



Aalborg University
Department of Electronics Systems

Binaural cues for monaural listeners

10th Semester - Master Science in Acoustics

Group 1066:

Unai Martínez de Estíbariz Cerdán
Daniel Fernández Álvarez

Supervisor:

Rodrigo Ordoñez

Abstract:**Title:**

Binaural cues for monaural listeners

Project period:

ACO10, spring semester 2009

Project group:

1066

Participants:Unai Mtz. de Estíbariz Cerdán
Daniel Fernández Álvarez**Supervisor:**

Rodrigo Ordoñez

Copies: 2**Number of pages:** 71**Enclosures:** CD-ROM**Date of completion:** 03/06/2009

Normal hearing people possess the ability to focus their attention into a specific source in noisy environments, which is known as cocktail-party effect. The auditory system compares interaural differences between ears to allow this skill. Therefore monaural people are not able to perform such processing.

Consequently, the present research focuses on manners to aid monaural listeners for improving their speech intelligibility in such environments through the use of binaural cues. The idea of this research is to study the concept in healthy listeners (availability matter), with a view to further on evaluate it in actual monaural listeners if improvements are drawn.

Two approaches to accost this hearing impairment have been developed. The first one is based on the selection of the channel with higher SNR channel (left or right) to be sent to the unimpaired ear, taking advantage of the head shadow effect. The second approach is based on lateral noise suppression, inspired by the work of Dr. Kollmeier.

The proposals are off-line evaluated by means of a listening test. Nine scenarios grouped depending on the number of simultaneous maskers (single, double or multiple) were designed for this evaluation. These scenarios aimed to resemble realistic and representative situations.

The statistical analysis of the results showed a significant improvement of the first approach. However, the second proposal did not yield any improvement, but no significant deterioration over the unaided monaural condition was found. These findings suggest the possibility to test the first proposal into actual monaural people and to implement it in real time.

The content of this report is freely available, but publication (with reference source) may only be pursued due to agreement with the respective authors.

Preface

This report is written by project group 1066 at the "Department of Electronics Systems", run by the "Study board for Electronics and Information Technology" at Aalborg University during the 4th semester at the Acoustics specialisation in the period spanning from February 1st to June 3rd, 2009.

The project "Binaural cues for monaural listeners" is proposed by group 08gr962.

The report consists of 7 chapters. Chapter 1 is an introduction to the topic, and background theory is presented in chapter 2. In chapter 3 the proposed approaches are explained. Chapter 4 and 5 deal with the design of the listening test and the evaluation of results respectively. Chapter 6 presents the conclusion of the current research and in chapter 7 possible future works are proposed. Last part of the report contains appendixes which are meant to help the overall understanding of the study.

The reader should notice that the "Harvard" method is used for citation. The bibliography can be found on page 55. A CD-Rom is enclosed, containing Matlab scripts and functions, as well as a digital version of this report.

Aalborg University June 3rd, 2009

Unai Mtz. de Estíbariz Cerdán
<unaim@es.aau.dk>

Daniel Fernández Álvarez
<danielf@es.aau.dk>

Table of Contents

1	Introduction	3
1.1	Problem statement	4
1.2	Description	4
2	Background Theory	7
2.1	Speech	7
2.2	Hearing losses	8
2.3	Binaural and Monaural Listening	10
2.4	Cocktail-party effect	16
2.5	Intelligibility improvement approaches	17
3	Proposed Aiding Methods	19
3.1	Global signal processing	19
3.2	Algorithm A	23
3.3	Algorithm B	24
4	Listening Test	27
4.1	Introduction	27
4.2	Methods	31
5	Results	41
5.1	Data presentation	41
5.2	Analysis	46
6	Conclusions	51
7	Future Work	53
A	Detailed Results	59
B	PTFs Measurement Report	61
B.1	Procedure	61
C	Graphical User Interface	65
D	Pilot Test	69

Introduction

In the last decades the population of hearing impaired people is increasing rapidly. Nowadays more than 500 million people suffer from different types and degrees of hearing impairments [HearIt09]. Adrian Davis, from the British MRC Institute of Hearing Research, estimated that the total number of people with hearing disabilities worldwide will exceed 900 million in 2025, as shown in figure 1.1. In Denmark this issue is afflicting the 15% of the population [Widex09].

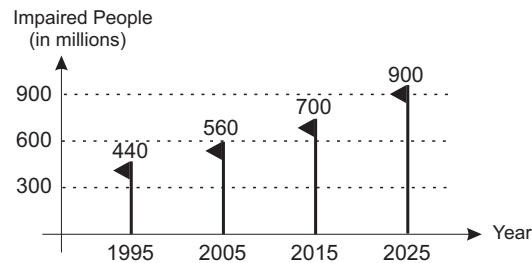


Figure 1.1: Estimation of the number of people suffering from hearing losses of more than 25 dB worldwide. Adapted from [HearIt09].

The awareness, and consequently its treatment, of this disability is increasing together with the growth of science in this field. Nevertheless it still turns not to be enough to decrease the number of people affected worldwide. Lifestyle is considered to highly influence this data due to nowadays urban context, where population is often exposed to loud noises in both leisure and working environments, giving rise to a great number of losses. Other causes of hearing losses are aging, infections or congenital (induced by illnesses and/or complications during pregnancy or birth). A clear indicator of the barrier that a hearing loss causes in a society is shown by the unemployment rates of impaired people (7.5%) versus normal hearing people (4.8%) [DISR03].

A survey for determining the difficulties of hearing impaired people in work places made by the Danish Institute for Social Research, indicated the existence of troubles when conversing with more than one work colleagues at the same time for the 77% of the participants [DISR03]. In children, audition losses can even affect the child's social and linguistic development [Widex09].

In the cases where the loss is severely pronounced in one of the ears, subjects are categorized as what is known in audiology as monaural listeners. For these people conventional amplification usually appears not to be useful enough, and their disadvantages are not

only related with the perceived loudness of the incoming sounds. At a neural level, normal hearing people is able to compare the two auditory signals arriving to each ear. These differences between auditory inputs are known as interaural differences and can be defined either by time and level [Moore04].

1.1 Problem statement

Different studies have demonstrated the reduced abilities of monaural listeners for locating sources, as wells as their difficulties to focus their attention into a particular source in a noisy environment (the latter commonly referred to as cocktail-party effect) [Blauert05]. It has been demonstrated that the accuracy of normal hearing people to focus the attention into a certain spatial margin is due to the processing of interaural differences [Moore04].

