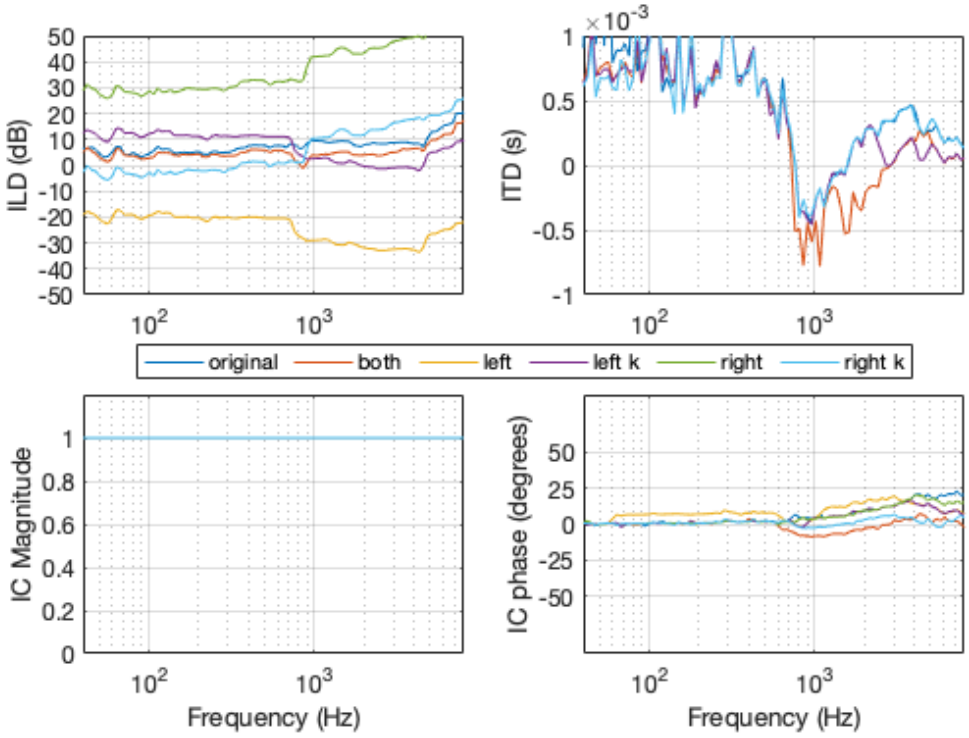


Figure 4.37: Binaural cues from signal F13-2A with S3 audiogram on both ears, left ear and right ear. Cues for unilateral losses are plotted with and without equalization by constant k.

4.3 Summarized conclusions

After a detailed observation of the results and careful listening of the generated audio tracks for all analyses in both experiments performed with signals A, B, C and D, the general conclusions are, in summary:

- Overall, the perception regarding sound localization for Analysis 1 when unilateral compression is applied cannot be compaired to that for signals with unilateral hearing loss generated for Analyses 2 and 3, since the strategy of equalizing sound power in both channels performed for Analyses 2 and 3 was not performed in Analysis 1. This equalization, however, was performed for signals generated with the Analysis 1 method for the subjective tests, described in the following chapter.

- Signals generated with the S3 audiogram (steep sloping moderate-to-severe hearing loss) sounded always much more damped and boomy than those with the N3 audiogram (moderately sloping moderate hearing loss). This is understandable, given that a person with an S3-like hearing loss would lose most of the high-frequency content of the signal, which is not fully compensated by the prescription procedures.

- Signals generated with the Analysis 3 method have both low and high frequencies emphasized, and have a more realistic sound quality than those generated for Analysis 2. This can be attributed to the compression applied by the fitting procedure used in Analysis 3, since the fitting procedure of Analysis 2 prescribes linear amplification only. Compression shrinks the dynamic range, making the signal sound denser, and it may also introduce superior harmonics that give the sensation of a brighter sound. As mentioned earlier, a later analysis of the gains prescribed by these two different fitting procedures showed a tendency of the non-linear procedure of indeed prescribing more gain to low and high frequencies than the linear procedure, at least for the signals in this study, which is also a factor that contributes to this difference in sound quality;

- No generalizations could be made as to which fitting procedure (linear amplification in Analysis 2 *versus* non-linear amplification in Analysis 3) produces the greatest changes in the spatial perception. This proved to be very signal dependent;

- In general, it was observed that there is more damage to the spatial perception and the sensation of externalization when the hearing loss and the fitting procedure are applied to the ear which is closer to the sound source.

Chapter 5

Subjective tests

After the theoretical analyses were performed for the anechoic and reverberant signals, subjective tests were applied to better evaluate the impact of the processing techniques on spatial perception. Two subjective tests (Test One and Test Two) were created: the first one aimed to evaluate the impact of compression alone (similarly to what was performed in Analysis 1), and the second one aimed to evaluate the impact of the non-linear fitting procedure (as performed in Analysis 3). Since there was not much difference in the changes on binaural cues in Analyses 2 and 3, and since signals generated for Analyses 2 and 3 seemed to present similar phenomena regarding spatial perception, the tests did not include linearly amplified signals as done in Analysis 2. Tests One and Two were done by the same group of 23 normal-hearing individuals, with ages ranging from 23 to 60 years.

Both subjective tests were implemented using an online form and were run on the same computer for all participants, in an acoustically treated room in the SMT laboratory. All participants wore the same pair of in-ear headphones (but different eartips) . The first test presented 33 signals, while the second presented 47 signals. After having heard each signal, which could be replayed at will, participants were asked two questions:

- Which horizontal direction (azimuth) does the sound seem to be coming from?
- How diffuse does the signal sound?

For the first question, participants had to choose an answer from the options presenting azimuths ranging from $-150°$ to $180°$ in steps of $30°$ (negative azimuths were attributed to the left hemisphere in relation to the head). Right after the question, a picture showing the head orientation and the azimuths was available to help visualization. The picture can be seen in Figure 5.1. Even if some of the signals have nonzero elevation angles, it was chosen not to evaluate elevation on the subjective tests, but rather focus on the azimuth alone. This decision made it possible to keep

the duration of the subjective tests within reasonable limits. Apart from that, since binaural cues are related to azimuth rather than elevation, isolating the effect of the processing strategies on azimuth was a way of maintaining the analyses focused on the binaural cues alone, rather than other factors such as monaural cues and other spectral colorations, which are beyond the scope of this work.

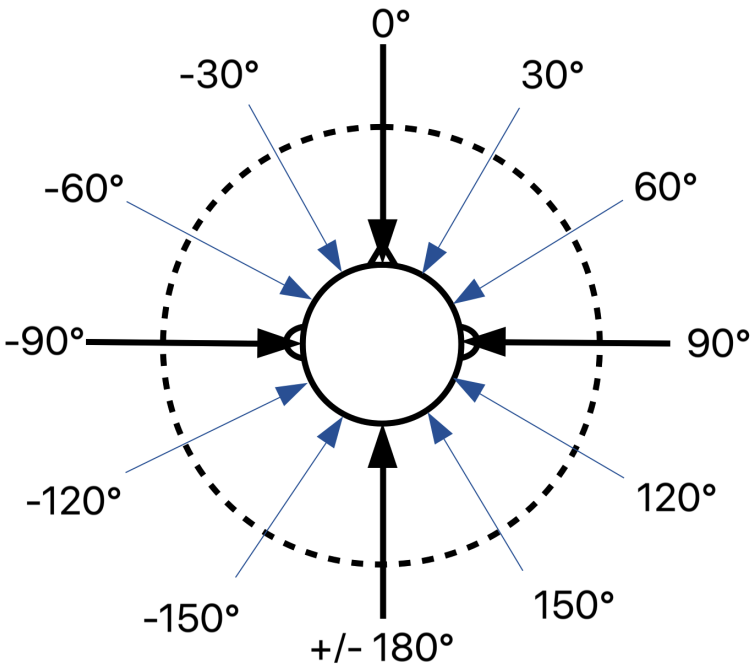


Figure 5.1: Picture showed in the listening test form presenting the head orientation and possible azimuths for the sound source location.

The second question aimed to extract from the participants their impression of the spreading of sound. It was also a multiple choice question, and listeners had to choose between three options: “Precise”, “More or less”, and “Diffuse”. Only for some signals (2 signals in Test One and 4 signals in Test Two), a third question was asked:

- Does the sound seem to be coming from more than one direction?

For this question, participants had to choose between “Yes” or “No”. These questions were added for signals which seemed to give a more perceptible impression of the “bouncing” phenomenon found for some signals analyzed in Chapter 4, and it was intended to verify if this bouncy quality was indeed perceived by the listeners.

The sound volume had to be calibrated so that the signals could be reproduced at an approximate typical sound level for speech signals (around 60-70 dB, with peaks at around 80 dB). All signals from both tests were first equalized so that they all presented the same approximate average sound power. After that, there was originally an attempt to adjust the signal reproduction and the computer volume

controls in order to make a 0 dB 2-kHz sinusoid (in the digital world) to be perceived at the threshold of hearing, so that the sound pressure level with which the speech signals would be reproduced would approximately match their average sound levels in the digital world (around 60-70 dB). Nevertheless, this calibration did not work very well with the computer that was being used, since the reference signal could not be heard at a reasonable reproduction volume due to performance limitations of the sound card. In the end, the signal reproduction and the computer volume controls were adjusted so that the signals were reproduced at a comfortable listening level, and the volume adjustments were kept the same for all signals.

Before the test started, some personal information about the participants were collected, such as age, gender and date of birth. They were also asked if they were musicians (either amateur or professional), professionals/researchers in the fields of Audio and Acoustics, or healthcare professionals in the area of hearing, since these occupations could imply a greater spatial hearing experience. After their personal information was collected, participants had to read the test instructions, which better explained how they could answer the questions and the meaning of some terms (such as "spreading", "precise", "diffuse"). Participants were also asked not to change the computer audio volume or the signal reproduction volume, which should be kept the same for the realization of all tests. The tests' durations ranged from 15 to 25 minutes, approximately.

The results are presented and analyzed in the following sections, for each test separately.

5.1 Test One

Test One aimed to evaluate the effect of compression alone on the binaural cues and on the localization ability. For this test, the original sentences F2 and F13 from the SMT database were used. Spatialized versions of these two sentences were generated for three positions in the anechoic chamber ((0°,-90°), (-10°,180°) and (20°,60°)) and for three positions in the cafeteria (A, C and D). The placement of these positions within their respective environments can be seen in Appendix A. Bilaterally compressed versions of these signals were then generated, by using compression thresholds 40 dB (for a more aggressive compression) and 60 dB (for a compression of mostly the louder parts of signal). The other compression parameters were kept fixed (CR = 3:1, AT = 10 ms and RT = 200 ms). Therefore, the test contained 33 signals in total:

- 3 uncompressed spatialized signals of sentence F2 and 3 uncompressed spatialized signals of sentence F13, as references (sentences chosen randomly for each

one of the three positions in the anechoic chamber and the three positions in the cafeteria);

- 12 signals with compression at both ears by using a 40 dB compression threshold (2 sentences × 2 environments × 3 positions in each environment);
- 12 signals with compression at both ears by using a 60 dB compression threshold (2 sentences × 2 environments × 3 positions in each environment);
- The remaining 3 uncompressed signals presented as the first three signals, for confirmation.

The signals were named Track 1–33 and presented in the same order for all participants, since the interface chosen to guarantee a friendly and easy way to realize the tests did not enable the randomization of sections. Since the bouncing effect provoked by compression seemed to be most noticeable for some of the compressed signals with $CT = 40$ dB, the signals for which the extra question was asked were Track 8 (F13 at position (20°,60°) in the anechoic chamber and $CT = 40$ dB) and Track 30 (F13 at position (0°,90°) in the anechoic chamber with $CT = 40$ dB).

5.1.1 Evaluation of azimuth

In order to enable statistical analyses, the chosen measurement to evaluate the answers concerning the azimuth was the squared angular error — that is, the squared difference between the perceived azimuth and the correct azimuth. When the squared angular error had to be averaged across participants or across signals and be represented by a single value, the RMS angular error $\overline{\varepsilon}$ was used instead. It is obtained according to Equation 5.1:

$$\overline{\varepsilon} = \sqrt{\sum_{i=1}^{N} \frac{(\theta_i' - \theta_i)^2}{N}} \tag{5.1}$$

where θ_i is the correct azimuth, θ_i' is the perceived azimuth and N is the number of participants or signals to be analyzed. The reason for the square root is to convert the values from degrees^2 back to degrees, in order to allow a better interpretation.

For all the statistical analyses, errors due to front-back confusion, regardless of azimuth, were not considered. That is, all azimuths in the rear hemisphere are mirrored to the frontal hemisphere before the angular error is calculated. That means that an azimuth which is perceived in a symmetrical position to the correct azimuth in relation to the interaural line (i.e. an azimuth of 60° perceived as 120°) is also considered correct and generates an angular error of 0. Following this logic, an azimuth of 180° which is perceived as 30° will produce an absolute angular error

of 30° only. It is worth noting that, regardless of the compression to which the spatialized signals are submitted, the correct azimuth is always considered to be the azimuth of the unprocessed spatialized signal (which is the position where the source is still supposed to be).

Figure 5.2 shows the RMS angular error for each signal, a result of the squared angular error averaged across participants for the respective signal. After verifying which signals presented the largest errors, it was seen that the following factors contributed for an increase in the RMS angular error: reverberant environment, sound coming from the back (especially in the reverberant environment), and compression. These are analyzed in more detail with the figures that follow.

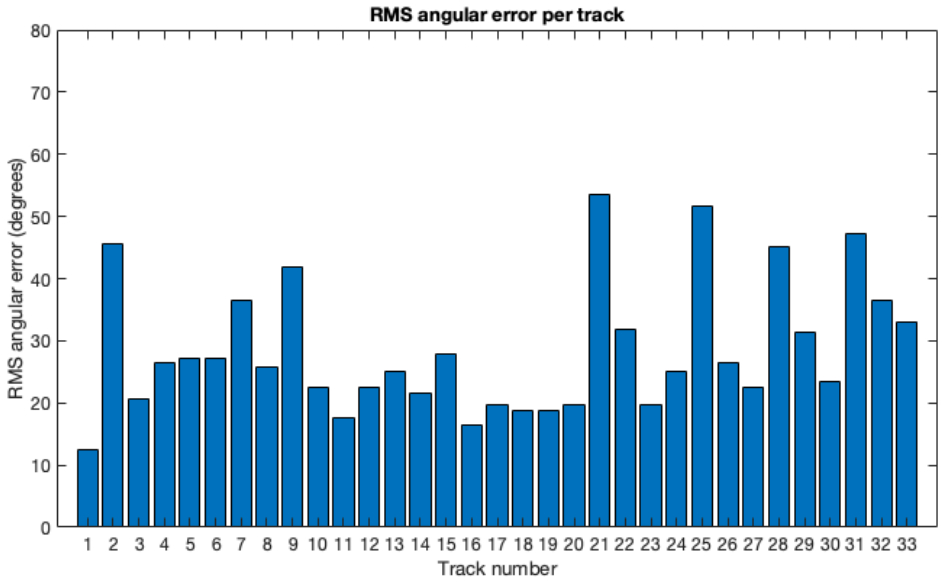


Figure 5.2: RMS angular error for each signal.

The angular error was analyzed in relation to position, environment, compression threshold, talker, and listener's experience. The Friedman Test [54, 55], which is a non-parametric alternative of the ANOVA, was used in this work to evaluate statistical significance between the difference in the means of three or more populations, that represent three or more levels of an independent variable. It can be used on a continuous or ordinal domain. When only two populations are evaluated, the Sign Test For Paired Data [55], which is also non-parametric and verifies the hypothesis that the difference between two populations generate a distribution with zero median, was used instead. With the Friedman Test, a further multiple comparison test following the Bonferroni method [56] was performed to verify the relation between each pair of populations separately. For both the Friedman and the Sign Test, the populations are represented by the mean square angular errors of each participant (or each signal, for the analysis of the influence of listener's experience). The RMS

angular error of each population is, therefore, the square root of the mean of those values.

Firstly, an analysis with the isolated effect of position in each environment was performed. Table 5.1 shows the mean square angular error of each participant for each position in the anechoic chamber, while Table 5.2 presents the same data for the cafeteria. For the cafeteria, positions A, D and C are replaced by the values of their approximate azimuths ($-90°$, $0°$ and $180°$, respectively). The RMS angular errors for each position are also shown in both tables. While there was not a very large difference between the RMS errors for each position in the anechoic chamber, a clearly larger RMS error can be seen for sounds coming from the back in the cafeteria. This can be explained due to a probable decrease in localization precision for sounds coming from the rear hemisphere in comparison to sounds coming from the frontal hemisphere, as observed in Carlile [57], Makous [58] and Mironovs [59]. The fact that the error at the left side is greater than the one for the frontal position may also be an indicator of the greater sensibility for azimuths near the frontal midline than for azimuths near the sides of the head, a phenomenon which is observed in the same three articles. For the anechoic chamber, a slightly larger error for the azimuth -90°might reflect the same decrease in sensibility near the sides of the head, and also be an evidence that an alteration of perceived source position caused by compression might be more noticeable when the source is near the side of the head, which can also explain the reasonably greater error at the left side in the cafeteria, when compared to the front.

According to the Friedman Test, the differences between the mean square angular errors, averaged across participants, of the three positions in the anechoic chamber are not statistically significant ($p > 0.05$). For the cafeteria, on the other hand, it was seen that the position at the back of the listener has a significantly greater RMS angular error than the other two positions ($p < 0.05$). For all other analyses that follow that use as populations the mean square angular errors of each participant (that is, all but the analysis of the influence of listener's experience, which uses the mean square angular errors for each signal), the table will be omitted for the sake of text conciseness and better visualization, and the bar plots illustrating the RMS angular errors of each population will be shown instead, in the same way that is shown in Figures 5.3 and 5.4, for the positions in the anechoic chamber and in the cafeteria, respectively.

Table 5.1: Mean square angular errors of each participant for each position in the anechoic chamber.

	Mean square angular errors (degrees^2)		
Participant	-90°	60°	180°
1	0	540	720
2	1800	360	360
3	180	540	1440
4	360	720	1080
5	180	540	0
6	540	720	1800
7	0	720	360
8	180	720	180
9	1080	720	180
10	360	360	360
11	360	360	540
12	900	540	180
13	0	540	360
14	3600	540	900
15	360	540	0
16	720	360	2160
17	1800	360	0
18	180	360	180
19	0	360	0
20	0	900	0
21	540	720	900
22	180	900	1800
23	4860	180	360
RMS angular error	28.11°	23.41°	24.55°

Table 5.2: Mean square angular errors of each participant for each position in the cafeteria.

	Mean square angular errors (degrees^2)		
Participant	-90°	0°	180°
1	360	0	2520
2	1260	180	1620
3	6840	1260	2520
4	360	540	1620
5	540	540	1800
6	1260	360	1440
7	720	540	720
8	360	180	2160
9	360	180	540
10	720	540	3960
11	540	720	1440
12	1800	0	360
13	540	0	1440
14	2340	540	540
15	360	0	1260
16	900	1800	2520
17	1980	2340	2160
18	360	360	1080
19	0	0	2700
20	180	720	3600
21	540	180	3600
22	1800	3240	4140
23	360	720	540
RMS angular error	32.62°	25.49°	43.88°

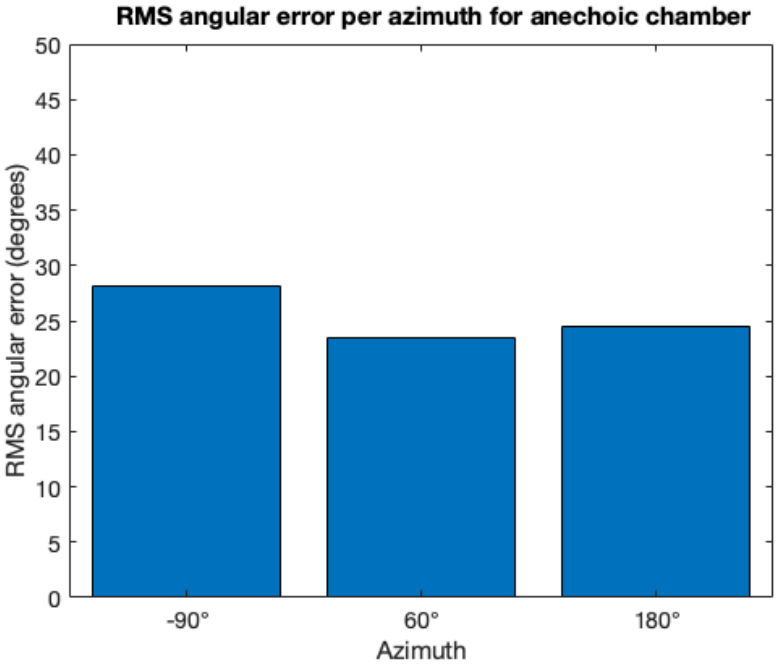


Figure 5.3: RMS angular error per azimuth in the anechoic chamber.

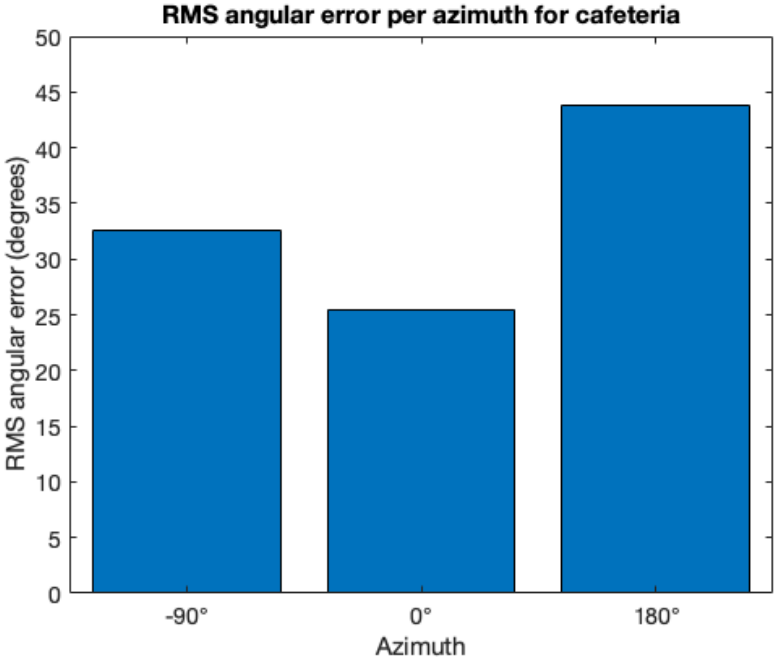


Figure 5.4: RMS angular error per azimuth in the cafeteria.

Secondly, an analysis of the influence of the environment was also performed. The Sign Test rejected the null hypothesis that the difference between the populations representing anechoic chamber and cafeteria generate a distribution with zero median ($p < 0.05$), which in this case means that the difference between the medians of the mean square angular errors in the two environments is statistically significant. That reflects the degradation in localization ability in an environment with a large reverberation time. The reflections occuring in nearer obstacles such as chairs and tables might also contribute to blur the sound source's location. Figure 5.5 shows the RMS angular error for each environment.

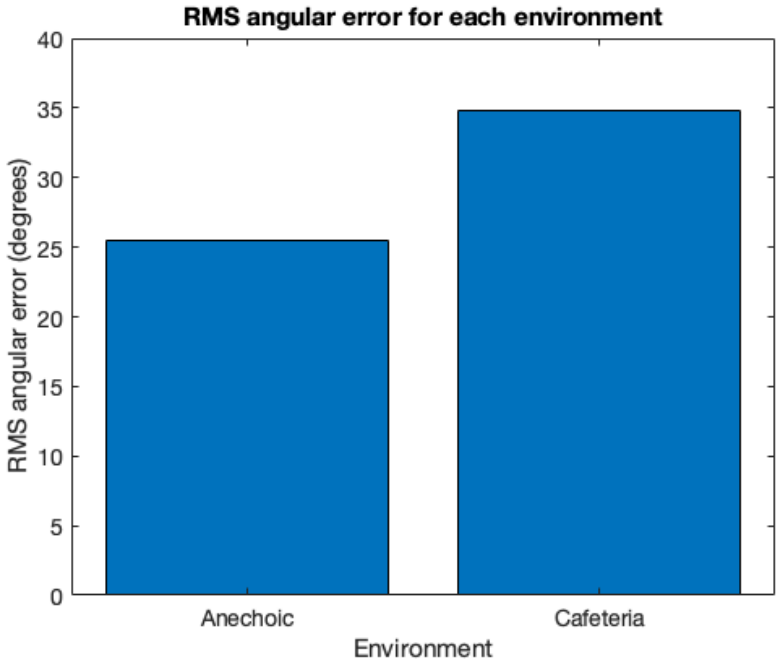


Figure 5.5: RMS angular error for each environment.

There was no significant difference between the mean square angular errors associated to each compression threshold (no compression, $CT = 40$ dB and $CT = 60$ dB, respectively), according to the Friedman Test ($p > 0.05$). In Figure 5.6, however, a reasonably larger RMS angular error for the signals with no compression can still be seen. This can be explained by the fact that compression makes the source location seem more blurred (as was observed in the evaluations regarding the spreading of sound, described in the following section). When this happens, the listener might tend to choose a position around the "middle" of the region where the source is supposed to be (which often might end up being the right one), while, for an uncompressed signal, a very precise spot might be harder to be correctly located by an insecure listener. Another possible reason for the lower RMS errors of the compressed signals is that compression squeezes the dynamic range of the signal, making it sound denser and, given that the sound powers of all signals are leveled

out, the compressed signals might be therefore perceived as louder, which can help localization.