Consequently, the so-called monaural listeners may have notable problems for thoroughly understanding a conversation in adverse auditory conditions (e.g. environments with simultaneous talkers). Since speech becomes of such importance, lack of intelligibility eventually leads to social problems in the daily life of monaural people, as their capability to fully understand the message is diminished [Moore04].

Along the reviewed literature several approaches for aiding hearing impaired people have been found [Kollmeier94], [Koehnke94]. However, none of them is specifically focused for monaural listeners. Hence, it becomes of great interest to research how binaural cues can aid the auditory skills of this type of hearing impaired subjects, with focus on improving speech intelligibility in noisy environments.

1.2 Description

To accost the stated problem, different algorithms aiming to improve the speech intelligibility are to be developed. For such purpose, the first step consists on a study of the binaural and monaural hearing. For the system implementation, binaural cues are obtained by means of a couple of microphones, one per ear. The acquired signals have to be researched about how to process them into a single signal which would feed the unimpaired ear.

Various representative scenarios, which try to resemble real noisy environments, are to be simulated so as to assess the performance of the proposed solutions under different conditions. These scenarios will depend on the number and location of masker sources interfering a target message. Eventually these are assessed in a listening test performed over a number of subjects.

All the results will be studied and statistically analyzed, and conclusions will be drawn so as to contribute to the research on monaural impairments. Additionally, future improvements arisen from this study will be mentioned.

Background Theory

A number of different topics related to the current research that were considered important to study are presented in this chapter. The study of these matters was valid to form a more precise image of what later on was going to be designed. Matters such as speech, type of hearing losses, binaural and monaural hearing, the cocktail-party effect and existing intelligibility improvement approaches are commented next.

2.1 Speech

Speech perception deals with the field of neuronal interpretation of complex acoustical patterns to perceive them as linguistic units. After many years of research in the field it can be stated that its perception cannot be based on simple extractions of acoustic patterns directly from the speech waveform. It is understood that these patterns vary in an intricate way based on preceding and following speech sounds. [Moore04]

Speech is a broadband sound with varying frequency spectrum in time. It shows different patterns for female and male voices. The male spectrum low frequency limit is settled down to 90 Hz, whereas for female voices this limit is around 150 Hz [Bronkhorst00].

When studying speech the most straight way to divide it is in sequences of words. Words can also be separated in smaller units in what is known as syllables, which at the same time can be divided into phonemes. These units do not necessarily carry a meaning or symbolize an object. The combination of them yields in the previously mentioned syllables and/or words. It is important to note that phonemes are not defined in terms of acoustics patterns but in what is actually perceived [Moore04].

Other characteristics such as consonants and vowels are another way of separating sounds, and they differ in their frequency range operation. Vowels range from 250 to 3 kHz, while consonants go from 450 to 8 kHz. Each type of consonants operates at different frequency intervals, thus modifying in one or other way the spectral information [Moore04].

Consonants and vowels also differ in the intensity level they are produced, tending the latter to possess a higher level. A term known as Consonant Vowel Ratio (CVR) was introduced to determine differences in level between consonants and vowels. According to [Sammeth99] different studies have studied the amplification effect of the level of consonants in impaired people, which yielded in higher intelligibility. In a listening test where six sensorineural hearing impaired and two normal hearing subjects were employed,

Sammeth himself proposed an approach where vowels were attenuated while consonants were held constant, but this research did not yield any intelligibility improvement. These findings suggests that the most important information of the speech context remains in consonants rather than in vowels.

Intelligibility is affected by many factors, both audiological and environmental. Audiological factors affecting the speech intelligibility are those related to the abilities of the listener to hear properly and they all have influence on the final perception of the speech. Frequency resolution or the dynamic range are abilities that are reduced in monaurally impaired people, yielding in a deterioration of the speech intelligibility [Dillon01].

Environmental factors, such as noise and reverberation, affect more monaurally impaired people because of their limited abilities. While normal hearing people have almost no difficulties at a SNR ranging from 0 to 6 dB, impaired people suffer remarkable intelligibility problems at the same range [Moore04]. Regarding reverberation, low levels of it are assumed to be positive to understand speech, but these levels which are appropriate for normal hearing can result counterproductive for people with hearing losses [PoissantEtAl06].

Other factors, such as the head orientation, also influence the listening experience of speech. [PerssonEtAl01] demonstrated that unilateral impaired people make use of their head orientation to improve their speech intelligibility in different environments: facing the target in quiet environments, while turning the unimpaired ear to the target in noisy environments. This last technique was studied by [EricsonEtAl88] and suggested that the shadow effect made by the head helps to improve the existing SNR.

2.2 Hearing losses

Hearing impaired individuals suffer from different type of hearing losses. Besides of the severity level, it is still possible to establish important differences regarding the origin of the disease, as well as to the distribution of the damage in both ears.

2.2.1 With respect to the origin

Fundamentally there are two types of hearing losses: conductive and sensorineural. Conductive hearing loss occurs because of the reduced transduction capabilities of the middle ear's ossicular bones. Diseases such as otitis media, otosclerosis or malleus fixation are frequent reasons of this type of loss. On the other hand, sensorineural loss is caused by abnormalities in the cochlea and/or the auditory pathway to the brain. Its origin can lie on acoustic neuroma, Manière's disease or ototoxic medications [Vestergaard04].

Conductive impairments can often be treated with the suitable medication, whereas sensorineural damage is permanent. The common way to lighten the latter is by means of amplification, as provided by hearing aids. Furthermore, reduced sensitivity can be a mix between the two mentioned types. In any case, sensorineural damage is the most common form of hearing loss [Vestergaard04].

Sensorineurally impaired listeners have difficulties to understand speech at low levels, even though their perception of high levels is similar to that of normal hearing people. Their loudness perception curve is not linear respect to the SPL, as it approximately happens with normal hearing listeners. Hence, a sudden increase of loudness is perceived when the input level raises a small amount, and this is referred to as *recruitment phenomenon* [Kollmeier94]. See figure 2.1 for a better understanding of this phenomenon.

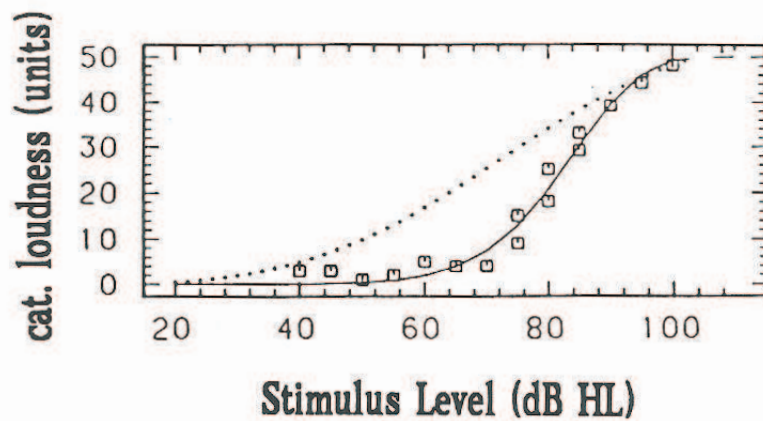


Figure 2.1: Recruitment phenomenon [Kollmeier94].

2.2.2 With respect to the distribution

Sensitivity losses affect each ear in a different manner. Moreover, one of the ears can stay totally healthy whereas the other shows a severe damage. Distinction among the following types of losses can be drawn:

- Symmetrical: both ears are affected in a similar degree
- Asymmetrical: affects each ear in a different degree
- Unilateral: just one of the ears contains a loss
- Bilateral: both ears are affected to some extend

According to this classification the target group of this research, previously entitled as monaurally impaired people, could be classified as asymmetrical and either bilateral or

unilateral distributed hearing loss. For this research the case studied is asymmetrical and unilateral loss, due to the use of normal hearing people in the listening test (further on explained in section 4.1. where the monaural condition was simulated by not reproducing any sound through the headphone channel of the "impaired simulated" ear.

2.3 Binaural and Monaural Listening

A state of the art of different abilities regarding both binaural and monaural listening is presented in the following sections, regarding basic information of topic in question.

2.3.1 Localization

The ability of locating sound has always been of great importance for human beings, as this ability is used to determine the direction of sources either to seek or to avoid, and to establish visual contact with the source in question. Localization is defined as direction and distance judgements to localize a sound source [Moore04]. For this task the human auditory system makes use of different cues, which can be both binaural and monaural.

According to [Moore04], localization performance can be divided in two main aspects: the ability to match the direction of a sound source to its actual direction; and the ability to detect a small spatial shift of a sound source. Regarding the first aspect, common errors in identifying sources direction lying in the median plane are common. The second aspect determines the resolution of the auditory system and the term known as Minimum Audible Angle (MAA) is used to define the smallest detectable change in angular position.

For locating sources in the space a reference is needed. For this case the head of the listener is used and a coordinate system based on the following three spatial planes is defined: median, frontal and horizontal. The point where these three planes intersect is the exact center of the head. Figure 2.2 shows an illustration of the mentioned coordinate system, where θ is the azimuth angle and δ is the elevation angle. The direction of an incoming sound is defined by these two angles, being azimuth the angle projected onto the horizontal plane, and elevation the angle projected onto the median plane [Moore04].

Binaural Cues

The auditory system makes use of dissimilarities between the two ears to locate a sound, where two classes can be found: the ones related to time, known as Interaural Time Differences (ITDs); and the ones related to their sound pressure level, known as Interaural Level Differences (ILDs).

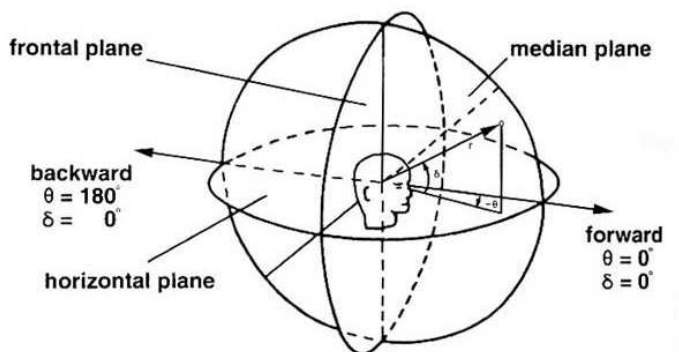


Figure 2.2: Coordinate system [Moore04].

ITD is a cue defined as the temporal displacement of the reproduced signal at one of the ears relative to the other ear [Blauert97], suggesting the existence of different arrival times to each of the ears. This difference is calculated by considering the path difference between the ears. Figure 2.3 shows a graphical representation of this difference.

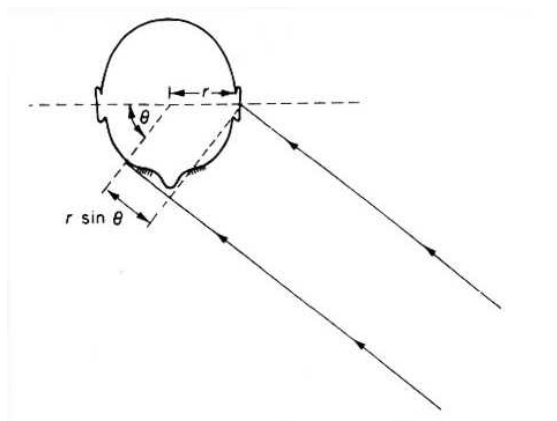


Figure 2.3: Interaural Time Difference path [Moore04].

where r denotes the radius of the head and θ the angle of the incoming source respect to the observer. The path difference, d , is computed by means of equation 2.1:

$$d = r \cdot (\theta + \sin\theta) \quad (2.1)$$

Once this difference has been obtained, the time to reach the ear is calculated, as shown in equation 2.2:

$$ITD = d/c \quad (2.2)$$

where c is the speed of sound in the correspondent environment. If considering a head diameter of 18 cm (as assumed as representative in [Blauert05]) an 343 m/s of sound

speed, according to equations 2.1 and 2.2 ITDs vary between 0 and $675 \mu\text{s}$.

For a better understanding of the ITDs, figure 2.4 shows the dependence of these differences as a function of the azimuth angle, where 0° , 90° and 180° corresponds to the source in front of, completely aside (right or left) and right behind the listener respectively.

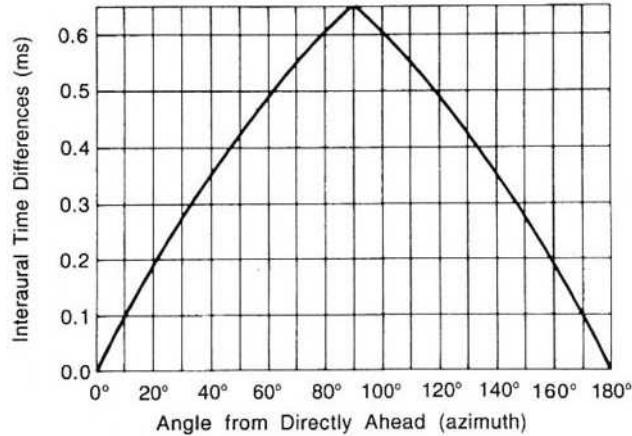


Figure 2.4: Interaural Time Differences [Moore04].