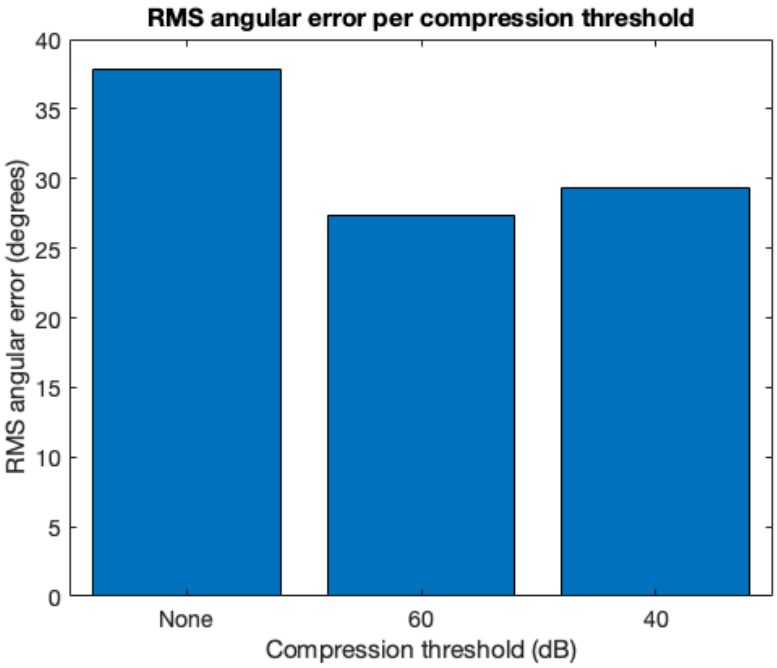


Figure 5.6: RMS angular error per compression threshold.

Regarding the influence of the talker on angular error, the Sign Test could not reject the null hypothesis that the difference between the mean square angular errors generate a distribution with zero median ($p > 0.05$). Although a couple of participants verbally expressed that the voice F1 (from sentence F13) sounds a bit louder and more precise than the voice F2 (from sentence F2), this generated no significant difference in the overall results. This can be considered a good thing, since both voices were recorded in the same conditions and since it shows that localization performance was more affected by the factors which are indeed evaluated in this study. Another good thing is that, since each talker speaks a different sentence, this result also suggests, by extrapolation, that the sentence itself did not have a significant influence on the angular error either. The RMS angular errors associated to both talkers are shown in Figure 5.7.

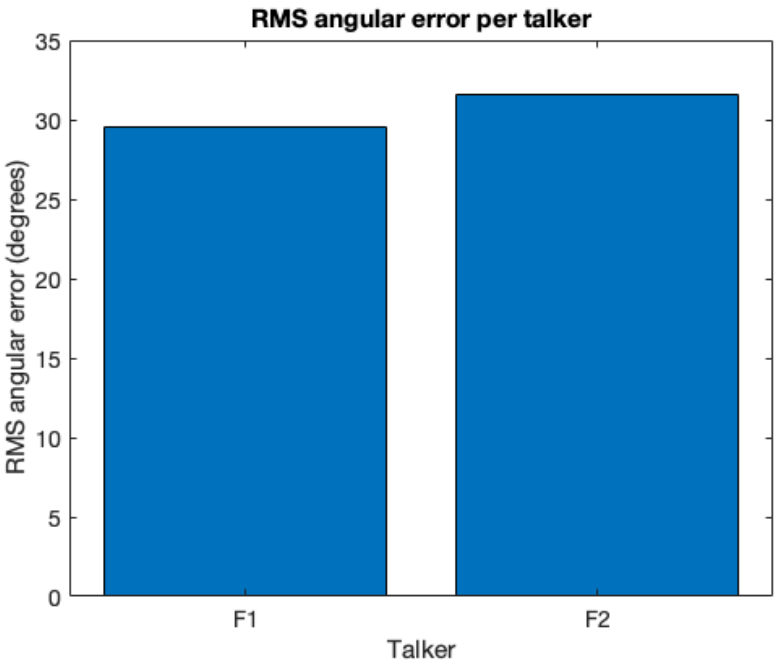


Figure 5.7: RMS angular error per talker.

Lastly, an analysis of the influence of the listener's experience on angular error was performed. Those participants that declared to be musicians and/or involved in research or business in the field of Audio or Acoustics were considered as "experienced listeners". There were 15 experienced listeners and 8 inexperienced listeners. Here, the populations considered were represented by the mean squared angular errors for each signal (that is, the squared angular errors averaged across participants for each signal), as can be seen in Table 5.3. Although there was a reasonably larger error for the inexperienced listeners, which was expected, the Sign Test could not reject the null hypothesis that the difference between the two populations generate a distribution with zero median ($p > 0.05$). The RMS angular errors for both listener groups are shown in Figure 5.8.

Table 5.3: Mean square angular errors of each signal for experienced and inexperienced listeners.

	Mean square angular errors (degrees^2)	
Signal	Experienced listeners	Inexperienced listeners
1	540	1012.5
2	780	675
3	780	675
4	840	2250
5	780	450
6	2040	1237.5
7	480	562.5
8	180	562.5
9	480	562.5
10	720	450
11	360	675
12	840	675
13	240	337.5
14	540	112.5
15	240	562.5
16	180	675
17	240	675
18	3120	2362.5
19	840	1350
20	240	675
21	540	787.5
22	2220	3487.5
23	660	787.5
24	660	225
25	1800	2475
26	480	1912.5
27	660	337.5
28	660	5175
29	600	2700
30	1080	1125
RMS angular error	28.18°	34.42°

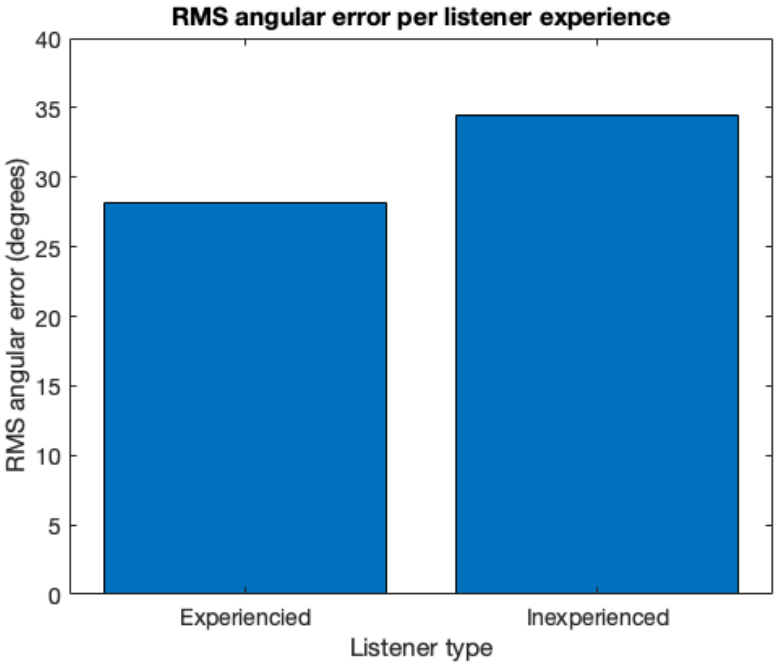


Figure 5.8: RMS angular error per listener's experience.

5.1.2 Evaluation of spreading

Alongside the evaluations regarding the angular error, the analyses regarding the influence of environment, compression threshold and listener's experience on the "diffusiveness" of sound were also performed. The influence of position and talker were not evaluated for this case, in order to avoid redundant information and to not stray away from the goal of this part of the work, which is to analyze the effect of compression (and non-linear fitting, for Test Two) in sound localization, rather than analyze the characteristics of human localization ability itself.

Figure 5.9 shows the percentage of answers "Precise", "More or less" and "Diffuse" associated to each environment. One can see that the anechoic chamber has the largest portion of answers in the "Precise" category, while the cafeteria is clearly associated to a larger quantity of "Diffuse" answers. This was predictable given the nature of these two environments and the fact that one could expect the signals to sound more spread-out in a very reverberant environment.

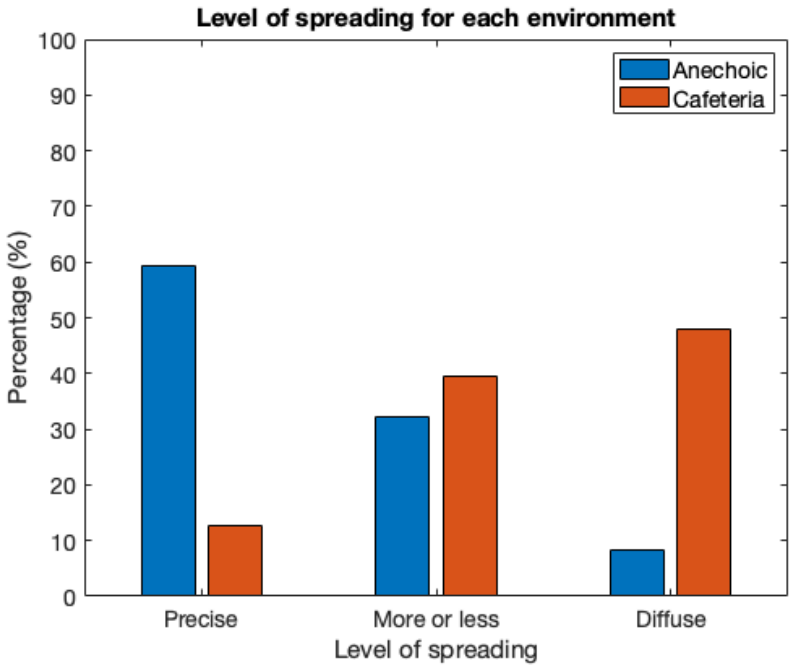


Figure 5.9: Percentage distribution of answers regarding the spreading of sound, for signals in the anechoic chamber and in the cafeteria.

For the analysis of statistical significance, however, a slightly different strategy was used. The verbal answers "precise", "more or less" and "diffuse" were converted to numbers 1, 2 and 3, respectively, in order to create a quantifiable scale, and were named as "grades". Then, in a similar way to what was done for the angular error, the mean grades for each participant, or each signal, compose the population to be analyzed. The Sign Test was applied when there were two populations, and the Friedman Test was used when there were three or more populations. For the case

of environments, the populations corresponding to the anechoic chamber and to the cafeteria are shown in Table 5.4. The Sign Test rejected the null hypothesis that the difference between the two populations form a distribution with zero median ($p < 0.05$). Therefore, one can affirm that sounds are significantly perceived as more diffuse in the cafeteria than in the anechoic chamber.

Table 5.4: Mean spreading grades per participant for each environment.

	Mean grades	
Participants	Anechoic chamber	Cafeteria
1	1.2667	2.6000
2	1.6000	2.1333
3	1.4000	2.4000
4	1.4000	2.6667
5	1.3333	2.5333
6	1.6000	2.5333
7	1.7333	2.6000
8	1.0667	2.5333
9	1.6000	2.2000
10	1.7333	2.2667
11	1.8000	2.1333
12	1.6000	2.1333
13	1.0000	1.8667
14	2.0000	2.4667
15	1.2667	2.4667
16	1.7333	2.6000
17	1.6000	2.4000
18	1.8667	2.3333
19	1.1333	2.5333
20	1.5333	2.6667
21	1.5333	1.9333
22	1.2000	1.9333
23	1.2667	2.1333
Population medians	1.5333	2.4000

Figure 5.10 shows the percentage of answers "Precise", "More or less" and "Diffuse" associated to compression threshold. From the picture, a closer look reveals that the uncompressed signals are considered the most precise, whereas the signals submitted to a 40-dB threshold compression are considered the most diffuse. This is coherent with our initial supposition that compression blurs the spatial perception, and a CT of 40 dB allows compression to act on a larger part of the signal, in comparison to compression initiating at 60 dB only. The Friedman Test revealed that there was significant difference between the means of the three populations ($p < 0.05$) and the multiple comparison test indicated that that the means of the three populations are all significantly different from each other, with $p < 0.05$. That reveals that, indeed, compression threshold is a factor that contributes to the perception of spreading of sound, even though it has not contributed to the increase in angular error itself. The lower the compression ratio — that is, the longer the signal stays in the compression region — the more diffuse the source is perceived. The mean grades per participant for each type of compression are shown in Table 5.5.

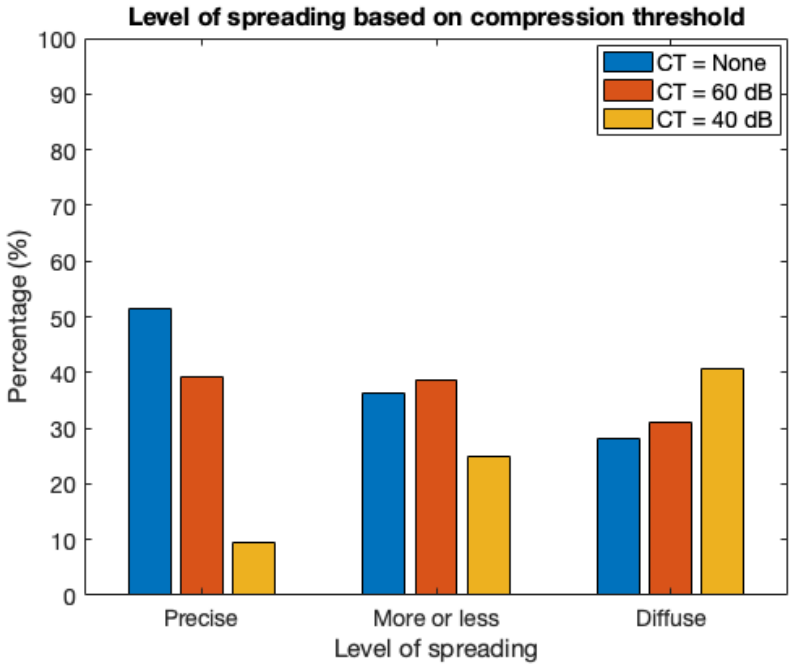


Figure 5.10: Percentage distribution of answers regarding the spreading of sound for signals with no compression, with compression starting at 40 dB, and with compression starting at 60 dB.

Table 5.5: Mean spreading grades per participant for each compression configuration.

	Mean grades		
Participants	CT = None	CT = 60 dB	CT = 40 dB
1	1.6667	1.9167	2.0833
2	2.0000	1.7500	1.9167
3	1.8333	1.7500	2.0833
4	1.6667	2.1667	2.0833
5	1.8333	1.8333	2.0833
6	1.8333	2.0000	2.2500
7	1.8333	2.0833	2.4167
8	1.6667	1.7500	1.9167
9	1.6667	1.9167	2.0000
10	1.3333	1.9167	2.4167
11	1.5000	2.0000	2.1667
12	1.3333	2.3333	1.6667
13	1.1667	1.3333	1.6667
14	2.0000	2.0833	2.5000
15	1.3333	1.9167	2.0833
16	1.8333	2.0000	2.5000
17	1.1667	1.8333	2.5833
18	1.5000	1.9167	2.5833
19	1.5000	1.7500	2.0833
20	1.8333	2.1667	2.1667
21	1.3333	1.9167	1.7500
22	1.1667	1.5000	1.8333
23	1.3333	1.5833	2.0000
Population means	1.5797	1.8877	2.1232

Lastly, the analysis of spreading perception based on listening experience was performed. Based on Figure 5.11, one could presume that experienced listeners (represented by the blue bars) are slightly more picky, since they have a slightly lower percent of "Precise" answers than the inexperienced group. Nevertheless, the Sign Test performed with the mean grades per signal, shown in Table 5.6, failed to reject the null hypothesis that the difference between the two populations originate a distribution with zero median ($p > 0.05$). Therefore, we can assume that listener's experience did not have a statistically significant influence on the perception of the spreading of sound.

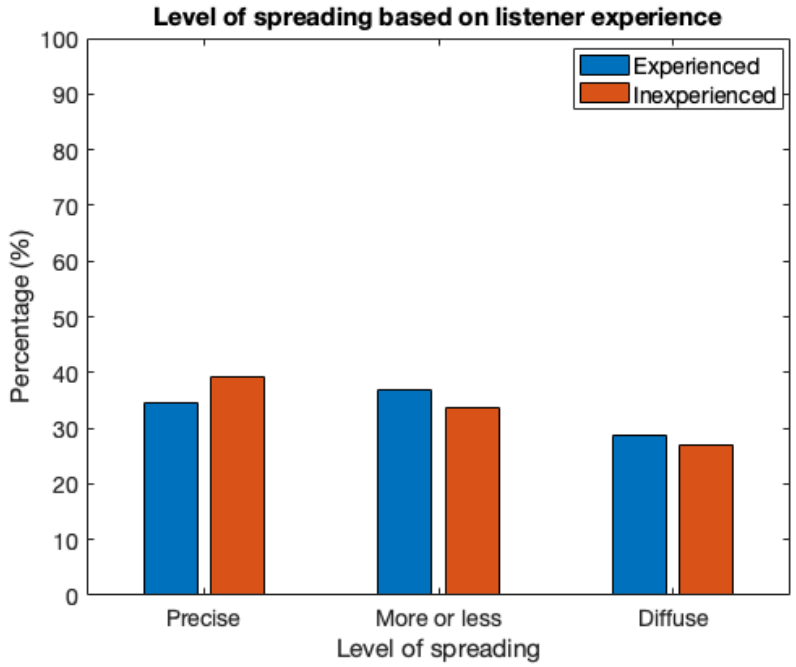


Figure 5.11: Percentage distribution of answers regarding the spreading of sound for experienced and inexperienced listeners.

Table 5.6: Mean spreading grades per signal for experienced and inexperienced listeners.

	Mean grades	
Signal	Experienced listeners	Inexperienced listeners
1	2.5333	2.3750
2	2.4667	2.5000
3	1.2000	1.5000
4	1.4000	1.1250
5	1.4000	1.1250
6	1.8000	1.8750
7	1.3333	1.3750
8	1.6667	1.0000
9	1.8000	2.0000
10	2.0000	2.3750
11	1.2000	1.0000
12	1.8667	2.3750
13	2.2000	2.7500
14	1.3333	1.1250
15	1.6667	1.1250
16	2.9333	2.8750
17	1.8000	2.0000
18	2.2000	2.0000
19	2.8667	3.0000
20	1.7333	1.6250
21	2.5333	2.1250
22	2.4000	2.3750
23	1.1333	1.0000
24	2.6000	2.6250
25	1.9333	1.7500
26	1.8000	1.8750
27	1.3333	1.2500
28	1.8000	1.7500
29	2.8000	2.5000
30	2.5333	2.0000
Population medians	1.8000	1.9375

Also presented in this subsection are the analyses of answers "Yes" and "No" for the extra question exhibited for Track 8 and Track 30. Judging from Figure 5.12, it seems that only Track 30 has a considerable amount of participants that answered "Yes" for whether the sound seemed to be coming from more than one direction. Indeed, Track 30 has a more perceptible bouncing effect than Track 8, particularly during the sibilant sound of the word "nosso". This can be explained due to the brief duration of this sibilant phoneme, as well as its higher level compared to the level of the other adjacent phonemes. The relatively high level of this phoneme triggers compression in the closest ear (given that the sound source is located at the $-90°$ azimuth) but, since the sound is very brief, the signal goes in and out of compression very quickly. The brief period in which compression is active produces a decrease in sound level in the left ear, and that sound momentarily seem to "bounce" to the right ear.

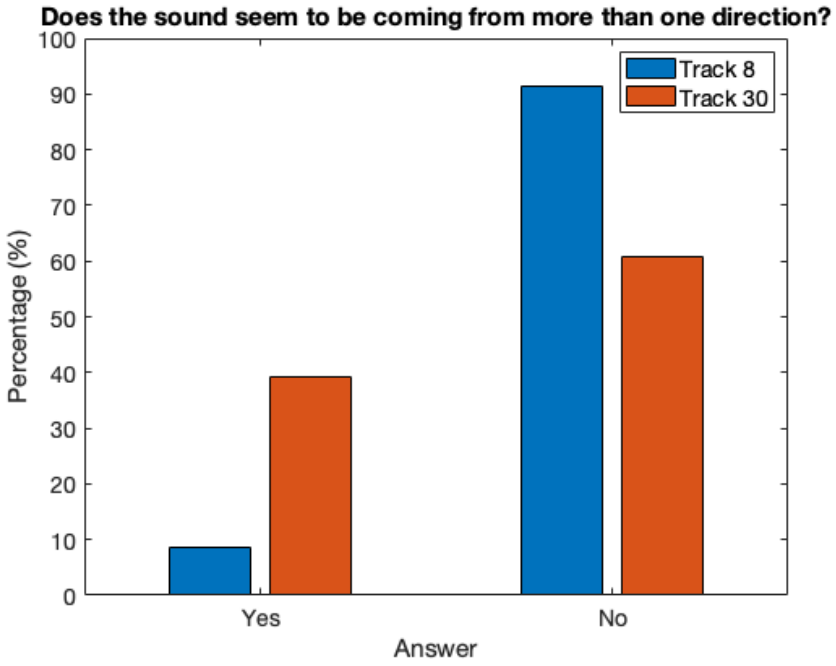


Figure 5.12: Percentage of "Yes" and "No" answers for the extra question formulated for Tracks 8 and 30.

5.1.3 Comparison with binaural cue measurements

The last analysis performed was a general comparison between the perceived azimuths and the binaural cue measurements. For that, the modes of the answers regarding azimuth for each signal were considered. The most frequently occured mistakes were front-back confusions, especially when the source was at the $0°$ or in the $180°$ azimuth. Nevertheless, as previously stated, it was chosen not to consider front-back confusions as mistakes, since the binaural cues of two symmetrical

positions on a cone of confusion are essentially the same.

In general, the uncompressed signals and their respective compressed versions produced the same modes. That reaffirms the findings in the previous sections that bilateral compression did not cause a significant change in the azimuth perception, but rather a change in the perception of the spreading of sound. When it comes to the binaural cue measurements, compression does indeed modify the ILD by some dBs (around 2 to 5 dB for most frequencies), and the change slightly increases with the decrease in compression threshold, as can be seen in Figure 5.13 for the binaural cue measurements of the compressed and uncompressed versions of sentence F2 in the anechoic chamber at the $-90°$ azimuth. Nevertheless, these changes do not seem to affect the azimuth perception correspondingly. For this and most other cases, the ITD remains virtually the same, except when its shape is already distorted in the uncompressed signal, as often happens for signals in the cafeteria environment, as was seen in Chapter 4. Since the magnitude and phase of the IC presented no different changes than those already described in Chapter 4, they are omitted from the figures in this chapter.

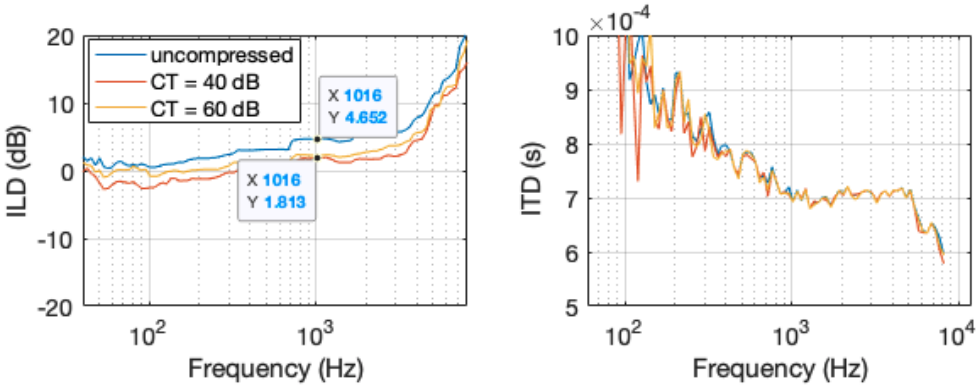


Figure 5.13: Measured ILD and ITD for the uncompressed and the two compressed versions of sentence F2 in the anechoic chamber at $-90°$ azimuth.

Aside from the front-back confusions, another systematic error that occurred was perceiving the $-60°$ azimuth as $-90°$ azimuth. Since the measured ILDs at 1 kHz and 4 kHz and the measured ITDs at 500 Hz and 1 kHz at $60°$ azimuth for an uncompressed signal are in accordance with their expected values for this azimuth, according to Figures 4.1 and 4.2, and since the changes in ILD due to compression are not changing the modes of the answers, it is safe to assume that this mistake is linked to perception itself. Given that people are worse at locating azimuths when they approach the sides of the head, an azimuth that is perceived "somewhere on the right side" might be interpreted already as being totally on the right side. Figure 5.14 shows the compressed and uncompressed versions of sentence F13 in the anechoic chamber at the $60°$ azimuth.

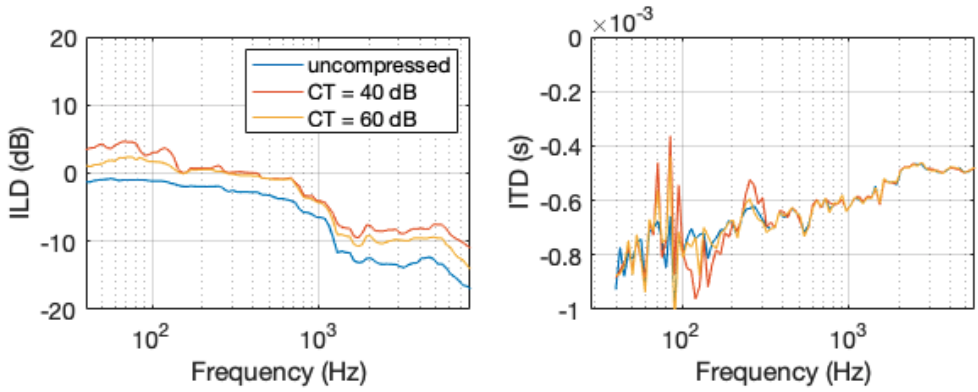


Figure 5.14: Measured ILD and ITD for the uncompressed and the two compressed versions of sentence F13 in the anechoic chamber at 60° azimuth.