ITDs occur all along the frequency range, but they become more representative in the low frequencies. In the high frequency range ambiguities regarding these differences are common as the length of the wavelength is no longer large compared to the size of the head. In this situation the auditory system can not determine which cycle corresponds to each ear, being 1.5 kHz the frequency limit for unambiguous locations.

The other binaural cue are the ILDs, also known as Interaural Intensity Differences (IIDs), and from now on, along this report they will always be referred as Interaural Level Differences. This cue is based on the level difference between both ears. This situation arises due to head, which partially shadows the contralateral ear. These differences are frequency dependent, as it is shown in figure 2.5.

Analyzing figure 2.5 it is clearly seen that diffraction decreases when increasing the frequencies under test and those differences yield in the known ILDs. Level differences can be interpreted as significant from 1 kHz and up.

In the case that a source is located along the horizontal plane, while the elevation angle is 0° , an experiment performed by [StevensAndNewman36] showed the frequency dependence of the localization accuracy under these circumstances. The range where there is a lack of accuracy covers the frequencies from 2 kHz to 4 kHz. Figure 2.6 illustrates these results.

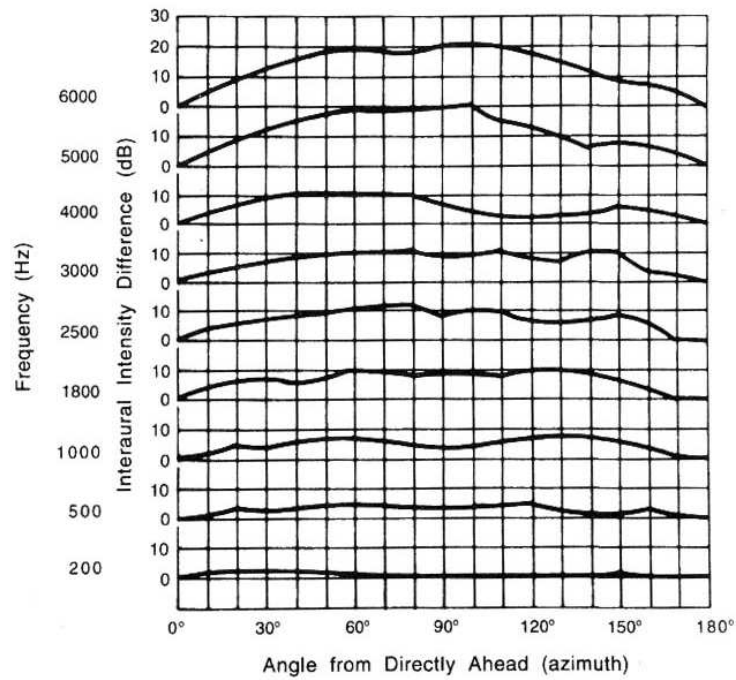


Figure 2.5: Interaural Time Differences [Moore04].

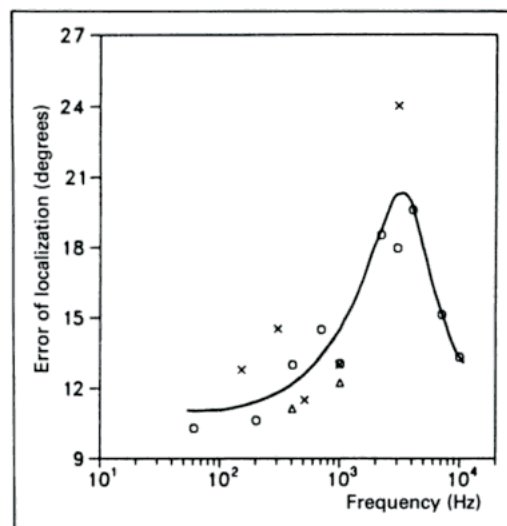


Figure 2.6: Localization in the horizontal plane [BuserAndImbert92] [StevensAndNewman36].

The results shown in this figure were the seed of the later pronounced duality theory, which basically stated that frequencies below 2 kHz were based on ITDs, whereas fre-

quencies above 4 kHz based their operation on ILDs. The frequency range between 2 kHz and 4 kHz was thought not to operate very efficiently at none of the interaural mechanisms commented before, explaining the errors of localization in this range [BuserAndImbert92].

Meanwhile, in the median plane confusions are caused when the source to be localized is placed in front of, behind, or above the subject's head. In these scenarios ITDs and ILDs do not provide conclusive results as these are null. [BuserAndImbert92] showed in an experiment the judgments of direction in the median plane for three different scenarios. Figure 2.7 shows the results of this experiment.

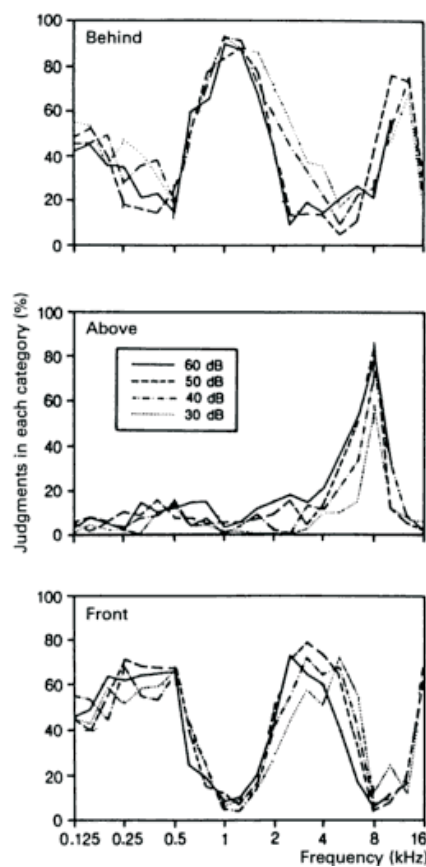


Figure 2.7: Localization in the median plane [BuserAndImbert92].

Trying to localize sounds away from the median plane can also derive in errors, the most common being an effect known as cones of confusion. This phenomenon occurs due to the non-deterministic nature of sound localization. For sounds lying in the surface of the cone, the location of them becomes ambiguous as there exist different spatial positions giving rise to the same ITDs [Mills72]. Figure 2.8 shows the mentioned cone of confusion in one ear.

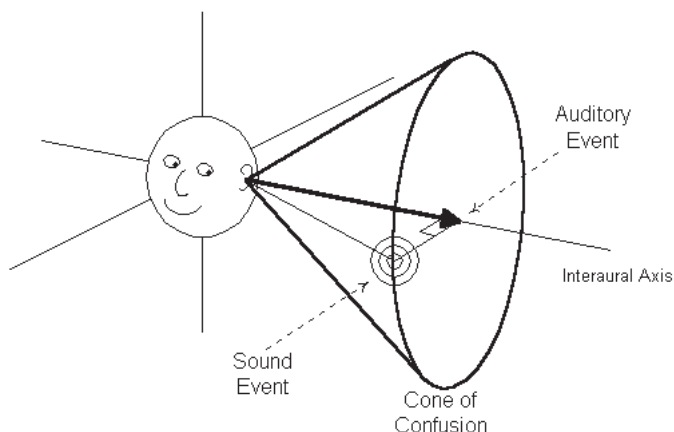


Figure 2.8: Cone of confusion [Mills72].

The most common way to resolve this situation is usually made by means of head movements. [Hirsh71] for example dealt with the improvement of the localization ability when moving the head. [FreedmanAndFisher68] demonstrated that monaural localization performs similarly to binaural localization, suggesting that other cues beside of interaural differences take place in the auditory system when moving the head.

Monaural Cues

Until now only binaural cues for localization have been presented, and in the following lines the cues based on monaural audition are described.

There are cases when neither interaural differences nor head movements provide enough information to predict the location of an incoming sound. [Butler69] suggested that the pinnae was actually used to judge ambiguities in the vertical direction, while [Batteau67] proposed the importance of this organ for every direction. To demonstrate this last statement, an experiment where an artificial pinnae was placed in a pair of headphones was performed. He showed the externalization of the sound, as subjects reported that sound was localized out in the space, instead of inside the head as explained before. Other researchers as [GardnerAndGardner73] studied the influence of occluding cavities of the pinnae, by filling them with moulded rubber plugs, concluding in a decrease of the ability to localize sounds.

According to the studies presented before it can be stated that the pinnae modifies the spectra of sounds in such manner that it depends on the incidence angle of the sound relative to the head. The pinnae together with the head and torso can be described as a complex direction-dependent filter [Moore04], better known as Head Related Transfer Function (HRTF). This transfer function is a complex pattern composed by a number of

peaks and dips highly dependent on the direction of the incoming source respect to the head.

HRTFs are unique for every direction reaching the head [Moore04], as well as for every single person. Different studies have performed experiments to demonstrate this last statement, as [Wenzel93], who measured virtual localization by using representative models of HRTF instead of the subject's custom transfer function; or as [Middlebrooks99] by swapping different HRTFs among subjects. Both concluded in the improvement of the localization ability when subjects listened through their own transfer functions.

As stated before, monaural localization ability can sometimes be comparable to that of binaural listeners [FreedmanAndFisher68]. This is a consequence of monaural adaptation, where other cues rather than interaural differences are exploited by the auditory system in order to be able to locate sounds accurately. [McPartland97] suggested that this adaptation takes several days (even months or years) for some people to be effective. [DíezAndChristensen06] developed a localization experiment where unilateral hearing loss was simulated in a number of subjects for a period of three hours, demonstrating that monaural adaptation was not fulfilled for that period of time.

2.4 Cocktail-party effect

This term refers to the fact that human listeners with healthy binaural-hearing capabilities are able to concentrate on one talker in a crowd of concurrent talkers and discriminate the speech of this talker from the rest [Blauert05]. This ability is extended to the enhancement of the target source within a noisy or reverberant environment, as well as suppressing sound coloration to a certain extent. Moreover, the term is also applied for target sources different than human voices, such as musical samples.

The amount of noise reduction depends on a number of factors, such as the number, position and spectral-temporal properties of the target and the interfering sound sources. In [Kollmeier94] the results of several experiments concerning these factors are analyzed (see figure 2.9). It was concluded that speech intelligibility increased when moving away the interferer source, being either noise or an additional speaker. However, this no longer holds when placing it at 180° or adjacent angles, where a clear decrease of the intelligibility can be noted. At these angles, the auditory system hardly can make use of the interaural differences, which values are practically null. Therefore, this leads to a notable influence of the spatial cues when treating source segregation.

The previous statement coincides with the first studies of the cocktail-party effect by [Cherry53] in the fifties, where it was considered that spatial separation was a major contributor for source segregation. However, recent studies point out that spatial hearing may not be the most relevant cue, as assured by [Yost94]. In any case, it is demonstrated

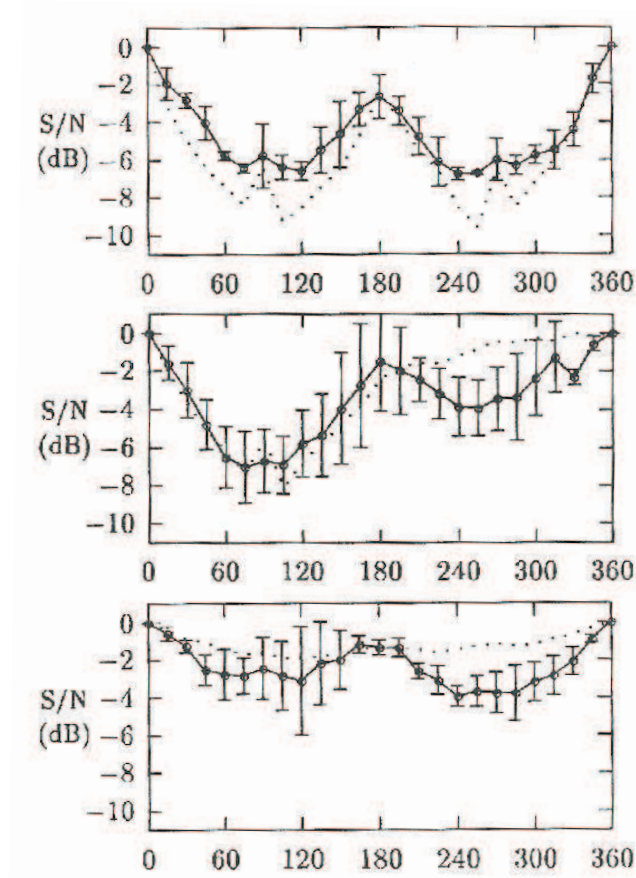


Figure 2.9: Azimuth of interferer speaker (degrees) [Kollmeier94].

that impaired subjects have stronger difficulties than binaural listeners when benefiting of the cocktail-party effect.