Lastly, another noticeable mistake was in the azimuth of position C in the cafeteria, present 3 out of 5 times in the test signals. Among these 3 signals, 2 of them had as modes an azimuth of −150°, and one had as mode the azimuth −120°. The correct azimuth for this position would be −180°. It should be noted that for this particular position in the cafeteria, however, even the unprocessed signal has an ILD and an ITD above zero, as already mentioned in Chapter 4 and as can be seen in Figure 5.15 for the signals generated from sentence F2, contrary to what was expected for this position. By examining the expected values from the ITD and the ILD for the −150° azimuth in Figures 4.1 and 4.2, it can be verified that, although slightly above zero, the ILD and ITD for these signals at position C are still smaller than they should be for a −150° azimuth. This could mean, as already stated in Chapter 4, that the loudspeaker position was not exactly at the back of the listener's head when the HRTFs from Kayser [12] were measured, but rather leaning a bit towards the left side, somewhere between the azimuths 180° and −150°, or that some reflections happening at the table and chairs contribute to the impression that the sound does not come directly from the back. Even though the cues indicate a deviation of less than 30° from the 180°, listeners who heard the signal not exactly at the back probably preferred to chose as answers the next azimuth at the left side. The modes were not changed by compression in this position. This change cannot be associated to a particular sentence, since it was observed at this position for both sentences F2 and F13. A smoothed-out version of the ITD of the unprocessed signal is represented in Figure 5.15 by a thick line, to allow a better visualization.

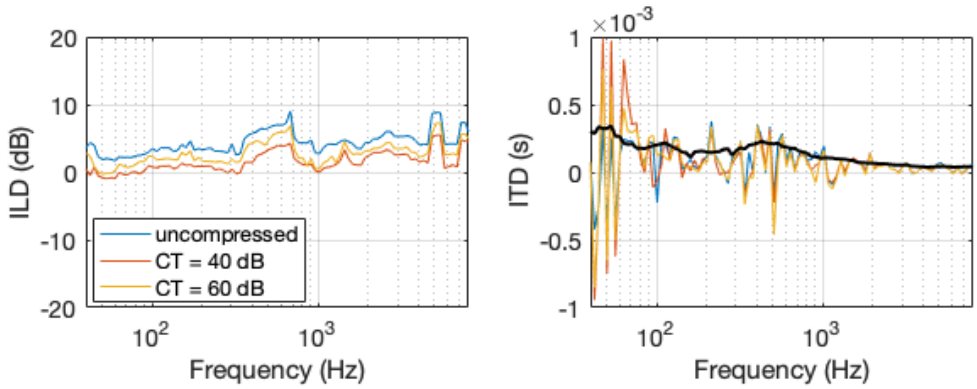


Figure 5.15: Measured ILD and ITD for the uncompressed and the two compressed versions of sentence F2 in the cafeteria at position C (back of the head). Smoothed-out ITD of the uncompressed signal is represented by a thick black line.

5.2 Test Two

Test Two focused on the effect of hearing loss and a non-linear amplification strategy on the localization performance. The original sentences F2 and F13 from the SMT database were, at the same positions in the anechoic chamber and in the cafeteria that were used for Test One. In order to modify the signals, the goal was to generate output signals that simulate a hearing loss configuration with a non-linear amplification strategy applied on top.

For that, the same procedure used to generate the signals from Analysis 3 in Experiments 1 and 2 in Chapter 4 was used — the use of specific audiograms in each ear as input for the FIG-6 fitting procedure, and a hearing loss compensation applied afterwards. Therefore, for each unprocessed spatialized signal, versions simulating hearing losses with a FIG-6 amplification were generated. Five hearing loss configurations were used: bilateral loss using audiogram N3, bilateral loss using audiogram S3, unbalanced hearing loss using audiogram S3 on the left ear and audiogram N3 on the right ear, and, for two signals only, unilateral hearing loss (audiogram S3 applied to the right ear, and audiogram N3 applied to the left ear).

The step of sound power equalization between channels, which was applied to the signals of Analysis 3 in both experiments of Chapter 4, was done here exclusively for the signals with unilateral loss, since those with bilateral loss do not need it, and those with unbalanced loss have a very small difference in sound power at both channels. It was also specifically desired to keep the small sound power difference between both channels, if there was any, for the S3N3 loss case, since most real life cases present an unbalanced hearing loss configuration that will inevitably imply a difference in sound power for the sound arriving at both ears.

The test was composed of 47 signals:

- 3 uncompressed spatialized signals of sentence F2 and 3 uncompressed spatialized signals of sentence F13, as references (sentences chosen randomly for each one of the three positions in the anechoic chamber and the three positions in the cafeteria);
- 12 signals with FIG-6 fitting applied to a bilateral N3 hearing loss configuration (2 sentences $\times$ 2 environments $\times$ 3 positions in each environment);
- 12 signals with FIG-6 fitting applied to a bilateral S3 hearing loss configuration (2 sentences $\times$ 2 environments $\times$ 3 positions in each environment);
- 12 signals with FIG-6 fitting applied to an unbalanced S3N3 hearing loss configuration — that is, S3 hearing loss at the left ear and N3 hearing loss at the right ear (2 sentences $\times$ 2 environments $\times$ 3 positions in each environment);
- 1 signal with unilateral N3 loss at the left ear and 1 signal with unilateral S3 loss at the right ear (both signals generated with sentence F13 at position $(-10°, 180°)$ in the anechoic chamber);
- 3 of the already generated signals presented at the beginning of the test, for confirmation (these signals appear twice in the test).

The signals were named Track 1–47 and were presented in the same order for all participants. The extra question was asked for Track 5 (bilateral S3 loss for sentence F13 at position $(-10°, 180°)$ in the anechoic chamber), Track 9 (unbalanced S3N3 hearing loss configuration for sentence F2 at position $(0°, -90°)$ in the anechoic chamber), Track 30 (unbalanced S3N3 loss for sentence F13 at position $(0°, -90°)$ in the anechoic chamber) and Track 33 (unprocessed sentence F2 at position $(0°, -90°)$ in the anechoic chamber). The first three of them were chosen for the extra question because they seemed to have the most noticeable bouncing effect. The last one was chosen as a reference, to test the participants' perception.

The answers given by one specific participant could not be considered for the analyses that follow, given that there was a mistake in the volume adjustment in this participant's turn. Therefore, the analyses were performed with the answers from the remaining 22 participants.

5.2.1 Evaluation of azimuth

The evaluation of the azimuth was done the same way as for Test One, with the squared angular error and the RMS angular error as the chosen measurements to quantify the participants' errors concerning the azimuth. Front-back confusions were also not considered here, and all azimuths were mirrored to the frontal hemisphere.

The correct azimuth is always considered to be the azimuth of the unprocessed spatialized signal, which is the position where the source is still supposed to be.

Figure 5.16 shows the RMS angular error for each signal, obtained from the squared angular errors averaged across participants. Almost all of the signals that present the biggest errors were those that have the sound coming from the back in the cafeteria, and those that have an unbalanced hearing loss configuration. In general, the RMS errors for each signal tended to be larger than those found for Test One. This can be explained due to a greater distortion of the shape of the binaural cues, as shown in Chapter 4 for Analysis 3 in both experiments, and a more evident change in sound quality due to the hearing loss simulation and the frequency-selective amplification.

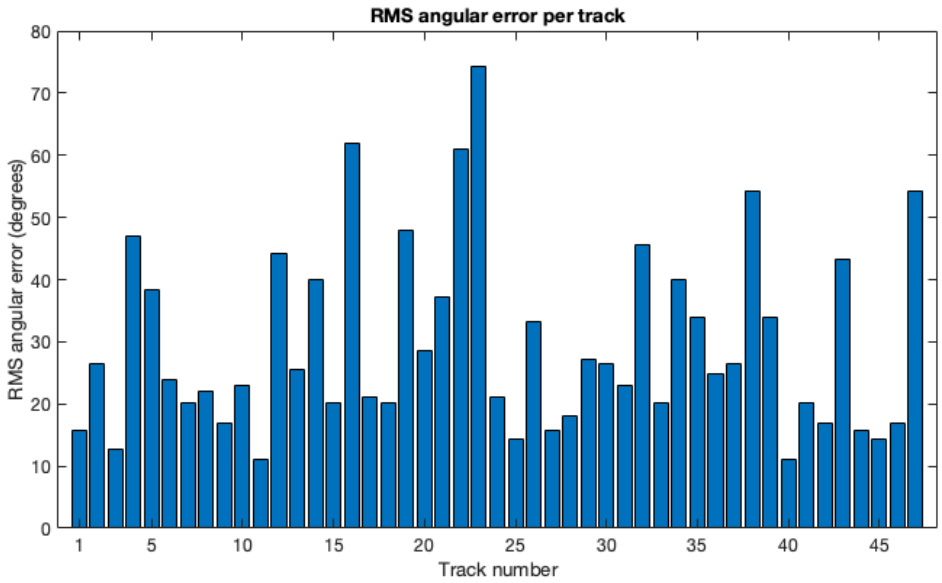


Figure 5.16: RMS angular error for each signal.

The angular error was analyzed with relation to position, environment, hearing loss configuration, talker, and listener's experience. The Friedman Test and the Signed Test for Paired Data were also used here in the same ways and conditions as for Test One, thus the tables showing the mean squared errors of the entire populations are omitted. Figure 5.17 shows the RMS angular error for each position in the anechoic chamber. Unlike the populations represented by the mean square angular errors for these positions in Test One, whose means had no significant differences between each other, here the Friedman Test and the multiple comparison test indicate significant difference ($p < 0.05$) between the mean square angular errors from sounds coming from the back and those from the other source positions. This behavior was actually more probable than in Test One, given the expected decreased localization ability for sounds coming from the back. The RMS angular errors shown in Figure 5.17 for all positions were slightly larger, indicating that the

signals of Test Two produced a larger distortion on the localization perception. For the positions in the cafeteria, as shown in Figure 5.18, the Friedman Test and the multiple comparison test revealed that the mean square angular errors for the azimuth 0° are significantly smaller than the mean square angular errors for the other two azimuths ($p < 0.05$). Figure 5.18 also presents a very similar result to the one illustrated by Figure 5.4 for Test One, indicating that there is indeed an increase in angular error with the increase in azimuth in the cafeteria.

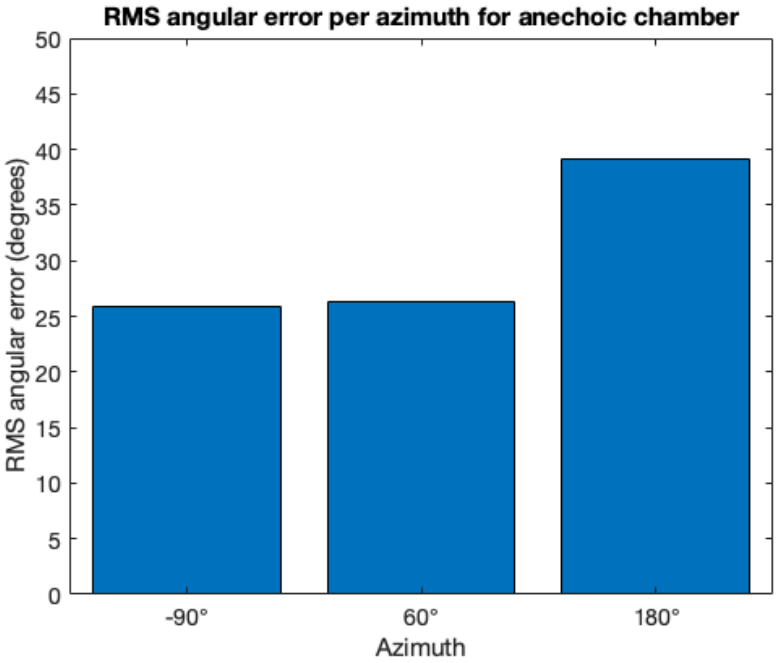


Figure 5.17: RMS angular error per azimuth in the anechoic chamber.

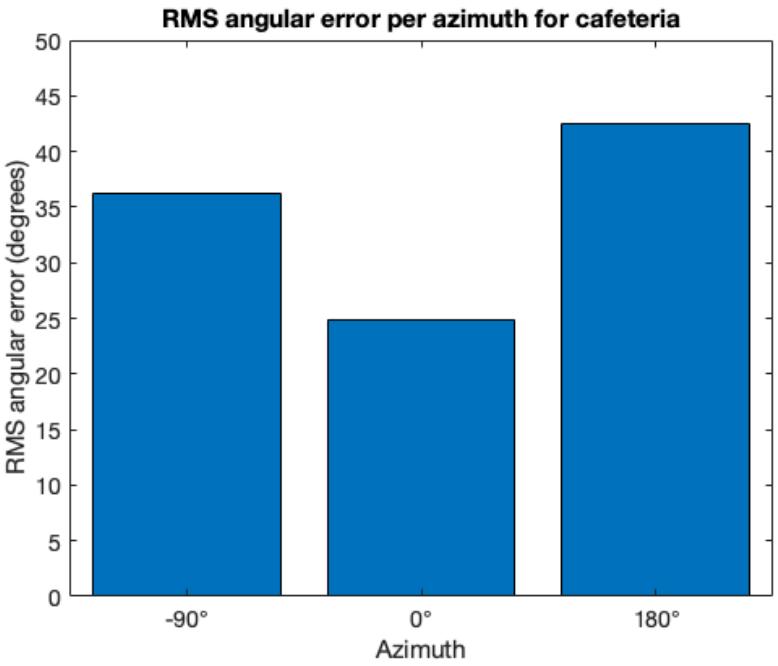


Figure 5.18: RMS angular error per azimuth in the cafeteria.

The Sign Test failed to reject the null hypothesis that the difference between the populations representing the mean square angular errors for anechoic chamber and cafeteria generate a distribution with zero median ($p > 0.05$). This means that, unlike Test One, Test Two failed to show a direct influence of environment in localization performance. This is probably due to the fact that the distortion caused by the processing in the signals of Test Two is greater than the distortion caused by compression in Test One and provides a similar degradation in localization performance for both environments, as illustrated by the RMS angular errors for each environment shown in Figure 5.19. The main difference between the results in this figure and the ones shown in Figure 5.5 is that the RMS angular error for the anechoic chamber is greater in Test Two, almost matching the RMS angular error for the cafeteria.

As far as the influence of hearing loss configuration is concerned, the Friedman Test and the multiple comparison test showed significant difference ($p < 0.05$) between the mean square angular errors of the S3N3 hearing loss and all the other configurations. This indicates that an unbalanced hearing loss configuration causes a larger increase in the angular error, since the inequality between the amplification strategies on both sides might cause a displacement of the perceived source location. The unilateral losses were not considered for the Friedman test, since there was only one signal as example of each, and a larger number of signals with unilateral losses would be necessary in order to allow a reliable statistical analysis. Nevertheless, their RMS angular errors are shown in Figure 5.20 along with the RMS angular er-

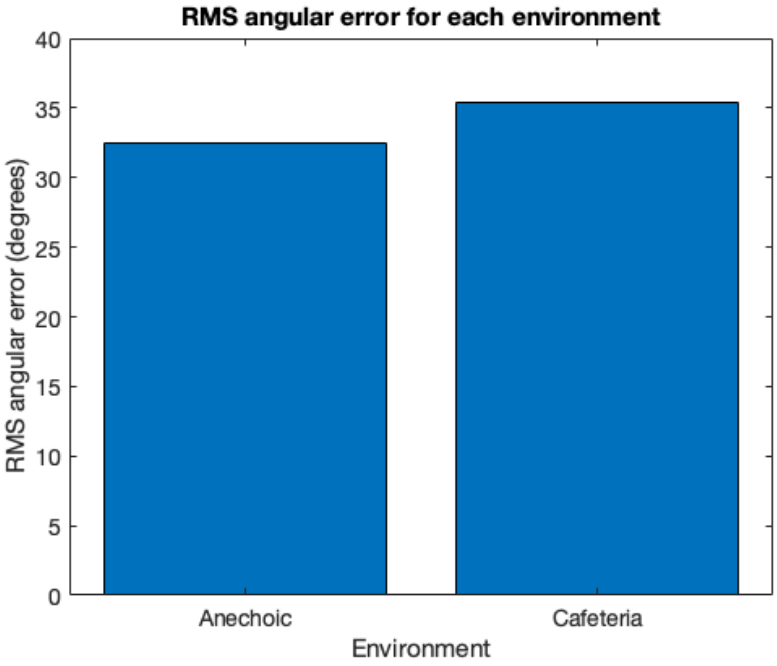


Figure 5.19: RMS angular error for each environment.

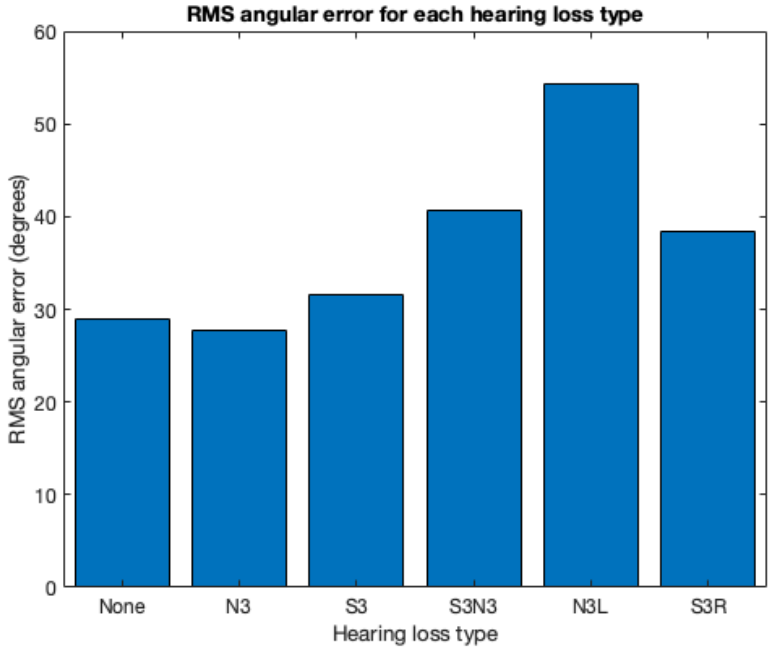


Figure 5.20: RMS angular error for each hearing loss configuration.

rors for the other hearing loss configurations. It can be seen that the unilateral losses also produce large RMS errors (particularly the signal with unilateral moderate loss at the left ear — N3L), probably for the same reasons associated to the unbalanced S3N3. For unilateral losses, this inequality is even more pronounced, given that the signal at the normal-hearing side is not adjusted by the fitting procedure at all.

Just as in Test One, the Sign Test failed to reject the null hypothesis that the difference in the populations represented by the mean square errors associated to Talker F1 and Talker F2 generate a distribution with zero median ($p < 0.05$). That means that, for both tests, the talker, and consequently also the different sentences, did not have a relevant influence on the angular error. The RMS angular error for each talker is presented in Figure 5.21.

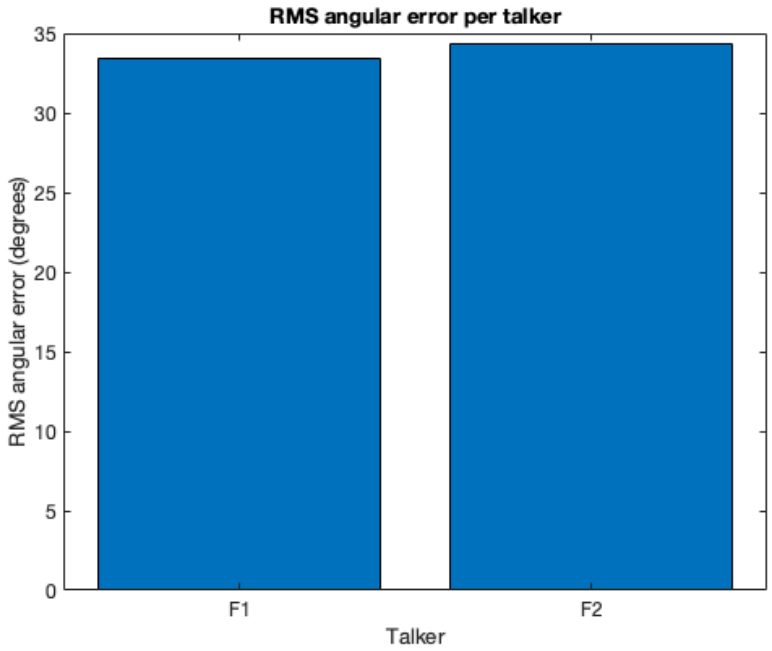


Figure 5.21: RMS angular error per talker.

Lastly, there was no relevant influence of the listener's experience on the angular error, as shown by the Sign Test, which rejected the null hypothesis that the differences between mean square angular errors of experienced and inexperienced listeners generate a distribution with zero median ($p > 0.05$). According to Figure 5.22, the difference in the RMS angular error for both listener groups in Test Two is smaller than that found for Test One, which, even so, was not considered statistically significant. In Test Two, the experienced group tended to have larger angular errors than in Test One.

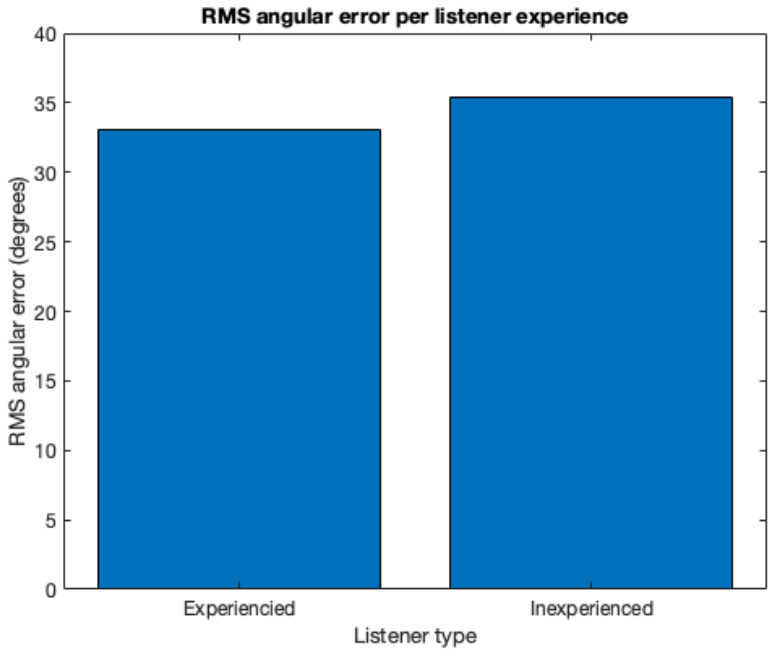


Figure 5.22: RMS angular error per listener's experience.

5.2.2 Evaluation of spreading

Similarly to what was done for Test One, the influence of some parameters on the responses regarding the spreading of sound was analyzed. Here, the parameters evaluated are environment, hearing loss configuration and listener's experience. This subsection also presents the evaluations of the Yes/No responses for the extra question asked for 4 out of the 47 signals in Test Two.

Figure 5.23 presents the percentage distribution of answers "Precise", "More or less" and "Diffuse" for the anechoic chamber and the cafeteria. The distribution is notably similar to that in Figure 5.9, indicating that the influence of the environment on the sensation of spreading of sound is virtually the same for Test One and Test Two. As in Test One, the answers "Precise", "More or less" and "Diffuse" were mapped into the values 1, 2 and 3, in order to enable statistical analyses. Table 5.7 shows the populations evaluated by the Sign Test, composed of the mean grades given by each participant to each environment. The Sign Test rejected the null hypothesis that the difference between the populations originates a distribution with zero median ($p < 0.05$). That indicates that the signals in the cafeteria are indeed perceived as more diffuse than in the anechoic chamber, even though there is no statistically significant difference concerning the angular errors themselves in both environments, as seen in the previous subsection.

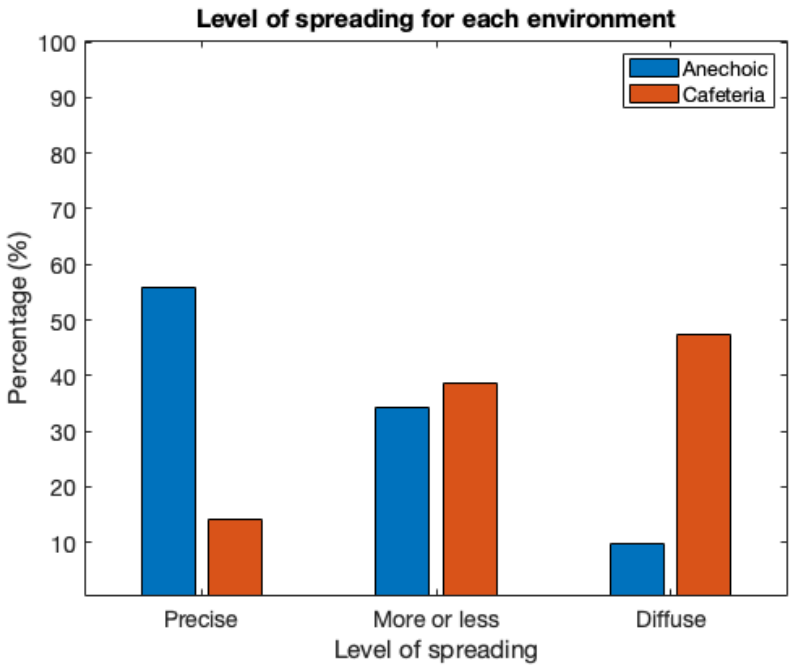


Figure 5.23: Percentage distribution of answers regarding the spreading of sound, for signals in the anechoic chamber and in the cafeteria.

Table 5.7: Mean spreading grades per participant for each environment.

	Mean grades	
Participants	Anechoic chamber	Cafeteria
1	1.3913	2.3333
2	1.4348	2.5714
3	1.6522	2.1429
4	1.0000	2.0952
5	1.5652	2.5238
6	1.3043	2.4762
7	1.2609	2.4286
8	1.1739	2.5238
9	1.2609	2.0000
10	1.5217	2.5714
11	1.7826	1.4762
12	1.7826	2.3810
13	2.1304	2.3810
14	1.5652	2.2857
15	1.8696	2.1429
16	1.9130	2.3333
17	2.0435	2.8095
18	1.0435	2.4762
19	1.8261	2.0000
20	1.1304	2.4286
21	1.6087	2.2381
22	1.6087	2.7143
Population medians	1.5652	2.3810

Figure 5.24 presents the percentage distribution of answers for each of the hearing loss configurations. The unilateral losses were not considered for the Friedman Test, since there is only one signal representing each, and it was chosen not to perform a statistical analysis based on only one representative of each case. Nevertheless, their median grades are shown in Table 5.8, along with the other medians and the populations formed by the mean grades per participant for each hearing loss configuration. According to the Friedman Test and the multiple comparison test, the mean grades for the hearing loss configurations N3, S3 and S3N3 are significantly higher than the mean grade for the no-hearing loss configuration ($p < 0.05$) and, by extension, the unilateral losses would probably also be significantly higher, based on their medians in Table 5.8. Therefore, one can confirm that all of the processed signals tended to have a more spread-out sound quality than their unprocessed counterparts, particularly those with a severe or unbalanced hearing loss configuration.