2.5 Intelligibility improvement approaches

The scientific community is working on manners to improve the speech intelligibility of impaired people by means of different approaches and algorithms. A very promising approach to overcome this problem is presented by Dr. Kollmeier in different publications [KollmeierEtA193-1] [KollmeierEtA193-2], where two different algorithms were developed to increase the speech intelligibility in noisy environments. The first approach was based on the suppression of lateral noise sources and it performed "surprisingly well" (in words of the researchers) in non-reverberant environments, highly improving the speech intelligibility, while its performance was not that accurate in noisy environments.

Another approach based on a dereverberation algorithm was developed and it yielded an

improvement in speech quality, but not in speech intelligibility. A combination of both as a future work improvement is suggested at the end of [KollmeierEtA193-1], and in [KollmeierEtA193-2] the design of such combination is shown, which results to "operate quite efficiently" under adverse acoustical conditions if a compromise between the two base algorithms is taken into account.

Proposed Aiding Methods

To enhance the speech intelligibility of monaural listeners two solutions have been developed. The corresponding algorithms have been designed and implemented in the mathematical software Matlab. Both of them work considering that the target speaker is always placed in front of the subject (i.e. 0° azimuth), as it usually happens in a common conversation. The so-called Algorithm A takes advantage of the head attenuation, whereas Algorithm B is an approach based on [KollmeierEtAl93-1]. The latter applies frequency dependent attenuation, aiming to diminish the energy coming from lateral directions. Both of them make use of the ITDs, computed by cross-correlating the signals recorded at both ears. Eventually, the processed signal is sent to the healthy ear.

Both algorithms share certain processing, which is explained in section 3.1. The different specifications for each are described next, in sections 3.2 and 3.3.

3.1 Global signal processing

The proposed methods are totally based on the computation of the ITDs. Along this section, the way they are obtained is described, paying special attention to the window properties (i.e., how the signal is segmented in order to obtain an ITD value every certain time). Prefiltering, applied at the beginning of both types of processing, is also explained at the end of the section.

3.1.1 ITDs computation

The sign of a ITD indicates whether the sound comes from one or other side. For the remainder of this report, a positive ITD will indicate a sound source located towards the right of the median plane, whereas a negative ITD will indicate a sound source placed towards the left side of the median plane. The ITDs are computed in the time domain by means of the cross-correlation method. This operation is used to determine the degree of resemblance between two signals. Equation 3.1 defines the cross-correlation function for discrete signals.

$$R[\tau] = \sum_{n=-\infty}^{\infty} x[n] \cdot y[n + \tau] \quad (3.1)$$

where x and y are the signals to be compared and τ is the progressive displacement applied y . When the displacement is such that both signals present a great similitude,

there is maximum in $R[\tau]$. Therefore, periodic signals lead to a maximum for every displacement multiple of the period. In practice, in speech analysis the cross-correlation function is computed over small segments. If a rectangular window, $W[n]$, is applied the formula corresponds to the equation 3.2. This window is defined by N samples of value 1 within the interval $(0, N-1)$, whereas is 0 outside.

$$R(\tau) = \sum_{n=0}^{N-1} \{x[n] \cdot W[n]\} \cdot \{y[n + \tau] \cdot W[n + \tau]\} \quad (3.2)$$

When the sequences under analysis are the two channels of a binaural signal, R_τ becomes maximum for a value of τ equal to the delay between ear, i.e. the ITD. Figure 3.1 shows an example of a 1 ms time delay among channels and its respective cross-correlation plot.

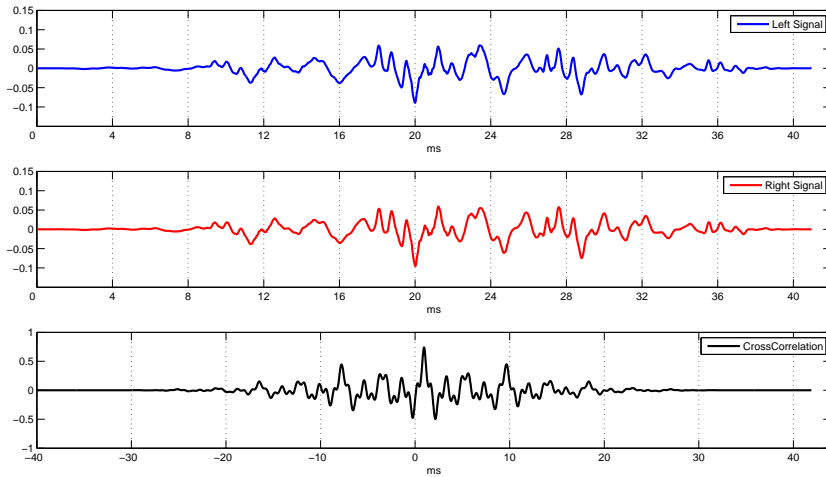


Figure 3.1: 1 ms time delay cross-correlation plot.

Once the ITD is computed, it has to be translated into an angle value. This can be performed according to the graphic showed in [Moore04], displayed previously in figure 2.4. This function considers a maximum ITD of $675 \mu\text{s}$. However, once the measurements were performed (see section 4.1.2), when recording a speaker placed at 90° from the manikin, the obtained value was $750 \mu\text{s}$. For that reason, the function in figure 2.4 was adjusted to assimilate the measuring conditions. The 3D perception matching eventually depends on the diameter of the head of each subject. However, this readjustment is still needed for the off-line processing, independent of the subject. Otherwise, the resulting ITDs would not correspond with the desired angles, and the algorithms calculations would deviate than those expected.

The corresponding ITD for the target speaker, placed at 0° azimuth, is theoretically 0 s. In practice, slightly different values than 0 s are commonly obtained in the measure-

ments. Figure 3.2 shows the ITD values obtained for a single speaker placed in front of the subject, with the window conditions specified in section 3.1.1. Most of the deviations correspond to $19.53 \mu\text{s}$, which is actually the resolution for the ITDs (since the sample frequency was 51200 Hz^\dagger). These deviations might be due to a non-exactly 0° position of the speaker. Sporadically, the system provides higher ITDs, being the most deviated value $78.13 \mu\text{s}$, which means an angle of 7.6° . Therefore, to ensure that other sources are interfering the communication, greater absolute values have to be obtained.

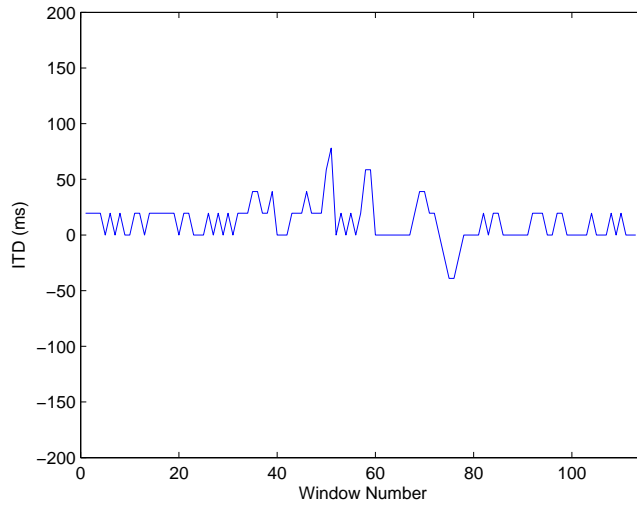


Figure 3.2: Measured ITD values for a speaker placed at 0° azimuth.

Mostly at high frequencies may happen that ITDs are not fully reliably obtained. This is due to the fact that the magnitude of the corresponding periodic components resemble that from the higher ITDs, as previously explained in 2.3.1. Thus, ITD values quite above $1000 \mu\text{s}$ were found at some specific cases, which can not really occur since the maximum ITD should be around $750 \mu\text{s}$. It was noticed that if the analysis range was limited to that physically possible, the ITDs reliability extremely increased. Hence, the limit was stated between -800 and $800 \mu\text{s}$.

Window length and shape

The window length needs to be carefully chosen so as to obtain productive results. A long window provides greater frequency resolution, since more points are used for the Fourier analysis. Furthermore, it reduces the computational load of the algorithm. However, this option would diminish the time resolution. Specifically for this algorithm, long windows are preferable since the determination of the ITDs becomes more unequivocal. Finally,

[†]This sampling frequency is that used by Harmonie, the measuring system for the recordings of the test signals. This frequency is not configurable in such system, and no need of resampling was found.

the compromise solution was set to a window length of 2048 points, which provides a frequency resolution of 25 Hz at the sampling frequency of 51200 Hz.

In order to increase the time resolution, it was decided to apply the overlapping method, in such manner that adjacent windows share part of the signal. 2:1 overlap was finally used, and is represented in figure 3.3. Thus, it was possible to increase the resolution to 20 ms. Along the reconstruction stage, it becomes necessary to compensate the excess of energy due to the overlap method. This is perfectly achieved if a Hanning window is used, which formula is displayed in equation 3.3 Opposite to the most intuitive rectangular window (see equation 3.4, which just cuts the corresponding part of the signal, this other type applies a gradual and symmetric attenuation at both extremes of the window. Additionally, if its frequency response is analyzed, it can be noted that in the Hanning window the side-lobes power decreases. These side-lobes are clearly undesirable since they may cause the spectral measurement to be corrupted by adjacent frequency components [Owens93]. However, the rectangular window has a narrower mainlobe and thus, for a given length, it should yield the sharpest transitions when a discontinuity occurs [Oppenheim89]. Figures 3.4 and 3.5 represent both windows in time and frequency domain.

$$w_h[n] = \begin{cases} 0.5 - 0.5\cos(2\pi n/M), & 0 \leq n \leq M \\ 0, & \textit{otherwise} \end{cases} \quad (3.3)$$

$$w_r[n] = \begin{cases} 1, & 0 \leq n \leq M \\ 0, & \textit{otherwise} \end{cases} \quad (3.4)$$

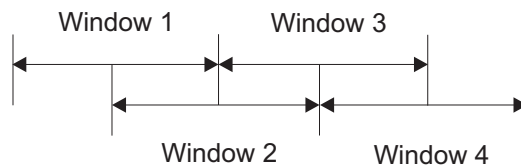


Figure 3.3: 2:1 overlap.

Prefiltering

First of all, the binaural signal is filtered between 100 Hz and 10 kHz. According to [Owens93], speech contains frequency components with significant energies up to about 10 kHz, even though the spectra of majority speech sounds only have significant content up to about 5 kHz. On the other hand, dispensing the content below 100 Hz can result beneficial since some reverberant components, which often hinder the intelligibility, would be suppressed. In any case, this filtering would not imply any decrease of the intelligibility, according to figure 3.6, extracted from [Poulsen05].

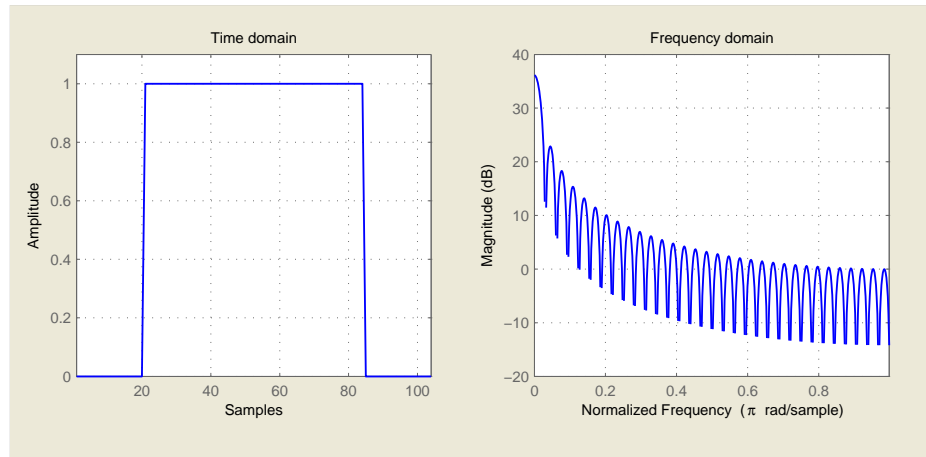


Figure 3.4: Time and frequency response of the rectangular window.