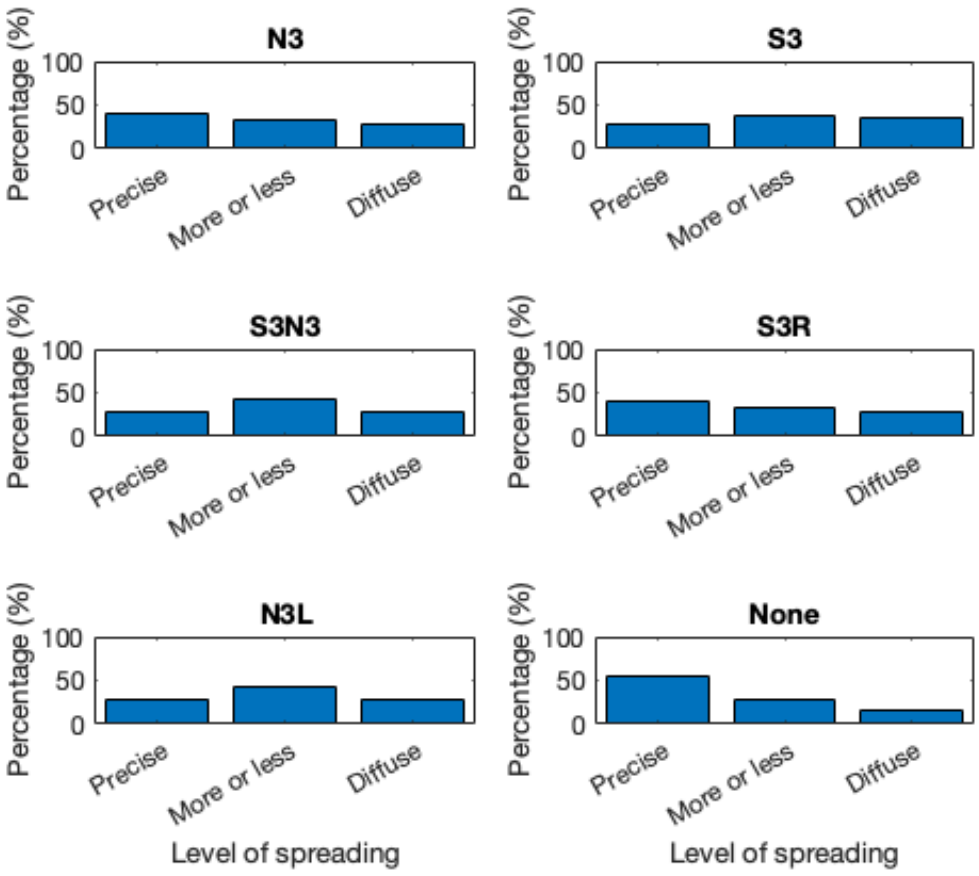


Figure 5.24: Percentage distribution of answers regarding the spreading of sound for each hearing loss configuration (hearing loss configuration appears on the title of each graph).

Table 5.8: Mean spreading grades per participant for each hearing loss configuration.

	Mean grades					
Participants	None	N3	S3	S3N3	S3R	N3L
1	1.3333	1.7500	2.1667	2.0000	1.0000	1.0000
2	1.1667	1.8333	2.4167	2.1667	2.0000	1.0000
3	1.5000	1.8333	1.9167	2.0000	3.0000	2.0000
4	1.3333	1.5000	1.6667	1.5833	1.0000	1.0000
5	1.8333	1.9167	2.2500	2.0833	2.0000	1.0000
6	1.6667	1.6667	2.0000	2.0000	2.0000	2.0000
7	1.5000	1.7500	2.0833	1.9167	1.0000	1.0000
8	1.6667	1.9167	2.1667	1.5833	1.0000	1.0000
9	1.3333	1.5833	1.7500	1.7500	1.0000	1.0000
10	1.6667	2.2500	2.1667	2.0000	1.0000	1.0000
11	2.1667	1.5000	1.4167	1.6667	2.0000	2.0000
12	1.5000	1.9167	2.0833	2.4167	2.0000	3.0000
13	1.3333	2.0833	2.8333	2.3333	3.0000	1.0000
14	1.5000	1.9167	2.1667	1.8333	3.0000	1.0000
15	2.0000	2.0833	1.9167	2.0000	2.0000	2.0000
16	1.8333	1.9167	2.0833	2.3333	3.0000	3.0000
17	2.1667	2.4167	2.5833	2.3333	3.0000	2.0000
18	1.5000	1.7500	1.7500	1.9167	1.0000	1.0000
19	1.5000	2.0833	2.0000	1.8333	2.0000	2.0000
20	1.3333	1.5833	1.9167	2.0833	1.0000	1.0000
21	1.5000	1.7500	2.0833	2.0000	3.0000	2.0000
22	1.8333	2.1667	2.2500	2.2500	1.0000	2.0000
Population means	1.5985	1.8712	2.0758	2.0038	1.8636	1.5455

Also following the procedure for Test One, an analysis of the influence of listener's experience based on the spreading of sound was performed. The percentage distribution of answers for each group of listeners is presented in Figure 5.25. Based on the figure, the experienced group, just like in Test One, seems to be slightly more demanding when it comes to localization precision, given their bigger percentage for "Diffuse" answers and smaller percentage for "Precise" answers, in comparison to the percentages associated with the inexperienced group. Nevertheless, just as in Test One, the Sign Test failed to reject the null hypothesis that the difference in the mean grades for both groups generate a distribution with zero median ($p > 0.05$), which means that the listener's experience does not significantly influence the perception of the spreading of sound. The populations used for the Sign Test, as well as their medians, can be seen in Table 5.9.

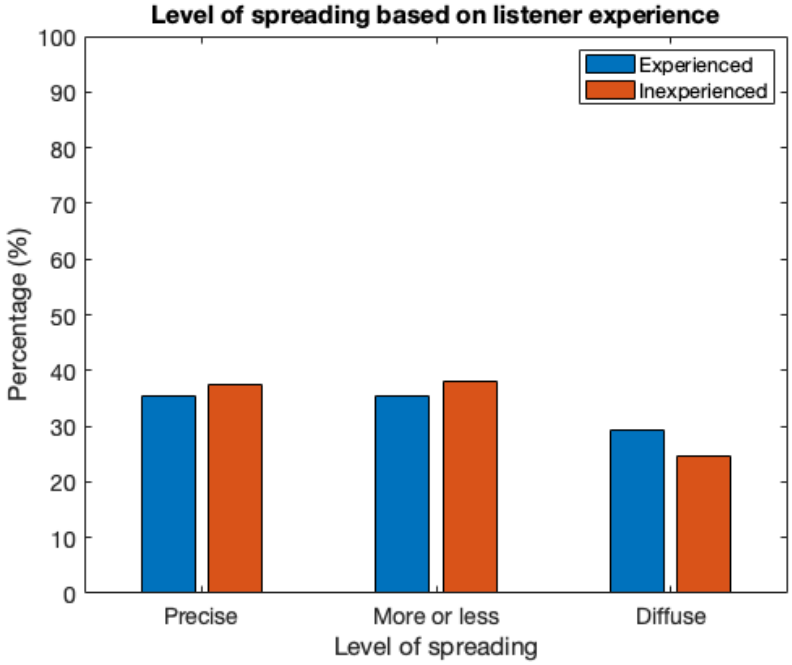


Figure 5.25: Percentage distribution of answers regarding the spreading of sound for experienced and inexperienced listeners.

Table 5.9: Mean spreading grades per signal for experienced and inexperienced listeners.

	Mean grades	
Signal	Experienced listeners	Inexperienced listeners
1	2.1818	1.8000
2	2.0000	1.6000
3	2.0000	1.4000
4	2.9091	2.4000
5	1.7273	1.6000
6	1.5455	1.4000
7	1.5455	1.6000
8	2.5455	2.6000
9	2.2727	1.6000
10	1.2727	1.0000
11	2.7273	2.6000
12	2.0909	2.0000
13	2.2727	2.4000
14	2.4545	2.4000
15	1.4545	2.0000
16	2.4545	2.2000
17	1.0000	1.0000
18	1.9091	2.4000
19	2.0909	2.4000
20	1.3636	1.2000
21	1.2727	1.8000
22	1.2727	1.4000
23	2.9091	2.4000
24	1.2727	1.2000
25	2.6364	2.0000
26	1.9091	2.2000
27	1.3636	1.2000
28	2.6364	2.6000
29	2.2727	2.4000
30	1.2727	1.4000
31	1.4545	1.8000
32	2.9091	2.4000
33	1.5455	1.2000
34	1.6364	2.0000
35	1.5455	2.4000
36	1.6364	1.6000
37	1.5455	1.8000
38	1.2727	1.4000
39	2.5455	2.4000
40	2.0909	2.0000
41	1.5455	1.4000
42	2.5455	2.8000
43	1.5455	1.6000
44	1.5455	1.4000
Population medians	1.9091	1.8727

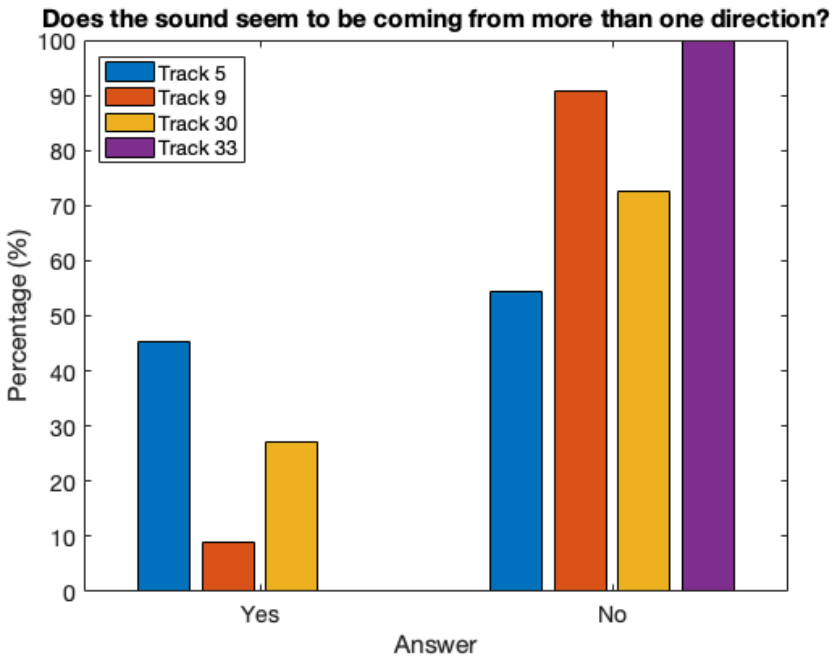


Figure 5.26: Percentage of "Yes" and "No" answers for the extra question formulated for Tracks 5, 9, 30 and 33.

Lastly, it was also desired to evaluate the percentage of Yes/No answers to the extra question, which was asked for Tracks 5, 9, 30 and 33. The percentage of answers "Yes" and "No" for each track are presented in Figure 5.26. It can be seen that the majority of participants chose "No" for all tracks, which means that most of them did not perceive sound coming from more than one direction. Track 33, which is an unprocessed spatialized version of sentence F2 at position $(0°, -90°)$ in the anechoic chamber, was used only as a reference, to verify the participants' perception and see if they were in any way influenced by the appearance of the extra question. Fortunately, there were no "Yes" answers for this track. What the other tracks have in common is the fact that they all have an unbalanced hearing loss configuration — Track 5 presents a severe unilateral loss at the right ear, while Tracks 9 and 30 present the S3N3 loss pattern. Prior to the listening tests, the unbalanced hearing loss was considered to be the main factor to cause a "bouncing" effect in the signals of Test Two. Judging from the percentage of answers in Figure 5.26, the highest percentage of "Yes" answers are coherent with this assumption, since they were for Track 5, which has the most unbalanced hearing loss configuration possible — that is, a severe unilateral loss. There could also have been an influence of position, since the source in Track 5 is at the back, while it is at the left side on the other tracks. The higher number of "Yes" answers for Track 30 in comparison to Track 9 is very likely due to the different sentences (F13 in Track 30 and F2 in Track 9). As seen earlier in this chapter, sentence F13, as spoken by talker F1, has a sibilant sound in the word "nosso" that is very sensitive to changes caused by compression. This

phoneme, in a similar way that was perceived for Track 30 in Test One, makes the soure position seemingly quickly shift from one side of the head to the other.

5.2.3 Comparison with binaural cue measurements

When analyzing the modes of the answers for azimuth, the most notable phenomena observed were:

- Frequent occurrence of front-back confusion for sounds coming from the front or from the back;
- Perception of the 60° azimuth at 90° azimuth;
- Systematic perception of position C in the cafeteria (which should be at 180° azimuth) at 150° azimuth;
- Variation of the modes for the position $(-10°, 180°)$ in the anechoic chamber for most signals with an unbalanced hearing loss configuration.

Firstly, the front-back confusions were disregarded, since it was chosen, for Test One as well as for Test Two, not to consider them as mistakes. The perception of the 60° azimuth at 90° is identical to the phenomenon observed for Test One. As previously checked for Test One, the ILDs (at 1 kHz and 4 kHz) and ITDs (at 500 Hz and 1 kHz) for the unprocessed signals at 60°azimuth are approximately matching the expected values for these cues at this azimuth illustrated in Figures 4.1 and 4.2. Since the change in the ILD in the upper frequencies caused by the non-linear amplification in the processed signals at this azimuth did not change the modes of the answers (which were already 90° instead of 60° for the unprocessed signals at this azimuth), one can assume that, as already explained in more detail in Subsection 5.1.3 for Test One, this perception of the 60° azimuth at 90° is probably linked to the decreased sensitivity in the perception of azimuths near the side of the head. Figure 5.27 shows the binaural cues for the unprocessed and processed versions (bilateral N3 hearing loss, bilateral S3 hearing loss and unbalanced S3N3 hearing loss) of sentence F13 at position $(20°, 60°)$ in the anechoic chamber. In the figure, "N3 both" and "S3 both" (as in "both ears") refer to the cues from the bilateral N3 and S3 hearing loss configurations, respectively, while "S3 N3" refers to the unbalanced S3N3 hearing loss configuration. Since there were no new relevant changes for the magnitude and phase of the IC, its plots are omitted in Figure 5.27 and in the others that follow.

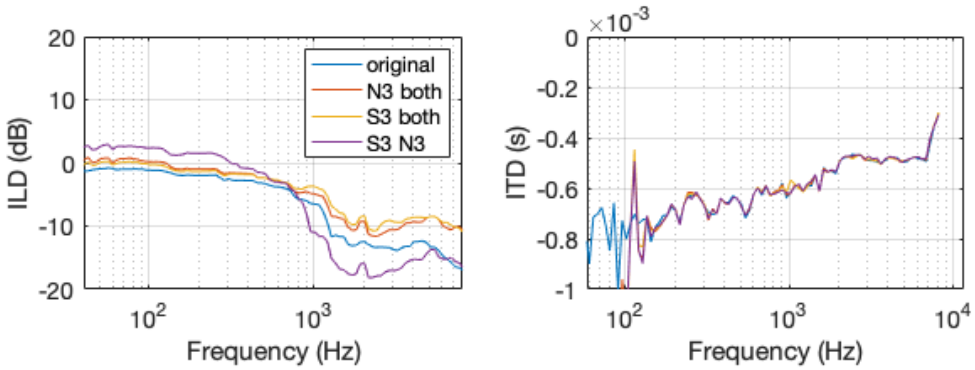


Figure 5.27: Measured ILD and ITD for the unprocessed and the three non-linearly fitted versions of sentence F13 in the anechoic chamber at 60° azimuth.

Similarly to what was observed for Test One, the modes for the signals at position C (at the back of the listener) in the cafeteria were systematically at −150°, although the source should theoretically be at the 180° azimuth. Since it was already shown in Subsection 5.1.3 that the ILD and the ITD of the unprocessed signals at this position indicate a source position between 180°and −150° azimuth, and that the alterations of the ILD caused by the non-linear fitting did not influence the modes, there is further evidence that maybe the loudspeaker position was not exactly at the back of the head, but very slightly leaning towards the −150° azimuth, or that reflections in the cafeteria contributed to that impression. Figure 5.28 shows all the variations of the sentence F2 at position C (back of the head) in the cafeteria.

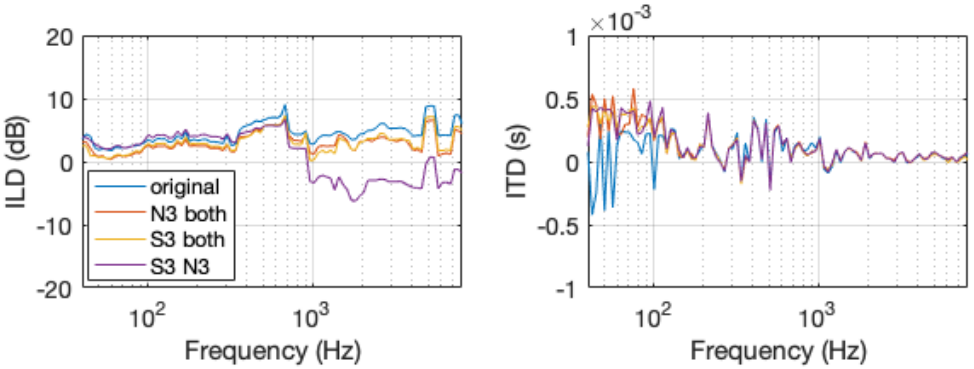


Figure 5.28: Measured ILD and ITD for the unprocessed and the three non-linearly fitted versions of sentence F2 at position C (back of the head) in the cafeteria.

It can be noted in both Figures 5.27 and 5.28 that the non-linear fitting applied to the S3N3 hearing loss configuration caused a noticeably large change in ILD, in comparison to the ILDs of the unprocessed signals, from around 1 kHz onwards. It is therefore impressive that such a difference caused no change in the modes for the two cases mentioned above. Nevertheless, the S3N3 and the unilateral loss configurations

produced a change in the modes for signals at position $(-10°, 180°)$ in the anechoic chamber, even though the differences for this position were much more discrete. This can be due to the fact that the perception of the sound coming from the back of the head is more precise for the signals in the anechoic chamber than for those in the cafeteria, and a slight difference in the ILD is therefore more easily perceived. For one of the two signals at this position with an S3N3 hearing loss configuration, the mode was 150°. For the signal at this position with unilateral severe hearing loss at the right ear, the mode was −150°, and for the signal at this position with unilateral moderate hearing loss at the left ear, the mode was 120°. This indicates a tendency to perceive the source closer to the ear that has the least hearing loss. Figure 5.29

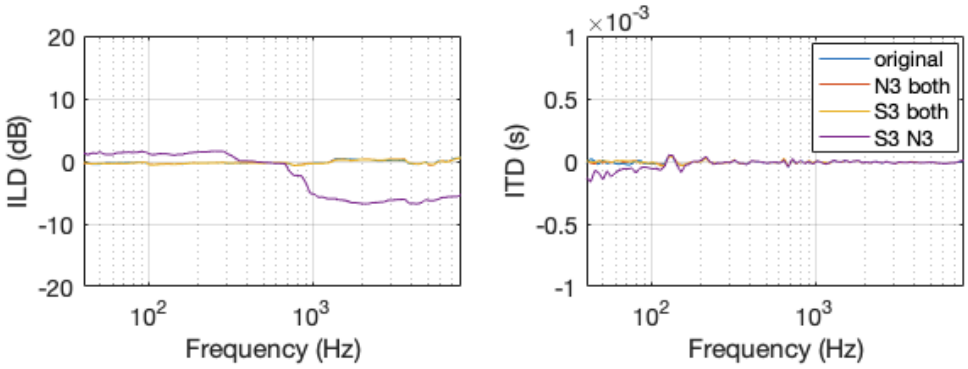


Figure 5.29: Measured ILD and ITD for the unprocessed and the non-linearly fitted versions of sentence F2 at position (-10°,180°) in the anechoic chamber.

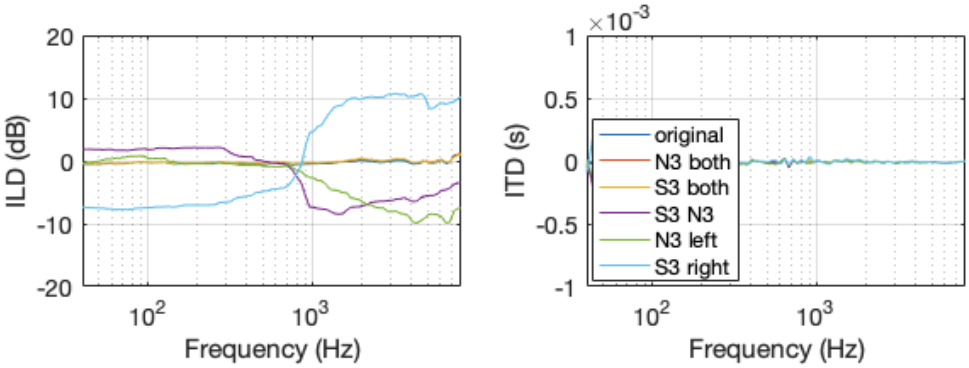


Figure 5.30: Measured ILD and ITD for the unprocessed and the non-linearly fitted versions of sentence F13 at position (-10°,180°) in the anechoic chamber.

shows the binaural cues for all the versions of sentence F2 at position $(10°, -180°)$ in the anechoic chamber, and Figure 5.30 shows the binaural cues for all the versions of sentence F13 at the same position (including the cases for the unilateral losses, represented by "N3 left" and "S3 right"). The ILD for the "S3N3" case in Figure 5.29 roughly matches, in absolute values, its expected value at 1 kHz and 4 kHz for a 150° azimuth according to Figure 4.2, if one disregards the small notch at 4 kHz

presented by the solid line, in comparison to the dashed line (this figure actually shows the values for the $-150°$, whose absolute values should be the same as those for 150°). The ILD for the "N3 left" case in Figure 5.30 roughly matches its expected value at 4 kHz for a 120° azimuth according to Figure 4.2, considering the dashed line and disregarding the notch in the solid line at that frequency — this evidences a shift towards the southeast direction. Lastly, the closest match for the ILD of the "S3 right" case in Figure 5.30 are the expected values presented by the dashed line at 1 kHz and 4 kHz for the $-120°$ azimuth according to Figure 4.2, which possibly indicates an even bigger deviation than the one observed in the mode for this signal, and corroborates a shift in source position towards the southeast direction. It must be noted that the ILDs for all unbalanced hearing loss configurations in both figures have a roughly approximate sigma-like shape. An example where this is very evident is the "S3 right" case in Figure F13, whose ILD has an approximate value of almost -10 dB until 800 Hz, and then becomes positive afterwards. This phenomenon can also contribute to the perception of a more "spread-out" sound when the non-linear fitting is applied, since there is a possibility that part of the spectrum seems to be coming from one direction, while the other part seems to be coming from another.

Chapter 6

Conclusions

This work aimed to analyze the effect of compression and frequency-selective amplification in hearing aids on the binaural cues and on the spatial perception, particularly in the horizontal plane. Spatialized speech signals were generated by using HRIRs recorded at an anechoic chamber, at an office and at a cafeteria. Three theoretical analyses were performed for signals in an anechoic chamber and in two reverberant environments: the isolated effect of a single-channel compression strategy (Analysis 1), the effect of a linear amplification strategy (Analysis 2), and the effect of a non-linear amplification strategy (Analysis 3) on binaural cues of the spatialized signals. Following that, two subjective tests (Test One and Test Two) were applied to normal-hearing participants in order to collect subjective perceptions regarding the azimuth and the spreading of sound in the environment.

The interaural level difference, the interaural time difference and the interaural coherence were calculated according to the equations described by Marquardt et al [10, 11] and Doclo [42]. The calculated ILDs were all in accordance, for the two frequencies chosen for analysis (1 kHz and 4 kHz), with the reference ILDs for each azimuth presented in Kayser [12]. The ITDs, on the other hand, presented an overestimation between 50 and 100 ms at the reference frequencies of 500 Hz and 1 kHz for azimuths between 0° and 180°. This deviation was associated to the fact that they were calculated based on the signals themselves, rather than the HRTFs as in Kayser [12]. Nevertheless, they showed the same pattern as in Kayser [12] regarding the relative differences between their values at 500 Hz and 1 kHz and regarding their behavior across azimuths. The observed pattern for the ITD both in this and the aforementioned work, due to the position of the ears at an azimuth of ±100° was an increase from 0° until 60°, then a decrease until 90°, another increase between 90° and 120° and then a continuous decrease until the azimuth reached 180°.

The magnitude of the interaural coherence had its median value around 1 for most cases, but showed a different behavior in the anechoic chamber and in the

reverberant environments. In the anechoic chamber, notches were observed at numerous frequencies, which was a possible indication of a lack of spectral content at those frequencies. In the reverberant environments, however, a different behavior was verified: while the magnitude of the IC was exactly 1 along the entire frequency range for the signals generated from the sentences with female voices, the signals generated from the sentence spoken by a male voice presented a considerable number of notches. Even though reverberation tends to keep spectral content active longer, mixing it in a way that can potentially "fill in" the gaps, the different spectral profile of a male voice is not capable to produce the same effect as the female voice. The phase of the interaural coherence had an approximately linear behavior for signals in the anechoic chamber (when seen with a linear scale), whereas, for signals in the reverberant environments, it usually exhibited a more distorted shape.

For Analysis 1, the effect of single-channel bilateral and unilateral compression strategies were evaluated. It was seen that, for bilateral compression, the changes in the binaural cues were subtle, and this translated into subjective perception in Test One as a sensation of a more "spread-out sound", with significant differences between all three groups of signals (no compression, compression with CT = 40 dB, and compression with CT = 60 dB), even though the presence of compression did not produce a significant difference in the angular error. The cases with unilateral compression in Analysis 1 for all environments, since there was no sound power equalization between channels, were evaluated primarily to see if the binaural cue measurements were responding properly. While there were only minor changes in the ITD, the ILD changed proportionally towards the uncompressed side, illustrating the perception of a source position shift or a leakage of part of the signal's content towards the uncompressed side. The influence of compression threshold and attack and release times were also evaluated for unilateral compression, for signals in the anechoic chamber. It was seen that the lower the compression threshold and the shorter the attack time, the greater the ILD. Conversely, the longer the release time, the greater the ILD. It must be noted, however, that the effect of attack and release times on the ILD were much subtler than that of compression threshold. An effect observed both for some bilaterally and unilaterally compressed signals was a "bouncing" effect. This bouncing effect was expected for unilaterally compressed signals since an equalization between the sound powers in both channels was not performed, but, for bilaterally compressed signals, it was a more interesting phenomenon, associated to compression being triggered independently in each channel.

In terms of subjective perception regarding bilateral compression, there were also other interesting findings based on the answers of Test One. It was observed that there was a greater difficulty in detecting the source location of sounds coming from the back, in comparison to the front, especially for signals in the cafeteria

environment. This is in accordance to statements in the literature that associated a worse localization performance for sounds at the back, in comparison to sounds at the front. The environment also directly influenced the participants' answers — both the angular error and the perception of spreading for signals in the cafeteria were significantly higher for signals in the cafeteria in comparison to signals in the anechoic chamber. The influence of neither the talker/sentence nor the listening experience was statistically significant on the angular error and on the perception of spreading. One out of the two signals for which the extra question ("Does the sound seem to be coming from more than one direction?") was asked presented a noticeable amount of "Yes" answers, due to a perceivable bouncing effect happening when a specific sibilant sound was delivered by the talker.

The modes of the answers for Test One were not influenced by the presence of compression, even though compression produced changes in the ILD. This indicates that the participants were apparently not sensitive to those changes. Systematic errors that were shown by the modes of the answers were front-back confusions, perceiving the 60° azimuth as 90° azimuth, and perceiving the position at the back of the head in the cafeteria at 150° azimuth. The perception of 60° as 90° was associated to the decreased localization ability described in the literature for sounds approaching the sides of the head, in comparison to sounds near the frontal midline. The perception of the position at the back of the head in the cafeteria at a 150° azimuth, however, has possibly more solid reasons. It is suspected that either the position of the loudspeaker was slightly leaning to the left hemisphere when the HRIRs were recorded, or that reflections on near obstacles could contribute to the impression that sound was coming from a southwest direction instead of coming directly from the back.

Analyses 2 and 3 evaluated the effect on the binaural cues of a linear and a non-linear fitting procedure, respectively. While there was no relevant change in the ITD for both bilateral and unilateral hearing loss configurations, the shape of the ILD for the unilateral loss configurations — now with signal power equalization between both channels — was noticeably distorted for both Analysis 2 and 3. For both bilateral and unilateral hearing loss configurations in both analyses, however, the source, or at least the most part of the signal's content, still seemed to be at the original position, even though there was a sensation of a less spatialized and more "spread-out" sound. For both analyses, the spatial perception was slightly more degraded when the hearing loss at the fitting procedure was applied to the ear closest to the sound source.

With those observations in mind, one can conclude that the linear and the non-linear fitting procedures generated similar results — therefore, the presence of compression in the non-linear fitting procedure did not worsen the spatiality of sound,

in comparison to linear amplification. The one aspect that really differed for the two analyses, however, was sound quality. It was observed that the non-linearly fitted signals of Analysis 3 sounded more realistic, brighter and with more low-frequency emphasis than the linearly fitted signals of Analysis 2. The brighter characteristic can be associated to superior harmonics introduced by compression, which is absent in Analysis 2, and the low-frequency emphasis can be due to the shrinkage in dynamic range caused by compression, which makes the signal sound denser. A later analysis also indicated that the non-linear fitting procedure indeed showed a tendency of prescribing more gain to low and high frequencies than the linear procedure for the signals in this study, which also contributed to the better resulting sound quality of non-linearly fitted signals. Differences in sound quality were also observed for the two types of audiograms used (one representing a moderate hearing loss and the other representing a moderate-to-severe steeply sloping hearing loss). The more severe hearing loss configuration always contributed to a much more muffled sound, and this negatively affected the spatial perception, as it was found by applying Test Two.

The subjective perceptions regarding non-linearly fitted signals evaluated with Test Two complemented the observations done for Analysis 3. The statistical analysis showed that the sounds coming from the back produced significantly greater angular errors in the anechoic chamber, while sounds from the front produced significantly smaller angular errors in the cafeteria. The increase in angular error with the increase in azimuth in the cafeteria was observed here as well as in Test One, and this is consistent with findings in the literature that associated worse localization performance with azimuth increase. Unlike in Test One, there was no significant influence of the environment in the angular error. This is due to the fact that the signals in Test Two produced in general larger errors, and the angular error for the signals in the anechoic chamber in Test Two are visibly higher than those for signals in the anechoic chamber in Test One. Nevertheless, like in Test One, signals in the cafeteria were significantly perceived as more diffuse than in the anechoic chamber.

All hearing loss configurations had a significant contribution for the perception of spreading of sound, and this was particularly noticeable for the S3N3 and the bilateral S3 cases. It was seen that the unbalanced hearing loss configurations (unilateral and S3N3-type configurations) are associated to a greater increase in angular error. Prior to Test Two, unbalanced hearing loss configurations were associated to a more noticeable bouncing effect in the signals, and this was proven true by the answers to the extra questions asked for four signals in Test Two. Three of these four signals had an unbalanced hearing loss configuration, and the signal which received the largest quantity of "Yes" answers had a severe unilateral hearing loss configuration, and also presented the same sibilant sound associated to the bouncing effect

in Test One.

Lastly, the influence of talker/sentence and listener experience on the angular error and the spreading of sound was verified. As in Test One, there was no significant influence of talker/sentence on the angular error, and no significant influence of listener experience neither on the angular error nor on the perception of the spreading of sound.

Most of the systematic errors made by the majority of participants in terms of azimuth are the same as those observed in Test One. These were frequent front-back confusions, perception of the 60° azimuth at 90° azimuth, and perception of the position at the back of the head in the cafeteria at a 150° azimuth. The first error is always expected in spatial perception tests and, as in Test One, was not considered, but the former two errors were associated to the exact same reasons for their occurence in Test One. A systematic error which happened particularly in Test One was the perception of the source position at the back of the head in the anechoic chamber either at the southwest or southeast direction, when an unbalanced hearing loss configuration was used. While the partipants were less sensitive to ILD changes related to hearing loss configurations in other cases, this case showed an increased sensitivity to the changes in the ILD, and the modes of the answers reflected a tendency of perceiving the source position closer to the side with the least hearing loss.

In a summarized way, it can be said that compression and frequency-selective amplification were not found to cause a drastic change in azimuth perception, but contribute negatively to the perception of externalization and generate the sensation of a more "spread-out" sound source. The degradation of spatial perception is increased when an unbalanced hearing loss configuration is used, and this should be considered with a lot of care, since the vast majority of hearing-impaired people have different hearing loss characteristics in each ear. For future works, a study on the influence of elevation on the spatial perception is desirable, as well as the opportunity to apply the listening tests on actual hearing aid users using their hearing aids instead of headphones. Another important next step is to include the noise reduction step prior to the fitting procedure, in order to analyze the amount of additional damage caused by denoising techniques. It is hoped that, in the near future, all this knowledge can make room for the creation of more comfortable and more realistic-sounding hearing aids, in order to provide hearing aid users the best quality of life possible.

References

[1] WORLD HEALTH ORGANIZATION. "Deafness and Hearing Loss". Available at `https://www.who.int/health-topics/hearing-loss`, 2022. Accessed 10/10/2022.

[2] WORLD HEALTH ORGANIZATION. "Fact sheets: Deafness and Hearing Loss". Available at `https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss`, 2022. Accessed 10/10/2022.

[3] MCCORMACK, A., FORTNUM, H. "Why do people fitted with hearing aids not wear them?" *International Journal of Audiology*, v. 52, n. 5, pp. 360–368, May 2013.

[4] DILLON, H., DAY, J., BANT, S., et al. "Adoption, use and non-use of hearing aids: a robust estimate based on Welsh national survey statistics", *International Journal of Audiology*, v. 59, n. 8, pp. 567–573, June 2020.

[5] KUK, F., KORHONEN, P. "Localization 101: Hearing Aid Factors in Localization". Available at `https://hearingreview.com/hearing-loss/hearing-disorders/apd/localization-101-hearing-aid-factors-localization`, 2014. Accessed 10/12/2022.

[6] KEIDSER, G., ROHRSEITZ, K., DILLON, H., et al. "The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers", *International Journal of Audiology*, v. 45, n. 10, pp. 563–579, October 2006.

[7] HASSAGER, H. G., WIINBERG, A., DAU, T. "Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment", *Journal of the Acoustical Society of America*, v. 141, n. 4, pp. 2556–2568, April 2017.

[8] DOCLO, S., MOONEN, M. "GSVD-based optimal filtering for single and multi-microphone speech enhancement", *IEEE Trans. Signal Processing*, v. 50, n. 9, pp. 2230–2244, September 2002.

[9] DOCLO, S., SPRIET, A., WOUTERS, J., et al. "Speech Distortion Weighted Multichannel Wiener Filtering Techniques for Noise Reduction". In: Benesty, J., Makino, S., Chen, J. (Eds.), *Speech Enhancement*, Springer-Verlag, cap. 9, pp. 199–228, Berlin, 2005.

[10] MARQUARDT, D., HOHMANN, V., DOCLO, S. "Interaural Coherence preservation in multi-channel Wiener filtering-based noise reduction for binaural hearing aids", *IEEE/ACM Trans. Audio, Speech and Language Processing*, v. 23, n. 12, pp. 2162–2176, December 2015.

[11] MARQUARDT, D., HADAD, E., GANNOT, S., et al. "Theoretical analysis of linearly constrained Multi-channel Wiener Filtering algorithms for combined noise reduction and binaural cue preservation in binaural hearing aids", *IEEE/ACM Trans. Audio, Speech and Language Processing*, v. 23, n. 12, pp. 2384–2397, December 2015.

[12] KAYSER, H., EWERT, S. D., ANEMÜLLER, J., et al. "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses", *EURASIP Journal on Advances in Signal Processing*, v. 2009, July 2009. Article ID 298605 (10 pages).

[13] FELS, J. "Lecture Notes to Medical Acoustics I + II". 2016. Institute for Technical Acoustics, RWTH Aachen University.

[14] FASTL, H., ZWICKER, E. *Psychoacoustics: Facts and Models*. 3 ed. Munich, Springer, 2007.

[15] DILLON, H. *Hearing Aids*. 2 ed. Turramurra, Boomerang Press, 2012.

[16] WORLD HEALTH ORGANIZATION. "Deafness and Hearing Loss – Fact sheet Nº 300". Available at `https://web.archive.org/web/20150516054114/http://www.who.int/mediacentre/factsheets/fs300/en/`, 2015. Accessed 10/20/2021.

[17] ASHA–THE AMERICAN SPEECH-LANGUAGE-HEARING ASSOCIATION. "Type, Degree, and Configuration of Hearing Loss". Available at `https://www.asha.org/siteassets/uploadedfiles/ais-hearing-loss-types-degree-configuration.pdf`, 2015. Accessed 10/20/2021.

[18] ZAHNERT, T. "The differential diagnosis of hearing loss", *Deutsches Ärzteblatt International*, v. 108, n. 25, pp. 433–444, June 2011.

[19] DANAHER, E. M., PICKETT, J. M. "Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss", *Journal of Speech and Hearing Research*, v. 18, n. 2, pp. 261–271, June 1975.

[20] DANAHER, E. M., PICKETT, J. M. "Sensorineural hearing loss and upward spread of masking", *Journal of Speech and Hearing Research*, v. 13, n. 2, pp. 426–437, June 1970.

[21] CHING, T. Y., DILLON, H., BYRNE, D. "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification", *Journal of the Acoustical Society of America*, v. 103, n. 2, pp. 1128–1140, August 1998.

[22] I DENTURE & HEARING. "Styles and Latest Hearing Aid Models". Available at `https://identureandhearing.com/hearing/super-page.php?id=10`, 2022. Accessed 8/2/2022.

[23] ANSI S3.22. *Specification of Hearing Aid Characteristics*. Norm, American National Standards Institute, 2014.

[24] DILLON, H. *Hearing Aids*. 2 ed. Turramurra, Australia, Boomerang Press, 2012.

[25] TITZE, I. R., BAKEN, R. J., *ET AL.*, K. W. B. "Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization", *Journal of the Acoustical Society of America*, v. 137, n. 5, pp. 3005–3007, May 1955.

[26] BOTELLA, J. "What is an audiogram and how to read it". Available at `https://www.hear.com/resources/all-articles/what-is-audiogram-how-to-read-it/`, 2021. Accessed 7/26/2021.

[27] LYBARGER, S. F. *Method of fitting hearing aids*. Technical report, U. S. Patent and Trademark Office, Washington, DC, 1983.

[28] MCCANDLESS, G. A., LYREGAARD, P. E. "Prescription of gain/output (POGO) for hearing aids", *Hearing Instruments*, v. 34, n. 1, pp. 16–21, January 1983.

[29] SCHWARTZ, D. M., LYREGAARD, P. E., LUNDH, P. "Hearing aid selection for severe-to-profound hearing loss", *The Hearing Journal*, v. 41, n. 2, pp. 13–17, February 1988.

[30] BYRNE, D., TONISSON, W. "Selecting the gain of hearing aids for persons with sensorineural hearing impairments", *Scandinavian Audiology*, v. 5, n. 2, pp. 51–59, April 1976.

[31] BYRNE, D., DILLON, H. "The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid", *Ear and Hearing*, v. 7, n. 4, pp. 257–265, August 1986.

[32] BYRNE, D., PARKINSON, A., NEWALL, P. "Modified Hearing Aid Selection Procedures for Severe/Profound Hearing Losses". In: Studebaker, G. A., Bess, F. H., Beck, L. (Eds.), *The Vanderbilt hearing aid report II*, York Press, pp. 295–300, Parkton, 1991.

[33] CORNELISSE, L., SEEWALD, R., JAMIESON, D. "The input/output formula: a theoretical approach to the fitting of personal amplification devices", *Journal of the Acoustical Society of America*, v. 97, n. 3, pp. 1854–1864, March 1995.

[34] BYRNE, D., DILLON, H., CHING, T., et al. "NAL-NL1 procedure for fitting non-linear hearing aids: Characteristics and comparisons with other procedures", *Journal of the American Academy of Audiology*, v. 12, n. 1, pp. 37–51, January 2001.

[35] KEIDSER, G., DILLON, H., FLAX, M., et al. "The NAL-NL2 prescription procedure", *Audiology Research*, v. 1, n. 1, pp. 88–90, May 2011.

[36] KILLION, M. C., FIKRET-PASA, S. "The 3 types of sensorineural hearing loss: loudness and intelligibility considerations", *The Hearing Journal*, v. 46, n. 11, pp. 31–36, November 1993.

[37] MOORE, B. C. J., ALCANTARA, J. I., STONE, M. A., et al. "Use of a loudness model for hearing aid fitting: II. Hearing aids with multi-channel compression", *British Journal of Audiology*, v. 33, n. 3, pp. 157–170, May 1999.

[38] MOORE, B. C. J. "Use of a loudness model for hearing aid fitting. IV. Fitting hearing aids with multi-channel compression so as to restore 'normal' loudness for speech at different levels", *British Journal of Audiology*, v. 34, n. 3, pp. 165–177, May 2000.

[39] VORLÄNDER, M. “Acoustic Virtual Reality: Lecture Slides”. 2015. Institute for Technical Acoustics, RWTH Aachen University.

[40] VORLÄNDER, M. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Berlin, Springer-Verlag, 2008.

[41] SCHLEICH, P., NOPP, P., D’HAESE, P. “Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant”, *Ear and Hearing*, v. 25, n. 3, pp. 197–204, June 2004.

[42] DOCLO, S., KLASEN, T. J., DEN BOGAERT, T. V., et al. “Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions”. In: *Procedings of 2006 International Workshop on Acoustic Echo and Noise Control (IWAENC 2006)*, pp. 1–4, Paris, September 2006.

[43] STRUTT, J. W. “XII. On our perception of sound direction”, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, v. 13, n. 74, pp. 214–232, February 1907.

[44] HARTMANN, W. M., RAKERD, B., CRAWFORD, Z. D. “Transaural experiments and a revised duplex theory for the localization of low-frequency tones”, *Journal of the Acoustical Society of America*, v. 139, n. 2, pp. 968–985, February 2016.

[45] SANDEL, T. T., TEAS, D. C., FEDDERSEN, W. E., et al. “Localization of sound from single and paired sources”, *Journal of the Acoustical Society of America*, v. 27, n. 5, pp. 842–852, September 1955.

[46] MILLS, A. W. “On the minimum audible angle”, *Journal of the Acoustical Society of America*, v. 30, n. 4, pp. 237–246, April 1958.

[47] AARONSON, N. L., HARTMANN, W. M. “Testing, correcting, and extending the Woodworth model for interaural time difference”, *Journal of the Acoustical Society of America*, v. 135, n. 2, pp. 817–823, February 2014.

[48] ALCAIM, A., SOLEWICZ, J. A., MORAES, J. A. “Frequência de ocorrência dos fones e listas de frases foneticamente balanceadas no português falado no Rio de Janeiro”, *Revista da Sociedade Brasileira de Telecomunicações*, v. 7, n. 1, pp. 23–41, December 1992. In Portuguese.

[49] BISGAARD, N., VLAMING, M. S. M. G., DAHLQUIST, M. "Standard audiograms for the IEC 60118-15 measurement procedure", *Trends in Amplification*, v. 14, n. 2, pp. 113–120, June 2010.

[50] VAIDYANATHAN, P. P. *Multirate Systems and Filter Banks*. Upper Saddle River, Prentice Hall, 1993.

[51] ZÖLZER, U. *DAFx: Digital Audio Effects*. 2 ed. United Kingdom, John Wiley & Sons, 2011.

[52] MCNALLY, G. W. "Dynamic range control of digital audio signals", *Journal of the Audio Engineering Society*, v. 32, n. 5, pp. 316–327, May 1984.

[53] ZÖLZER, U. *Digital Audio Signal Processing*. 2 ed. United Kingdom, John Wiley & Sons, 2005.

[54] FRIEDMAN, M. "The use of ranks to avoid the assumption of normality implicit in the analysis of variance", *Journal of the American Statistical Association*, v. 32, n. 200, pp. 675—701, December 1937.

[55] DEEP, R. *Probability and Statistics*. London, Elsevier, 2006.

[56] BONFERRONI, C. E. "Teoria Statistica delle Classi e Calcolo della Probabilità". 1936. In Italian.

[57] CARLILE, S., P.LEONG, HYAMS, S. "The nature and distribution of errors in sound localization by human listeners", *Hearing Research*, v. 114, n. 1–2, pp. 179–196, December 1997.

[58] MAKOUS, J. C., MIDDLEBROOKS, J. C. "Two-dimensional sound localization by human listeners", *Journal of the Acoustical Society of America*, v. 87, n. 5, pp. 2188–2200, May 1990.

[59] MIRONOVS, M., LEE, H. "On the accuracy and consistency of sound localisation at various azimuth and elevation angles". In: *144th Convention of the Audio Engineering Society*, Milan, May 2018. Audio Engineering Society.

[60] DEPARTMENT OF MEDICAL PHYSICS AND ACOUSTICS, CARL-VON-OSSIETZKY UNIVERSITY OLDENBURG. "Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses". Available at `http://medi.uni-oldenburg.de/hrir/index.html`, 2013. Accessed 10/16/2022.

[61] BATISTA, N. A. R. *Estudo sobre Identificação Automática de Sotaques Regionais Brasileiros Baseada em Modelagens Estatísticas e Técnicas de Aprendizado de Máquina*. Master's thesis, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas (UNICAMP), Campinas, 2019. In Portuguese.

[62] NASCENTES, A. *O Linguajar Carioca*. 2 ed. Rio de Janeiro, Organizações Simões, 1953. First edition published in 1922. In Portuguese.

[63] ROMANO, V. P. "Áreas lexicais no Centro-Sul do Brasil sob uma perspectiva geolinguística", *Revista de Estudos da Linguagem*, v. 26, n. 1, pp. 103–145, January 2018. In Portuguese.

[64] HARSANYI, Z. *A Vida de Galileu: O contemplador de estrelas*. Rio de Janeiro, Editora José Olímpio, 1957. In Portuguese.

[65] SILVA, A. P., NETO, A. F., PIRES, L. S., et al. "Corpus CEFALA1: Primeira Base Bimodal do Laboratório CEFALA". Available at `https://www.cefala.org/arquivos/Protocolo%20de%20Coleta%20Corpus%20CEFALA-1.pdf`, 2017. Accessed 5/18/2020. In Portuguese.

[66] ASHA–THE AMERICAN SPEECH-LANGUAGE-HEARING ASSOCIATION. "Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)–ASHA Special Interest Division 3, Voice and Voice Disorders". Available at `https://www.asha.org/uploadedFiles/members/divs/D3CAPEVprocedures.pdf`, 2002. Accessed 5/18/2020.

[67] PUC-SP–PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO, LABORVOX. "Protocolo - Consenso da Avaliação Perceptivo Auditiva da Voz (CAPE-V) - ASHA 2003, SID3". Available at `http://www4.pucsp.br/laborvox/dicas_pesquisa/downloads/CAPEV.pdf`, 2003. Accessed 5/18/2020. In Portuguese.

[68] RASO, T., MELLO, H. *C-oral-Brasil I: Corpus de Referência do Português Brasileiro Falado Informal*. Belo Horizonte, Brazil, Editora UFMG, 2012. In Portuguese.

[69] RASO, T. *Specifications on the C-ORAL-BRASIL Informal Corpus*. Technical report, Faculdade de Letras, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, 2012.

[70] YNOGUTI, C. A. *Reconhecimento de Fala Contínua Usando Modelos Ocultos de Markov*. PhD thesis, Faculdade de Engenharia Elétrica e de Com-

putação, Universidade Estadual de Campinas (UNICAMP), Campinas, 1999. In Portuguese.

[71] ALENCAR, V. F. S. *Reconhecimento Distribuído de Voz Contínua com Amplo Vocabulário para o Português Brasileiro*. PhD thesis, Programa de Pós-Graduação em Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, 2009. In Portuguese.

[72] CIRIGLIANO, R. J. R., MONTEIRO, C., BARBOSA, F. L. F., et al. "Um conjunto de 1000 frases foneticamente balanceadas para o português brasileiro obtido utilizando a abordagem de algoritmos genéticos". In: *XXII Simpósio Brasileiro de Telecomunicações - SBrT'05*, pp. 544–549, Campinas, September 2005. Sociedade Brasileira de Telecomunicações. In Portuguese.

Appendix A

Additional figures

A.1 Drawings from the Kayser HRIR database

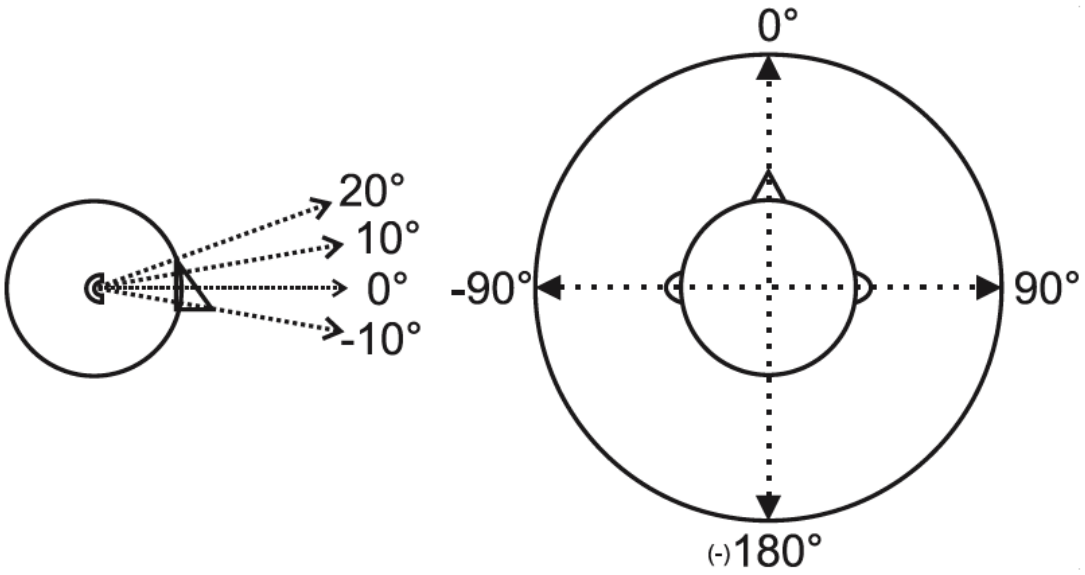


Figure A.1: Coordinate system with azimuth and elevations for HRIRs recorded in the anechoic chamber. Source: Oldenburg University [60].

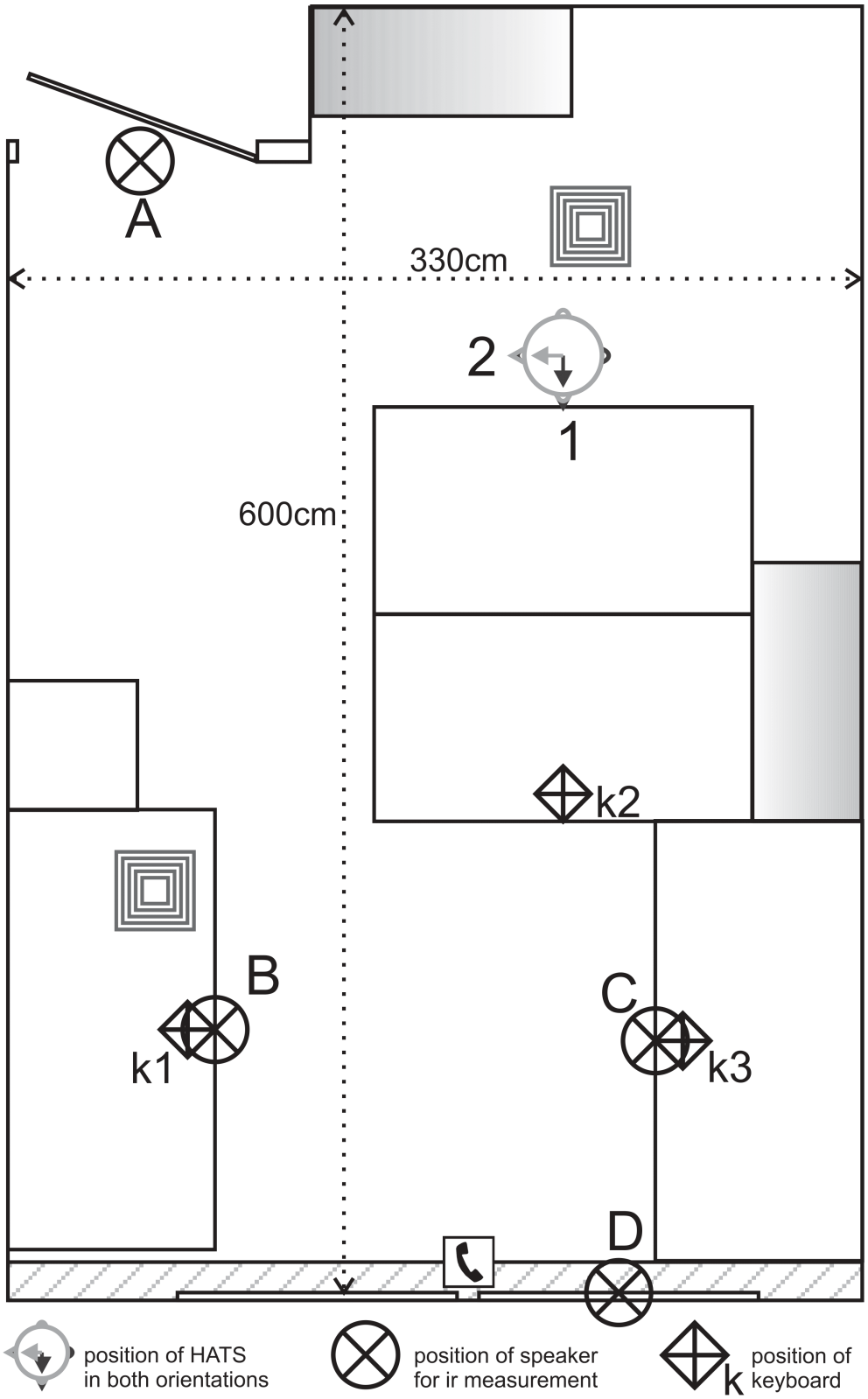


Figure A.2: Floor plant of Office II, with source and HATS positions. Source: Oldenburg University [60].

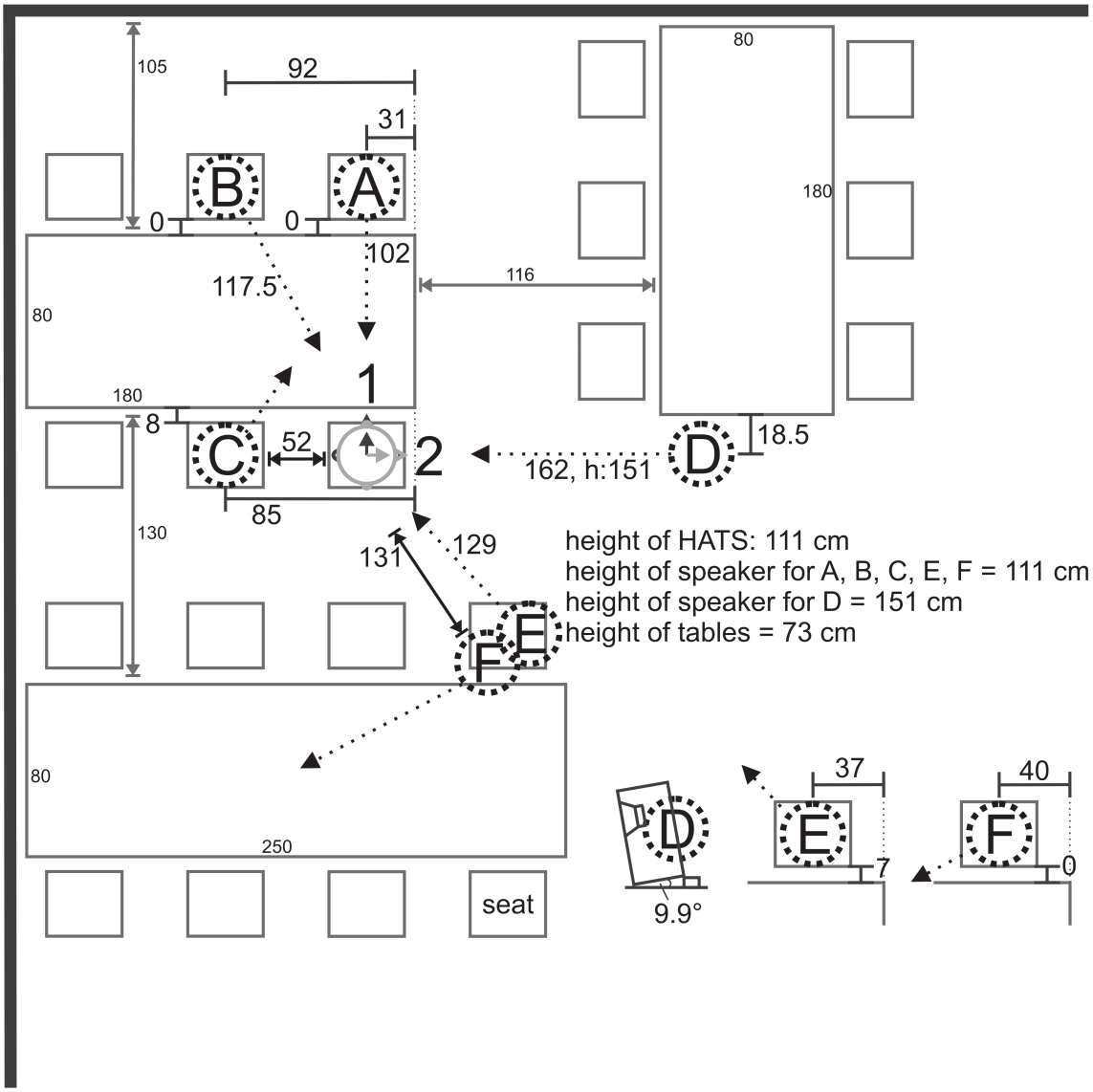


Figure A.3: Floor plant of Cafeteria, with source and HATS positions. Source: Oldenburg University [60].

A.2 Binaural cues for Experiment 1–Analysis 2

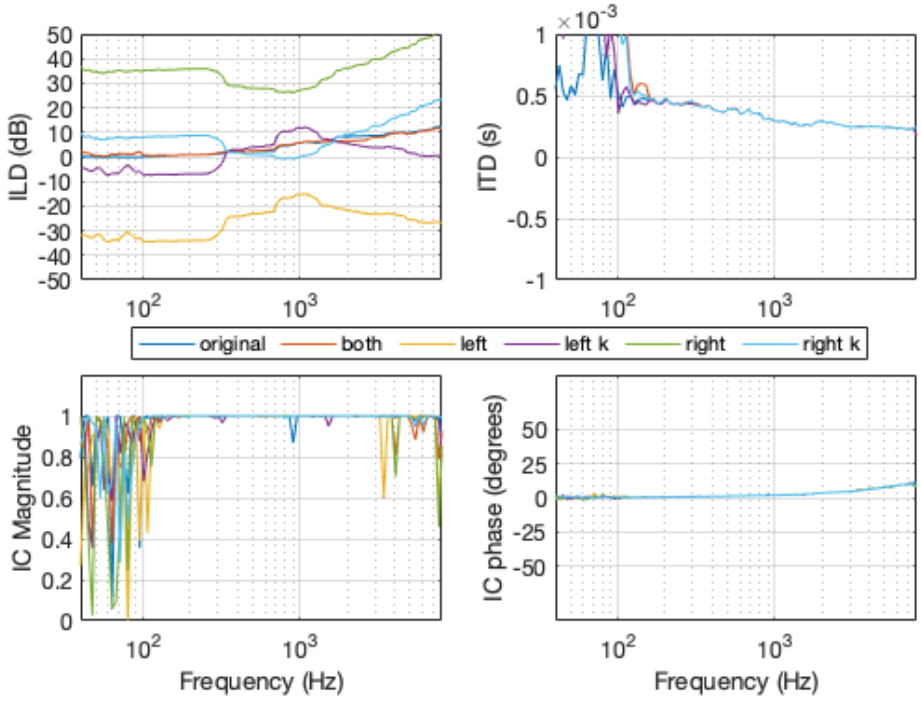


Figure A.4: Binaural cues from signal B with N3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

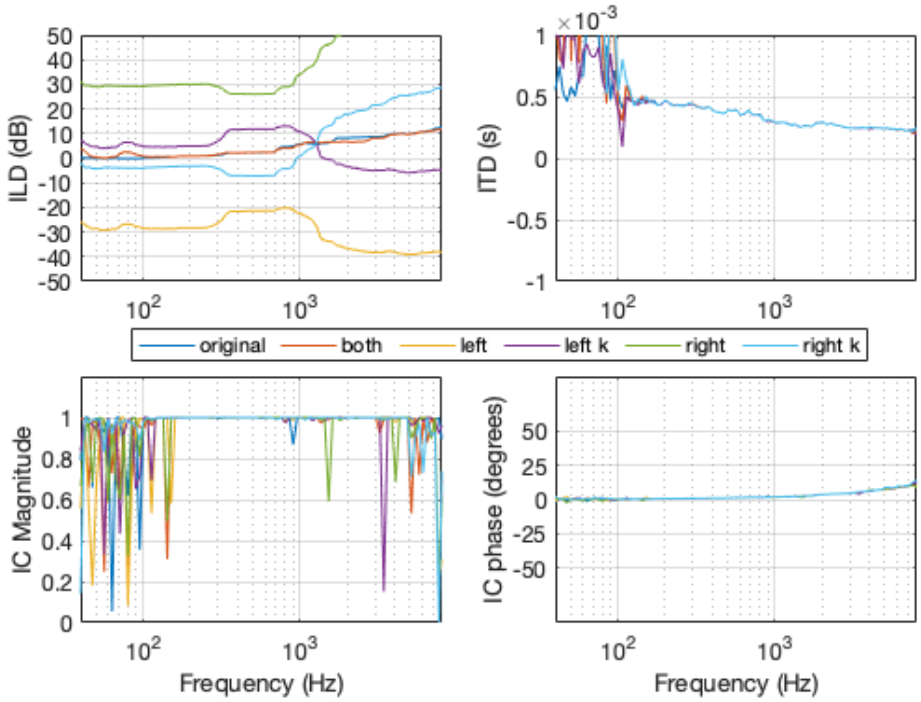


Figure A.5: Binaural cues from signal B with S3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

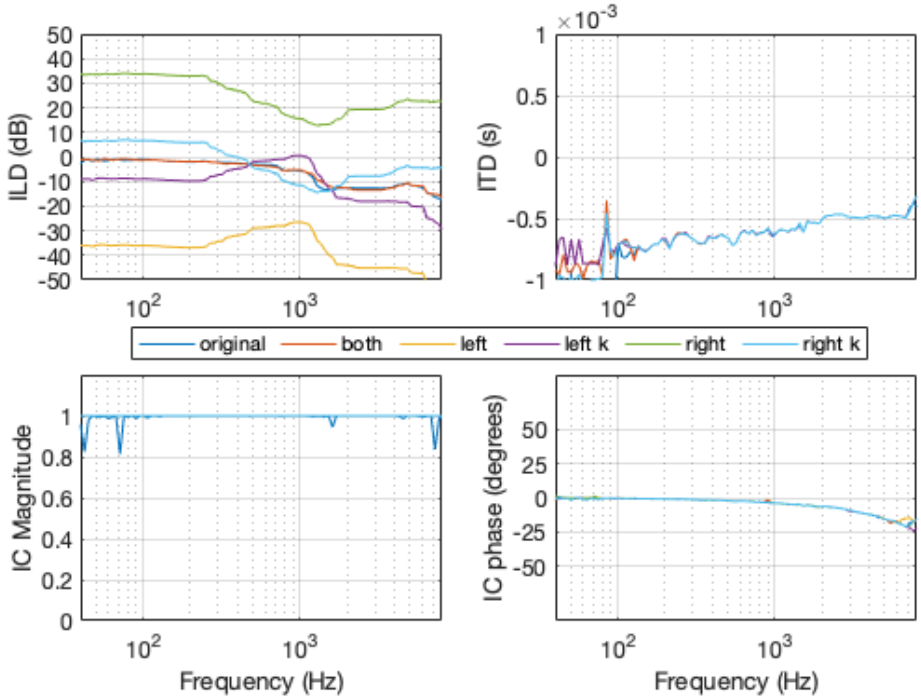


Figure A.6: Binaural cues from signal D with N3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

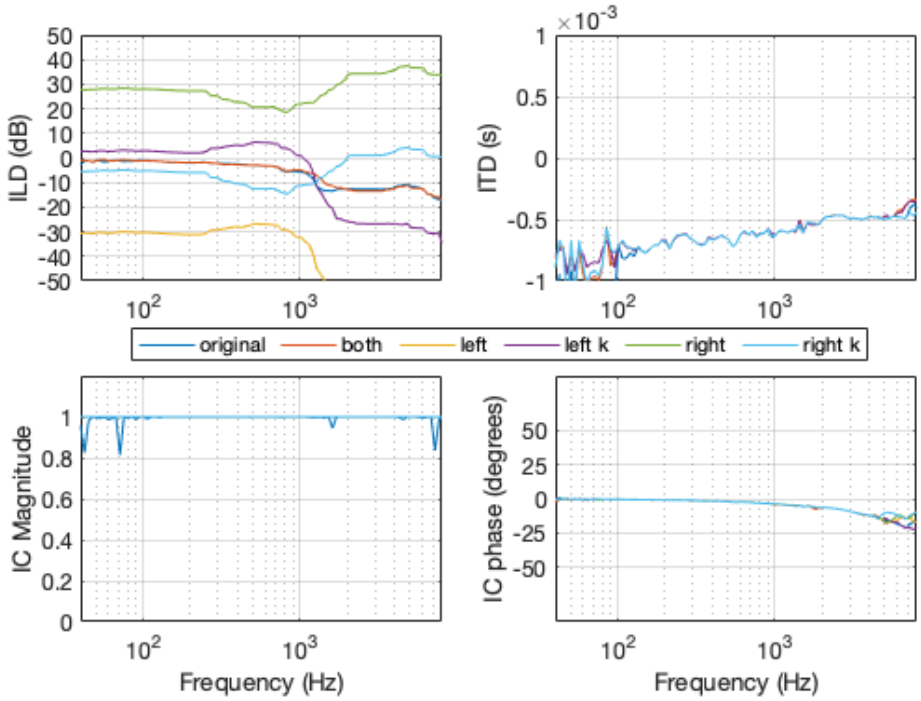


Figure A.7: Binaural cues from signal D with S3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

A.3 Binaural cues for Experiment 1–Analysis 3

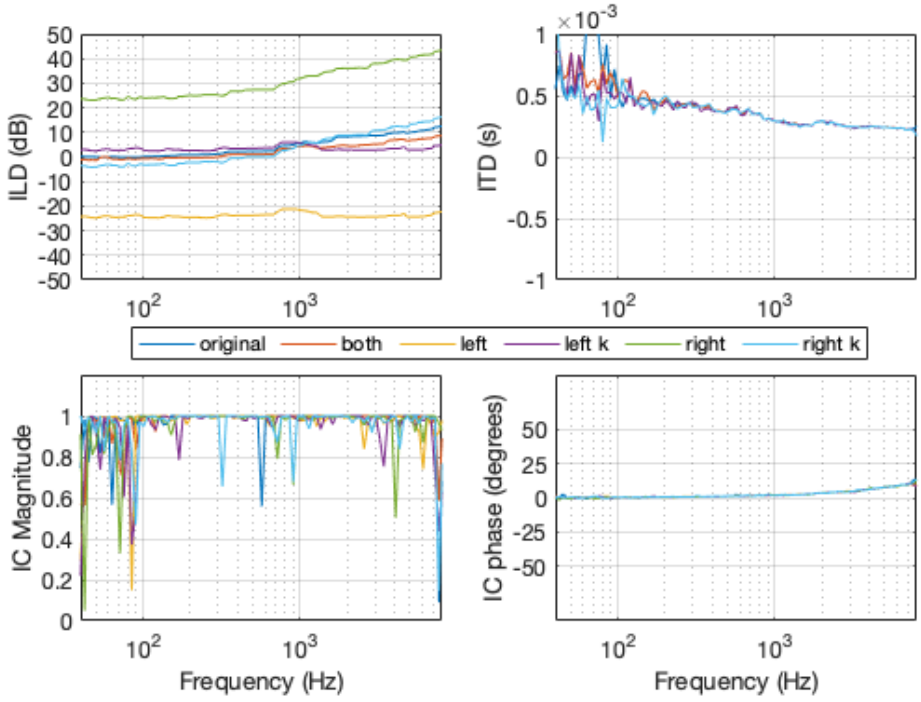


Figure A.8: Binaural cues from signal B with N3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

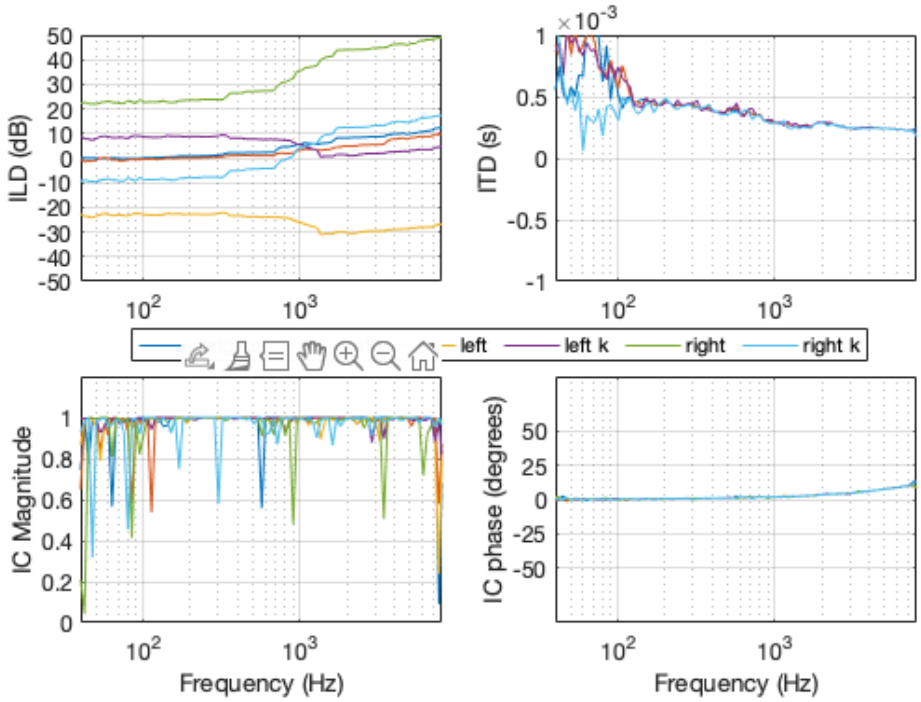


Figure A.9: Binaural cues from signal B with S3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

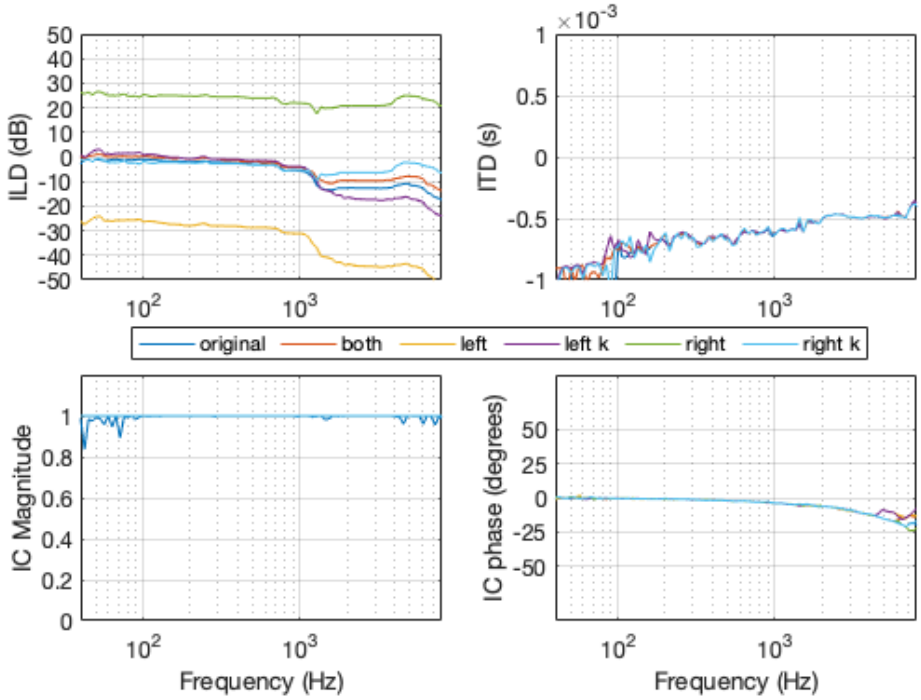


Figure A.10: Binaural cues from signal D with N3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

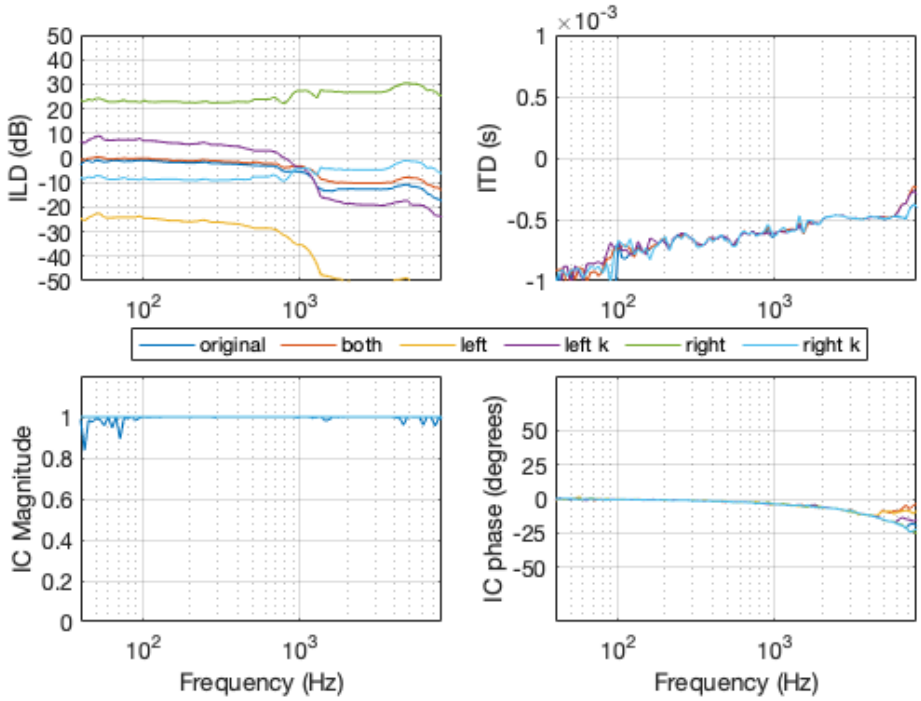


Figure A.11: Binaural cues from signal D with S3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

A.4 Binaural cues for Experiment 2–Analysis 1

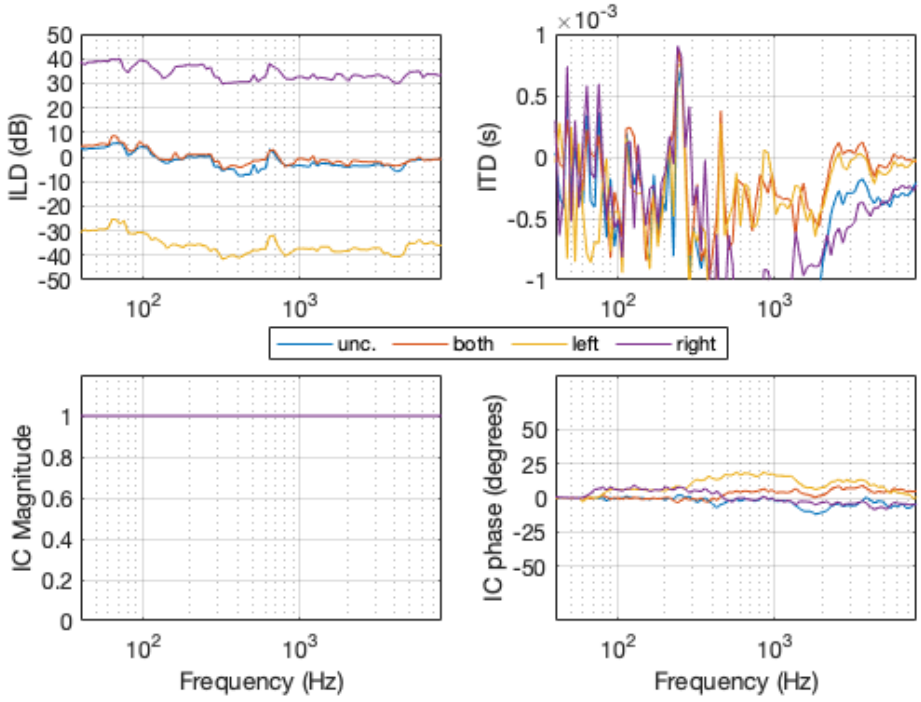


Figure A.12: Signal F13-1B (Office II) uncompressed and with compression applied at both ears, only at the left ear, and only at the right ear.

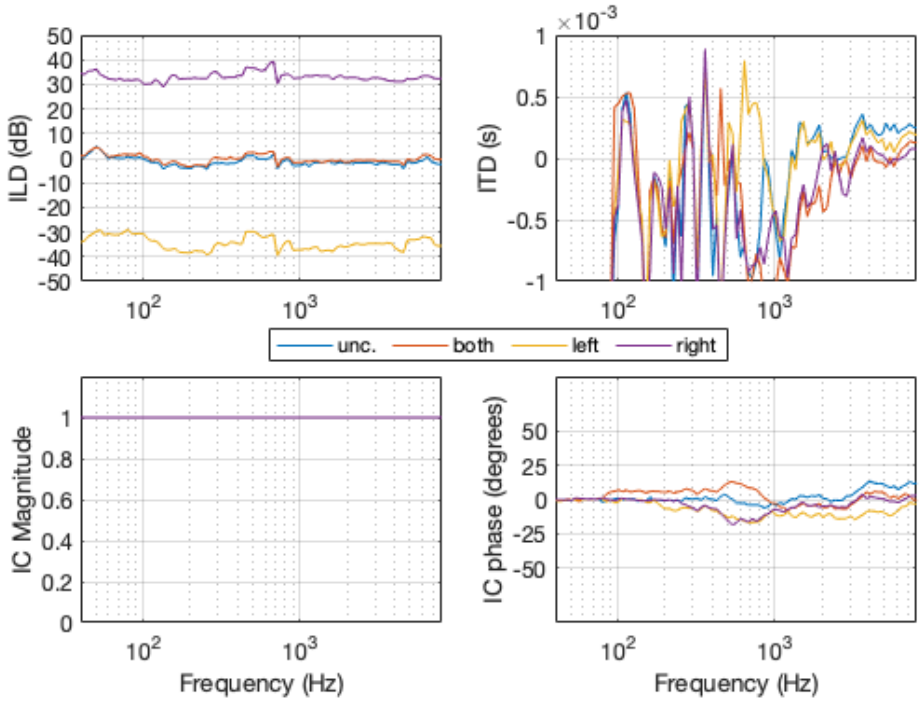


Figure A.13: Signal F13-1D (Office II) uncompressed and with compression applied at both ears, only at the left ear, and only at the right ear.

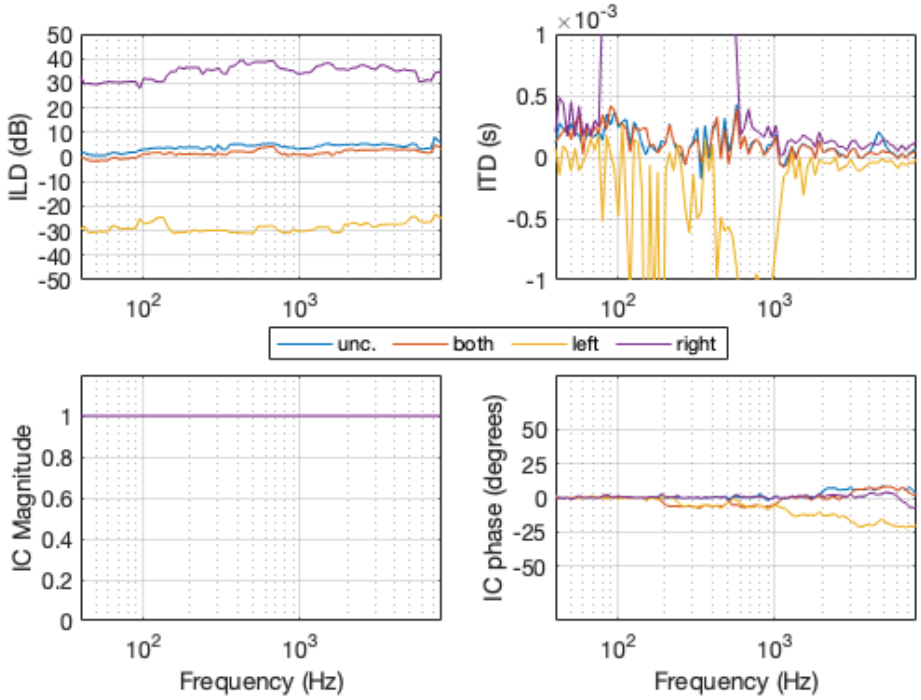


Figure A.14: Signal F13-2C (Cafeteria) uncompressed and with compression applied at both ears, only at the left ear, and only at the right ear.

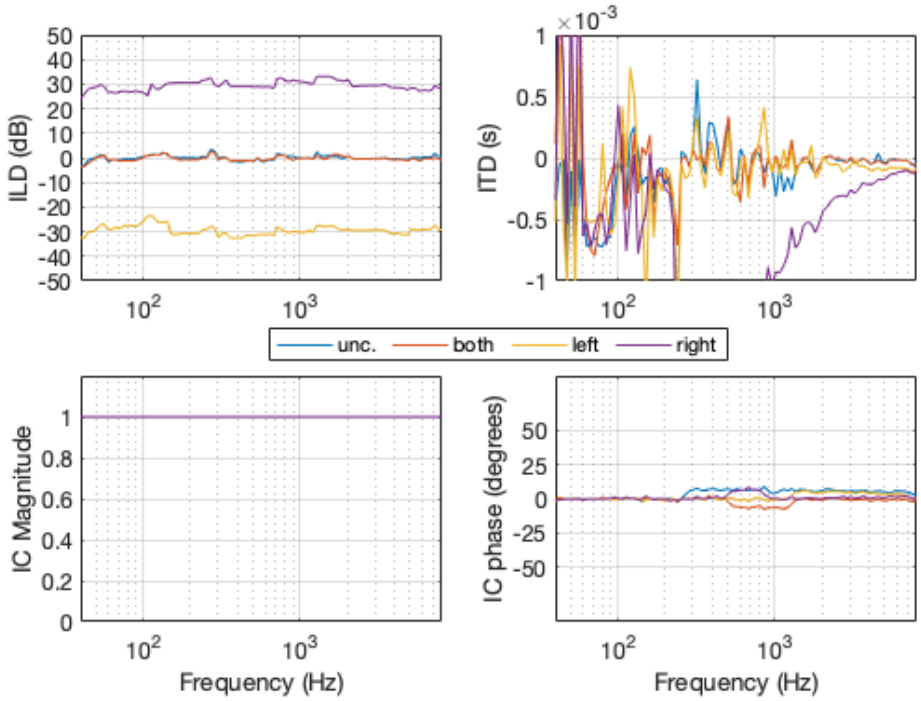


Figure A.15: Signal F13-2D (Cafeteria) uncompressed and with compression applied at both ears, only at the left ear, and only at the right ear.

A.5 Binaural cues for Experiment 2–Analysis 2

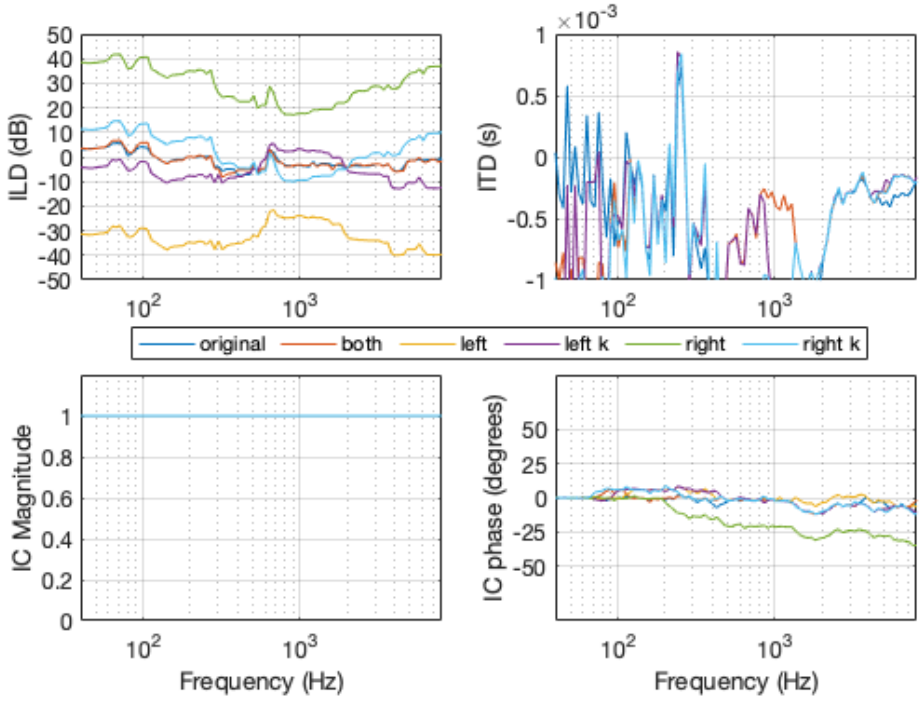


Figure A.16: Binaural cues from signal F13-1B (Office II) with N3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

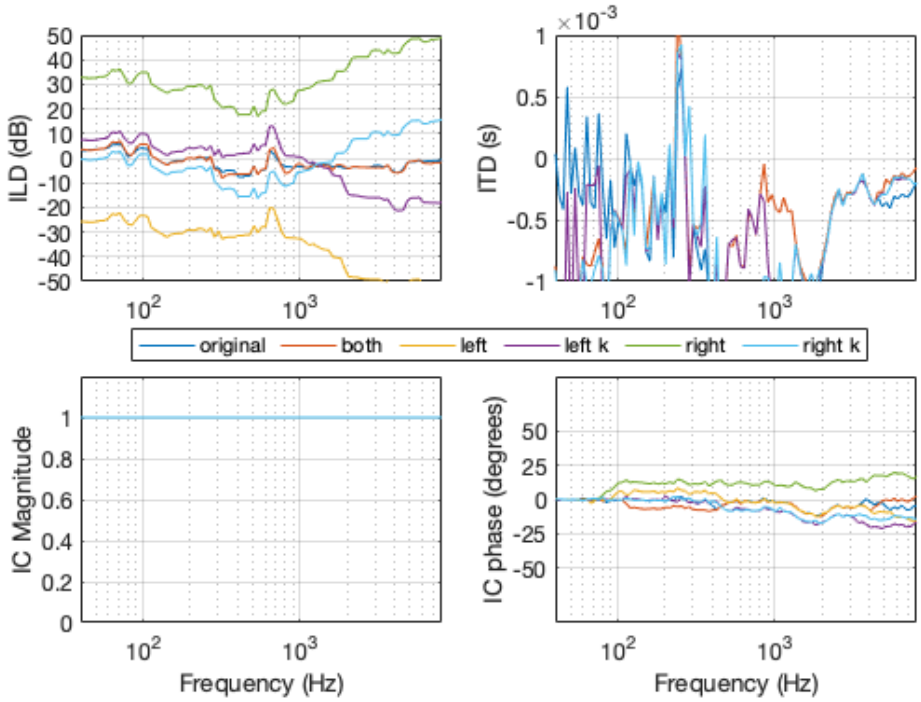


Figure A.17: Binaural cues from signal F13-1B (Office II) with S3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

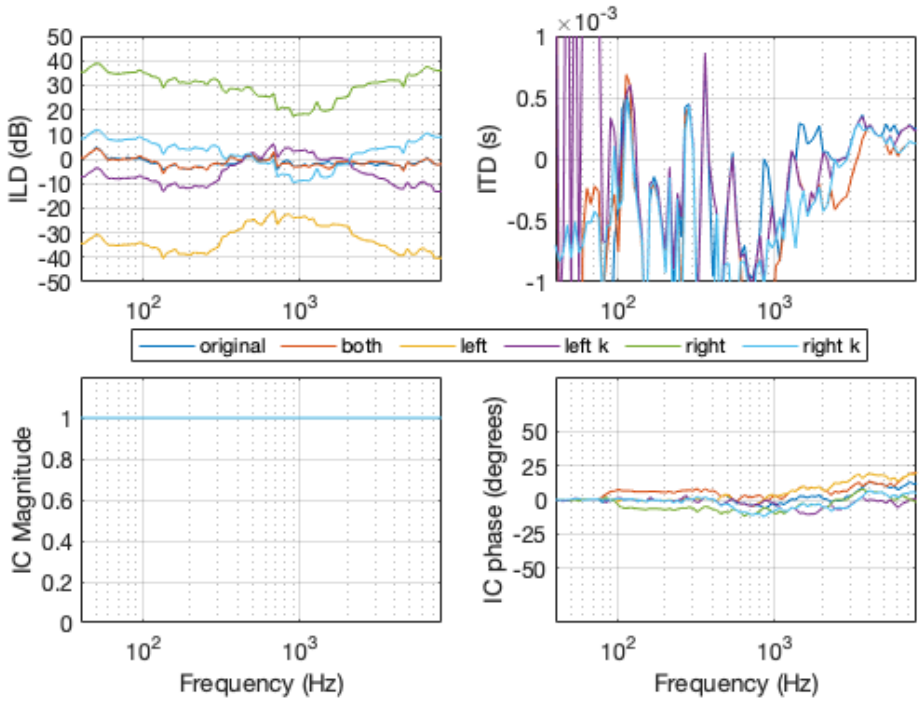


Figure A.18: Binaural cues from signal F13-1D (Office II) with N3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

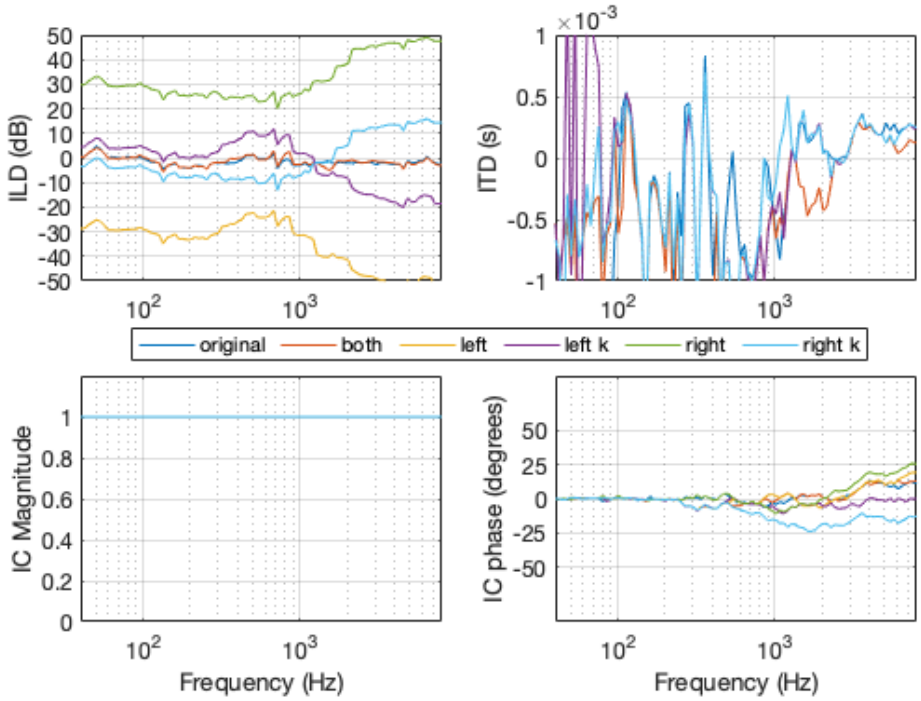


Figure A.19: Binaural cues from signal F13-1D (Office II) with S3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

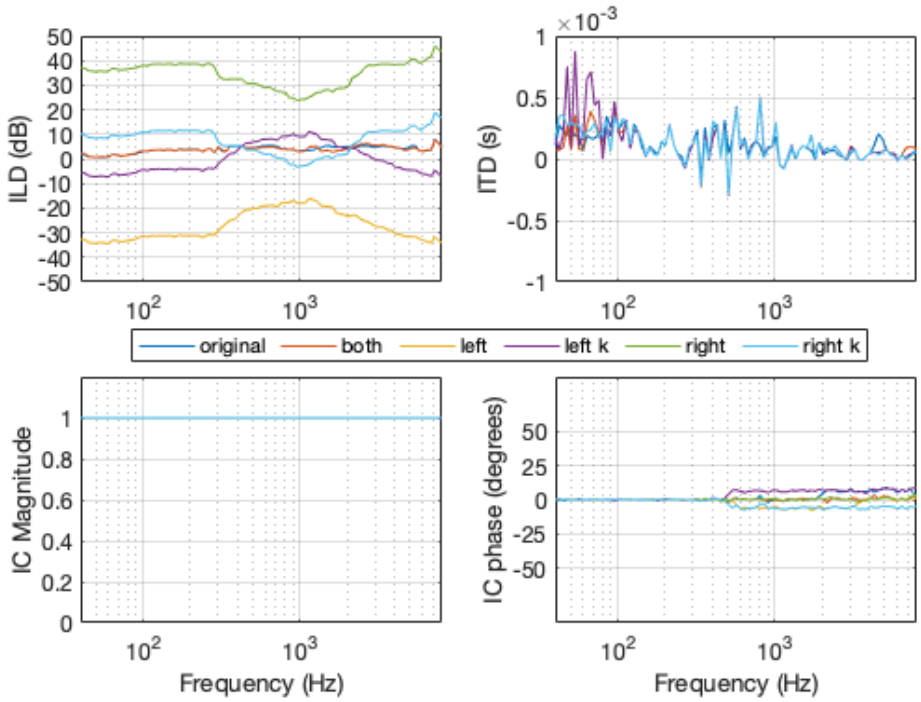


Figure A.20: Binaural cues from signal F13-2C (Cafeteria) with N3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

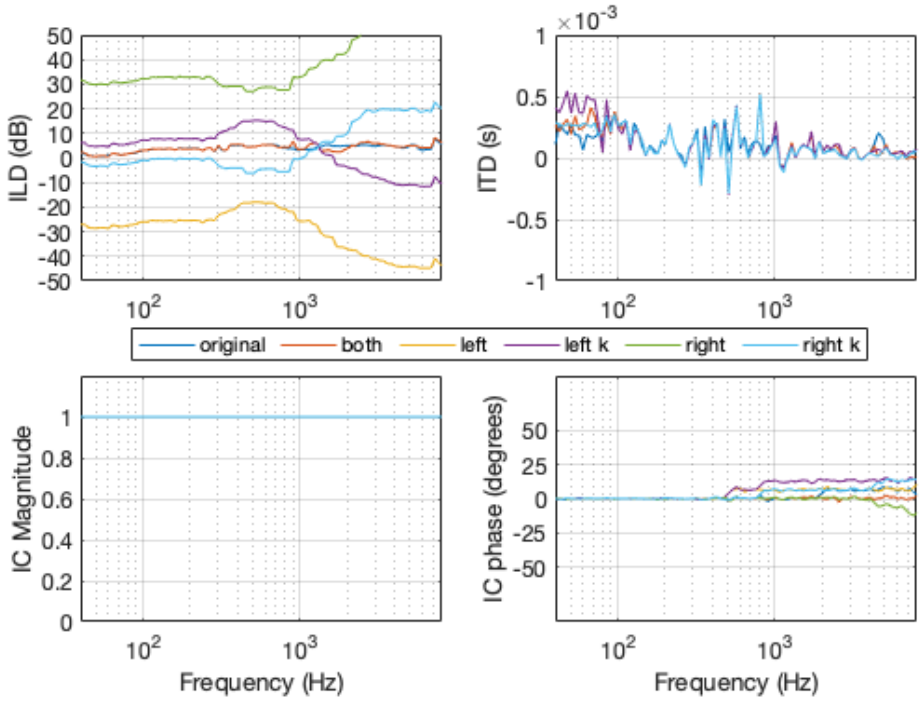


Figure A.21: Binaural cues from signal F13-2C (Cafeteria) with S3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

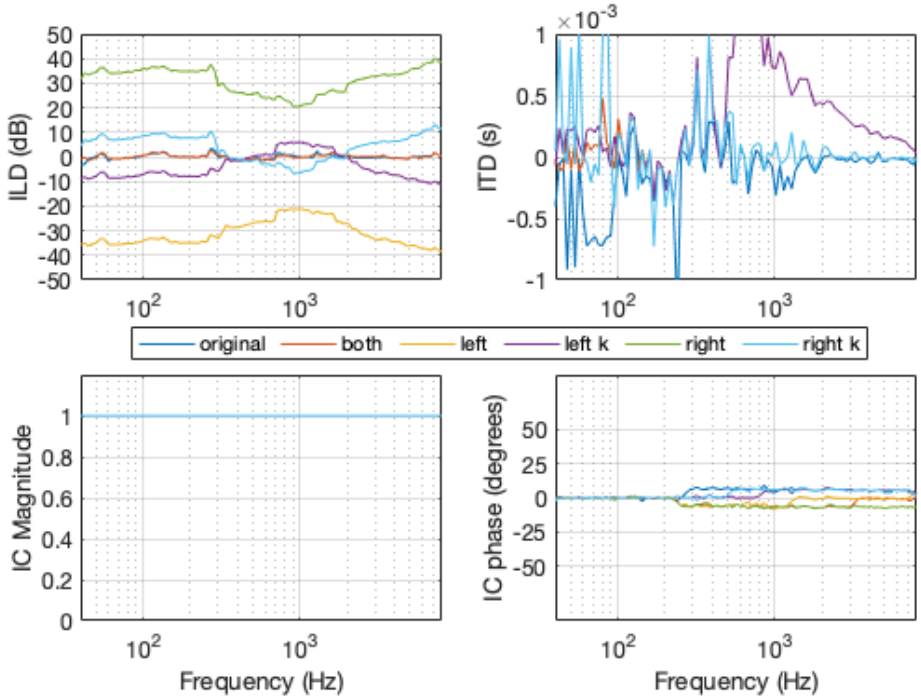


Figure A.22: Binaural cues from signal F13-2D (Cafeteria) with N3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

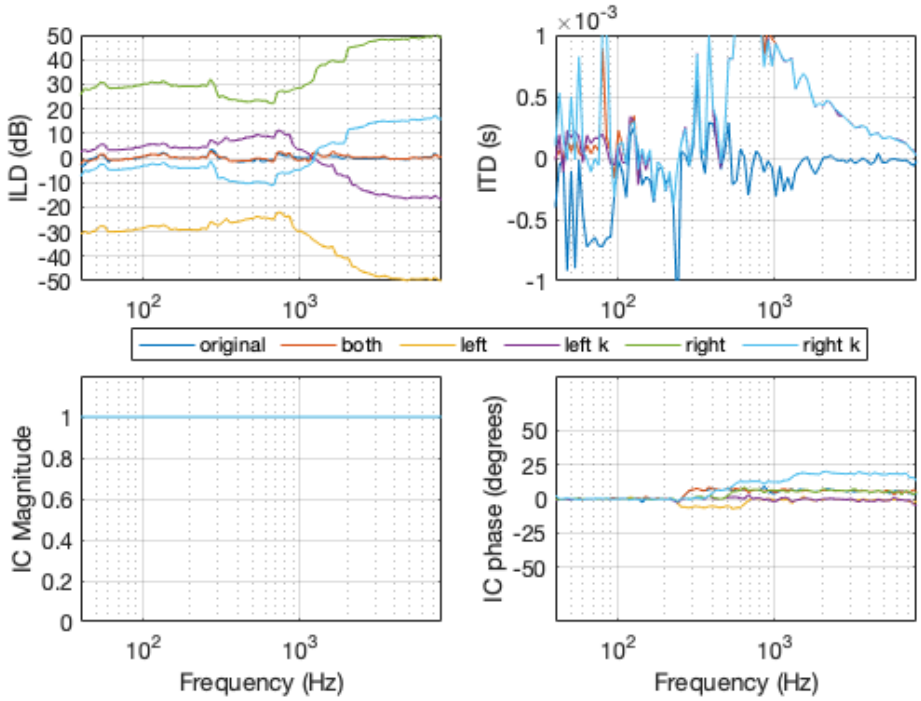


Figure A.23: Binaural cues from signal F13-2D (Cafeteria) with S3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

A.6 Binaural cues for Experiment 2–Analysis 3

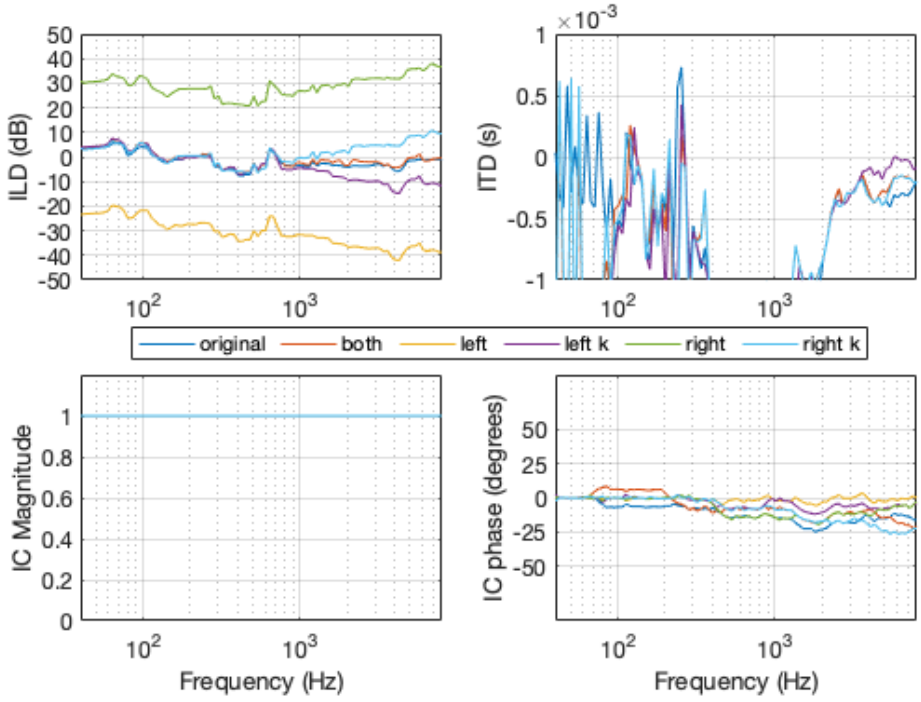


Figure A.24: Binaural cues from signal F13-1B (Office II) with N3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

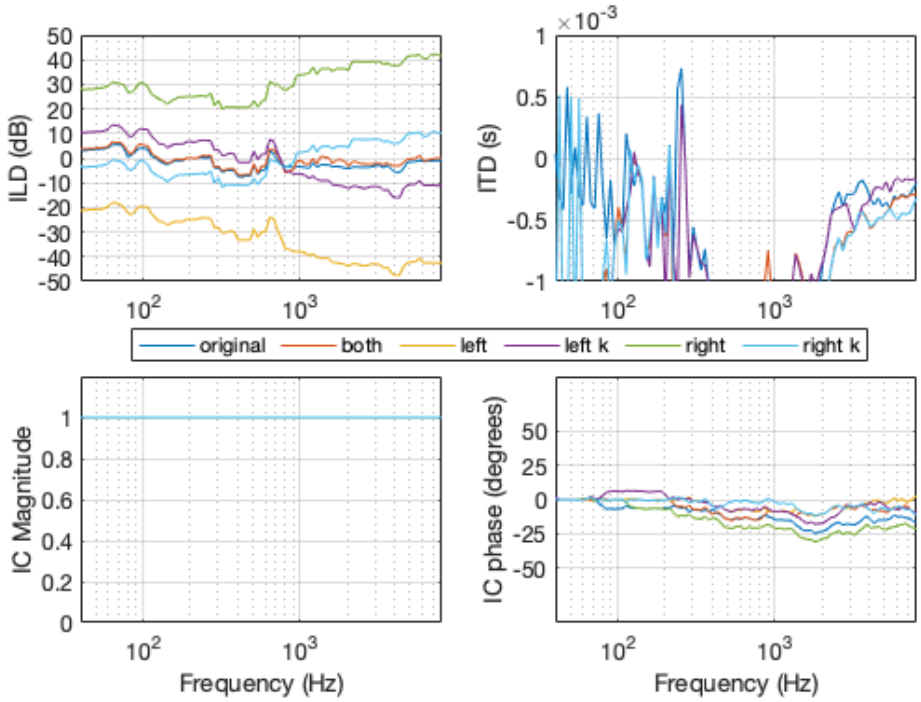


Figure A.25: Binaural cues from signal F13-1B (Office II) with S3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

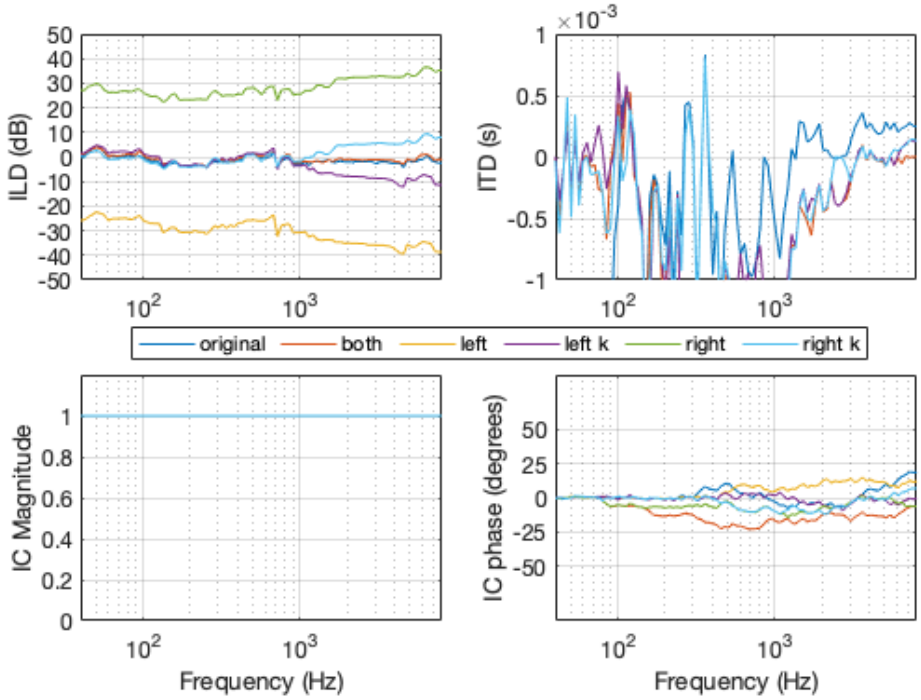


Figure A.26: Binaural cues from signal F13-1D (Office II) with N3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

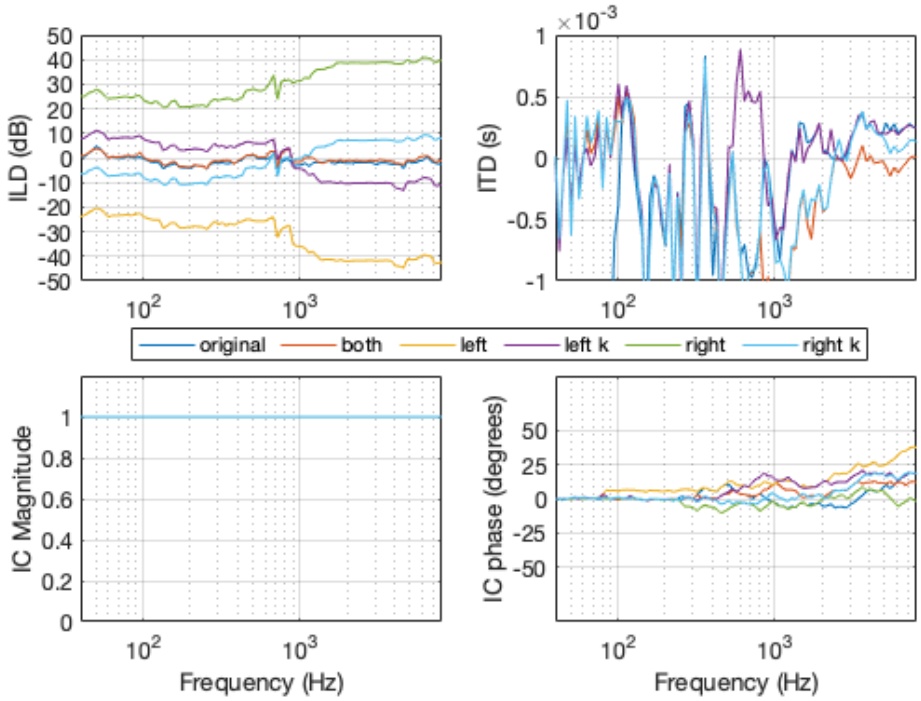


Figure A.27: Binaural cues from signal F13-1D (Office II) with S3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

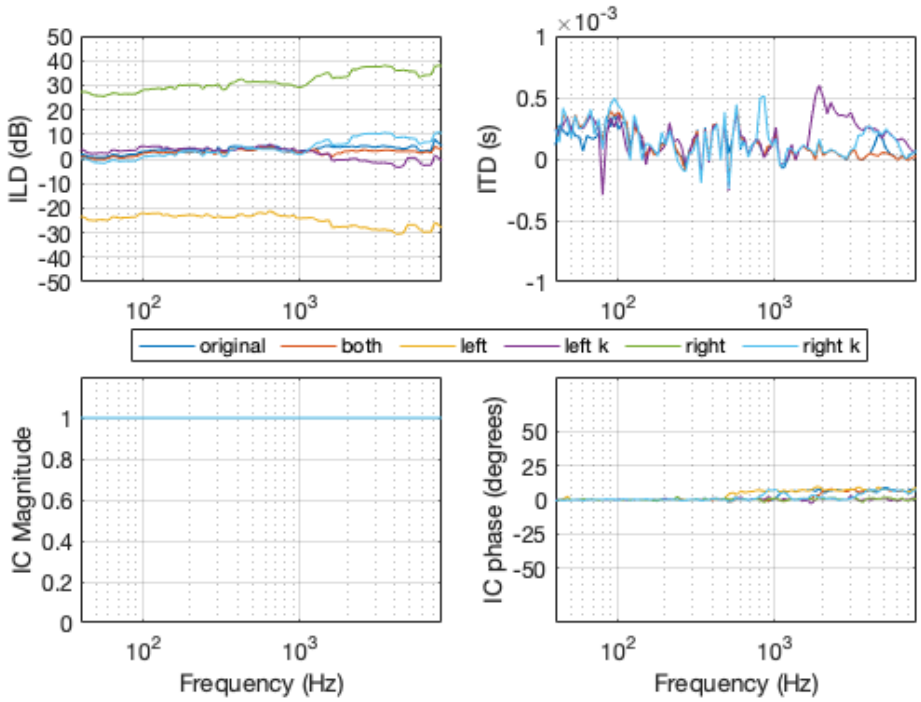


Figure A.28: Binaural cues from signal F13-2C (Cafeteria) with N3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

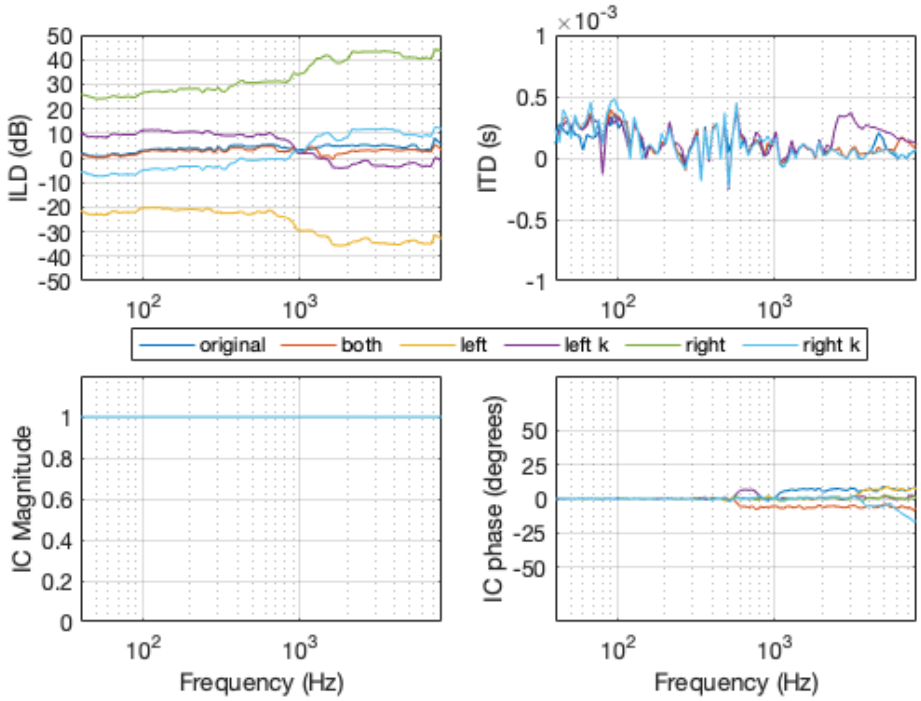


Figure A.29: Binaural cues from signal F13-2C (Cafeteria) with S3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

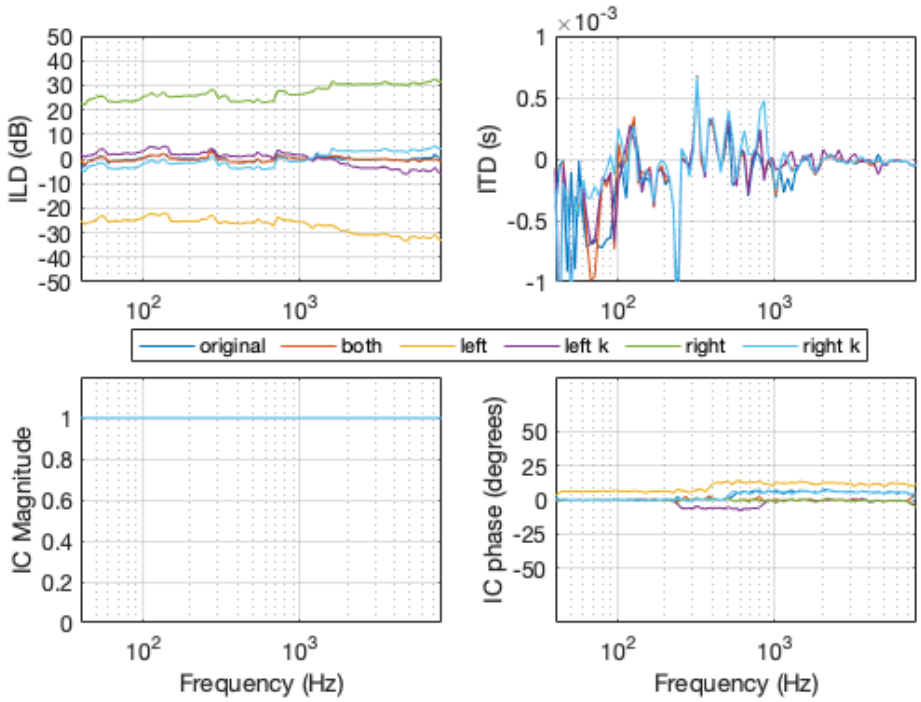


Figure A.30: Binaural cues from signal F13-2D (Cafeteria) with N3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

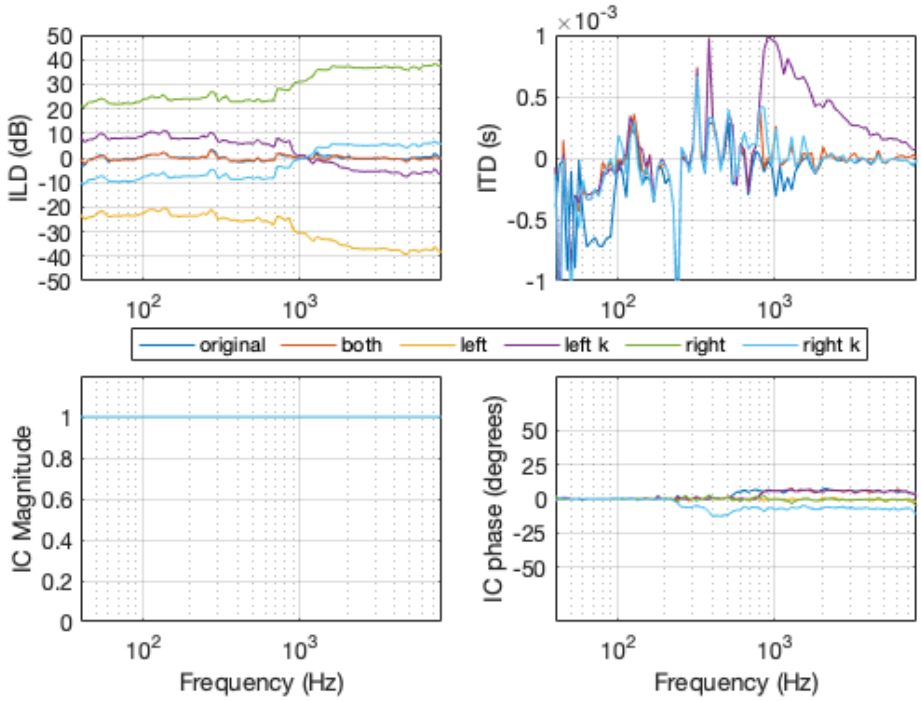


Figure A.31: Binaural cues from signal F13-2D (Cafeteria) with S3 audiogram. Cues for unilateral losses are plotted with and without equalization by constant k.

Appendix B

Brazilian Portuguese speech databases

B.1 Braccent

The Braccent database was built by Batista [61] and encompasses multiple accents of Brazilian Portuguese. The accents are divided into subgroups, which are named after the accent subgroups described in the work of Nascentes [62]. These are: *baiano* (typical from the Brazilian state of Bahia – here indicated as "BA"), *carioca* (from the city of Rio de Janeiro – "CA"), *fluminense* (from the state of Rio de Janeiro outside the city of Rio de Janeiro – "FL"), *mineiro* (from the state of Minas Gerais – "MI"), northern ("NO"), northeastern ("NE") and southern ("SO"). Figure B.1 illustrates the geographical location where these accent subgroups are present.

The database has male and female speakers of different educational levels and currently contains approximately 142 speakers and 1750 audio tracks. Each speaker reads a list of sixteen phonetically balanced short paragraphs, which is presented in Batista [61]. There is one short paragraph in each audio track, so that each speaker produces up to sixteen audio tracks.

The audio tracks were recorded as stereo with either a 48 kHz or 44.1 kHz sampling rate. All tracks were then converted to mono 16 kHz@16-bit PCM Wave files. Since the recordings were made in several non-controlled environments through a web application, by built-in microphones of computers and smartphones, they already contain natural reverberation and background noise from the environment itself. Detailed data about the number of recordings divided by accent subgroup and gender are shown in Table B.1.

Figure B.1: Accent subgroups as organized by Nascentes [62]. Dash-dot line indicates international borders; dashed lines indicate state borders; solid lines indicate accent subgroup borders. Source: Nascentes [62] *apud* Romano [63].

Table B.1: Number of recordings of the Braccent database divided by accent subgroup and gender.

	Number of recordings		
Accents	Female	Male	Total
BA	102	80	182
CA	47	35	82
FL	63	51	114
MI	63	85	148
NE	161	198	359
NO	8	19	27
SO	436	410	846
Total	880	878	1758

B.2 CEFALA-1

The CEFALA-1 *corpus* was developed by the Center of Speech, Acoustics, Language and Music Studies (CEFALA) from the Federal University of Minas Gerais (UFMG). It has 104 speakers, of whom 49 are female and 55 are male. There is no detailed

information about the speakers and where they come from, but it is known that the great majority is formed by UFMG students and professors born and raised in the state of Minas Gerais. There is one audio track for each speaker, each of them with the following structure: i) approximately 2 minutes of spontaneous speech (e. g. personal or non-personal narrative, comment on any topic, description of a daily activity); ii) reading of a 153-word excerpt from the book "A Vida de Galileu: O contemplador de estrelas" [1] [64], with about 60 s duration; iii) final part of about 50 s with the reading of 20 sentences. The sentence list can be seen in the instruction pages available in Portuguese in Silva *et al.* [65]. It mainly contains sentences belonging to Alcaim *et al.*'s phonetically balanced sentence lists [48] and commonly used phrases from the Brazilian Portuguese versions of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) [66, 67].

The speech signals were recorded in an acoustically treated professional studio and captured simultaneously by the following five microphones:

- M1: DYLAN lavalier microphone, model DL-09, positioned on the speaker's chest approximately 20 cm from the lips;
- M2: STANER wireless microphone, model SW-481, 2 meters away from the speaker to capture ambient noise;
- M3: Brüel & Kjaer condenser microphone, model 1065, positioned 15 cm away from the speaker at the height of the speaker's lips, 45^o to the right of the sagittal plane;
- M4: Samsung Galaxy S2 Lite GT-i9070 smartphone, capturing ambient noise, simulating audio acquisition from a mobile device–positioned 1 meter away from the speaker, 70 cm above the floor;
- M5: GoPro Hero 3+ Black Edition camera, used for video and audio recording, positioned in front of the speaker at a distance of 1 meter.

Therefore, five audio files were generated for each speaker. They all have the same spoken content, but different quality and SNRs, due to the characteristics of each microphone. All files are in mono 44.1 kHz@16-bit PCM Wave format.

B.3 CFPB

The Forensic Corpus of Brazilian Portuguese (CFPB–Corpus Forense do Português Brasileiro, in Portuguese) was provided by the Federal Police's National Criminalistics Institute (Instituto Nacional de Criminalística da Polícia Federal). The database

[1] by Hungarian writer Zsolt Harsanyi, translated to Portuguese.

contains 206 spontaneous speech recordings and 205 sentence reading recordings[2] from 206 male speakers. The spontaneous speech part is always about some topic previously brought up by the interviewer, who does not speak during the recording. On the other hand, most of the reading part's content is made of commonly used phrases from the Brazilian Portuguese versions of CAPE-V [66, 67] and sentences taken from Alcaim *et al.*'s phonetically balanced sentence lists [48]. The speakers come from many different regions of Brazil and the *corpus* and all of Nascentes's accent subgroups [62] are present in this corpus.

Recordings were made in office rooms with no acoustic treatment, with a Shure SM58 microphone, an Edirol UA25 or UA25E audio interface, and Adobe Audition 3.0 software. Audio tracks were recorded as stereo with 44.1 kHz sampling rate and were then converted to mono 22.05 kHz@16-bit PCM Wave (spontaneous speech files) and to mono 8 kHz@16-bit PCM Wave (sentence reading files) formats.

B.4 C-ORAL-BRASIL-I

C-ORAL-BRASIL-I is a vast database of informal spontaneous speech in Brazilian Portuguese, developed under the coordination of Prof. Tommaso Raso (Federal University of Minas Gerais) [68]. It has an analogous structure to that of the C-ORAL-ROM database, which features corpora for several Romance languages (French, Italian, European Portuguese and Spanish).

Given the focus on spontaneous and informal speech, C-ORAL-BRASIL-I has audio samples recorded in a wide variety of public and private environments. The corpus contains monologues (produced by 1 speaker), dialogues (featuring 2 speakers) and conversations (more than 2 speakers). It presents 362 speakers (158 male and 203 female), of diverse age groups (ranging from 18 to over 60 years old), educational levels, and occupations, from several geographical locations. It must be noted, however, that more than 60% of the speakers come from the state of Minas Gerais, and roughly 32% of all speakers have at least one non-informed feature (sex, age, level of education, or geographical origin).

The speech signals have variable audio quality due to the distinct acoustical conditions of each type of environment. Examples of recording locations are acoustical cabin, cafeteria, supermarket, domestic environment, school library, and the interior of a car. The audio files are in either mono or stereo 22.05 kHz@16-bit PCM Wave format. They follow the given recording processes, as described by Raso [69]:

a) *dialogues* and *monologues*: 44.1 kHz stereo or mono digital recording with unidirectional microphones, converted to mono or stereo 22.05 kHz@16-bit

[2]One of the reading recordings was clipped and is therefore not used in this work.

PCM Wave files;

b) *conversations*: mono recording with omnidirectional microphone in mono audio files or up to eight channel stereo recording with unidirectional microphones downmixed to stereo audio files. Recordings done with 44.1 kHz sampling rate and converted to 22.05 kHz@16-bit PCM Wave files.

More detailed information about the number of recordings and the speakers' origins are given in Tables B.2 and B.3, respectively.

Table B.2: Number of recordings of the C-ORAL-BRASIL-I database divided by environment type and recording type.

Environment type	Recording type	Number of recordings
family/private	monologue	36
	dialogue	35
	conversation	34
public	monologue	14
	dialogue	11
	conversation	9
	Total	139

Table B.3: Speakers of the C-ORAL-BRASIL-I database divided by origin.

Origin	Number of speakers
Belo Horizonte	138
Other cities of Minas Gerais	89
Other Brazilian states	19
Other countries	2
Unknown	114
Total	362

B.5 SMT

This database was created by the Audio Processing Group (GPA) from UFRJ's Signals, Multimedia and Telecommunications Laboratory (SMT), motivated by a project in collaboration with HP Labs which took place between 2010 and 2013. The corpus's purpose was to contribute to the development of automatic tools for quality assessment, with and without reference, of acoustically degraded full-band audio signals. Therefore, a phonetically balanced database with clean signals was needed.

The chosen material for the database were the 20 phonetically balanced sentence lists presented in Alcaim *et al.* [48]. Each list contains 10 sentences, so that each

speaker reads all 200 sentences, generating one sentence per file. The sentences are grouped according to the respective list to which they belong. The database has 2 female and 2 male speakers, all four of them from the city of Rio de Janeiro.

The recordings were done in a professional studio with a high-quality microphone placed about 0.3 m away from the acoustical source. Files were stored as mono 48 kHz@24-bit PCM Wave format.

B.6 Ynoguti

The database developed by Ynoguti [70] contains 71 speakers of both genders. They read several phonetically balanced sentences of approximately 3 s duration, and each audio file contains one sentence. Considering an average of 30 sentences per speaker, there are approximately 30 audio tracks recorded by each speaker.

The speech signals were recorded in a relatively quiet and non-reverberant environment with a good quality directional microphone. The database has male representatives of the BA, FL, MI, NE and SO accent subgroups, and female representatives of the BA, MI and SO subgroups. All recordings are in 16-bit PCM Wave format, and each recording has a 22.05 kHz and a 16 kHz version. Detailed information about the number of recordings divided by accent subgroup and gender is shown in Table B.4 (only the 22.05 kHz versions are counted, so that half of the total number of files is presented).

Table B.4: Number of recordings of the Ynoguti database divided by accent subgroup and gender.

	Number of recordings		
Accents	Female	Male	Total
BA	26	25	51
FL	0	114	114
MI	52	155	207
NE	0	81	81
SO	519	1020	1539
Total	597	1395	1992

B.7 Alcaim-Alencar

The database named as 'Alcaim-Alencar' is a spoken corpus of Brazilian Portuguese developed in Alencar [71] for speaker-independent continuous voice recognition systems for Brazilian Portuguese with wide vocabulary. It uses a phonetically balanced list of 1000 sentences generated by genetic algorithms in Cirigliano *et al.* [72], based on a written corpus of a Brazilian newspaper.

The Alcaim-Alencar database contains 100 speakers (50 women and 50 men), and each speaker reads all 1000 sentences, therefore generating 1000 separate audio files. The speakers' age range from 17 to 65 years old, some of them being voice professionals, others trained for audiobook recordings, and others lacking any theoretical or practical experience in voice recording. There is no information about the speakers' geographical origins.

The recordings were done in an acoustically treated professional studio and their specifications were chosen to match that required by voice encoders used in mobile and VoIP networks. Therefore, they are stored as 16 kHz@16-bit PCM Wave files and have a 50-7000 Hz bandwidth.