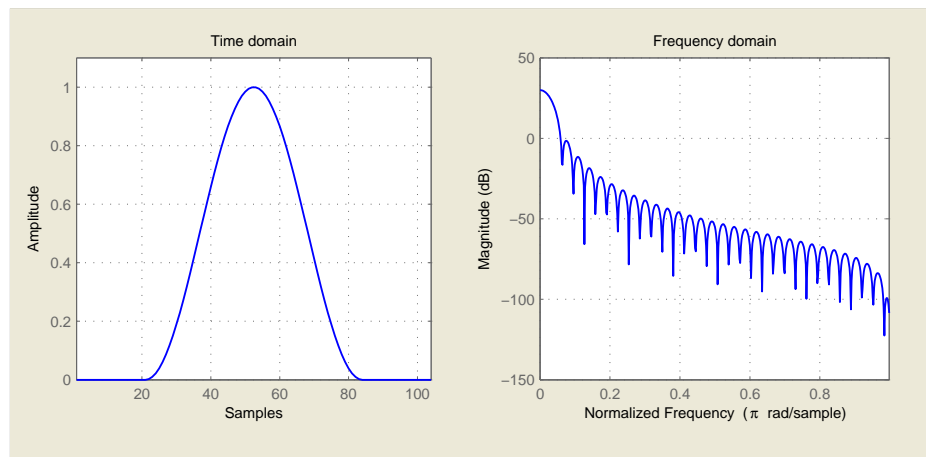


Figure 3.5: Time and frequency response of the Hanning window.

3.2 Algorithm A

When a source is placed laterally with respect to the listener, the wave at the contralateral ear arrives attenuated due to the shadow effect of the head. This attenuation affects mostly to mid and high frequencies and is angle dependent, as it is reflected in figure 2.5, extracted from [Moore04]. According to that figure, shadow effect starts to appear at 500 Hz, with attenuations about 3 or 4 dB. Attenuations up to 20 dB occur from 5 kHz on.

Since the target is placed in front of the listener, any wave coming from other directions can be considered as noise. Hence, this first proposal aims to improve the intelligibility by choosing the channel with higher SNR so as to send it to the healthy ear: that opposite to the side the noise comes from. Whatever channel is chosen, the target perception is

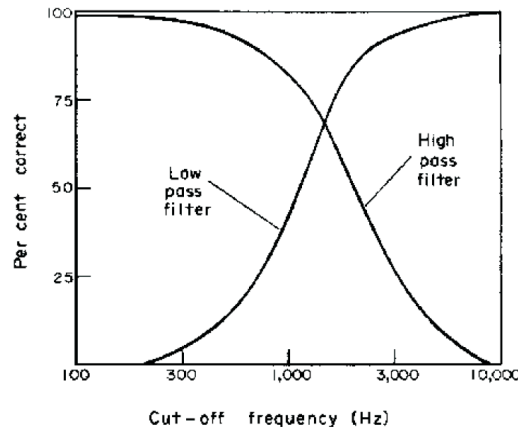


Figure 3.6: Intelligibility for low- and high-pass filtered speech, [Poulsen05].

not affected, since theoretically arrives to each ear under the same conditions.

Algorithm A is characterized by its simplicity, since it hardly requires more processing than that already mentioned. When a segment of signal provides an ITD greater than 0 s, the signal recorded at the left ear is chosen, and viceversa. Therefore, it is a promising solution for a real-time application.

3.3 Algorithm B

In [KollmeierEtAl93-1], the author introduces the basis of his method for lateral noise suppression. In this research, recordings are performed at a sampling frequency of 20 kHz. Signals are segmented into blocks, each Fourier transformed and converted to logarithmic magnitude and phase spectra. For each of these frequency components, the interaural level and phase differences are computed and compared with those the target would provide. According to this deviation an attenuation of up to 20 dB is applied.

Algorithm B is inspired in Kollmeier's proposal. Each window is passed by a filter bank, and the ITD is computed for each of the outputs. Those bands which show a sufficiently noticeable ITD are attenuated. The margin, in terms of the ITDs, where the energy is preserved varies depending on the scenario. Further specifications are explained next.

Band analysis

The human ear resembles a bank of filters with $1/2$ - $1/8$ bandwidths ([Moore04]). Therefore, a $1/3$ octave band analysis seems reasonable. However, some difficulties were found to implement such accuracy level. On one hand, the computational load extremely increased when compared with a octave band analysis. Additionally, the ITDs compu-

tation resulted more efficient when using an octave bank filter. This is due to that the cross-correlation is applied to a broader signal, in terms of frequency width. The reason is that, mostly at high frequencies, some components can lead to a certain ambiguity in the ITD determination (previously explained in section 2.3.1). This phenomenon easily diminishes as the signal becomes broader, since more frequency components may help to disambiguate. Hence, an octave band filter bank was finally applied.

Such bank was implemented by means of a Butterworth band-pass filters of order 8. Their frequency response is represented in figure 3.7.

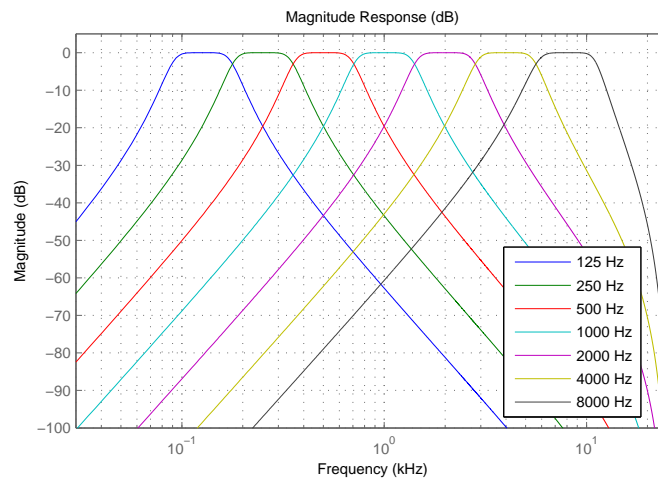


Figure 3.7: Magnitude response of the octave filters.

Margin

Let the scenario consist on a single masker placed 90° with respect to the listener, and the target still at 0° . When both speakers are talking, and assuming a balance of energies, the resulting global ITD will approximately determine an angle around 45° . Hence, if the ITD indicates an angle greater than 45° , it would actually mean that the masker is being stronger than the target, so attenuation must be applied. Therefore, for a suitable performance of the algorithm, the border angle for applying or not attenuation should be dependent on the masker location. Consequently, this limit has always been set halfway from the closest masker location.

Attenuation

Different values of maximum attenuation were considered. The higher the more efficient was the splitting between maskers and target, therefore easier to pick the desired message. However, artifacts occurred, resulting into unnatural reconstructed signals. A

balance between intelligibility improvement and naturalness was aimed. Even though the optimum value slightly oscillated depending on the scenario, an amount of 14 dB was fixed for all the cases.

In order to avoid abrupt attenuation changes between either bands or windows, a transition area was set. Medium attenuation, i.e. 7 dB, is applied when the ITD results into an angle 10° above or below the border angle. Figure 3.8 represents the applied attenuation function when there is a masker located 90° with respect to the listener.

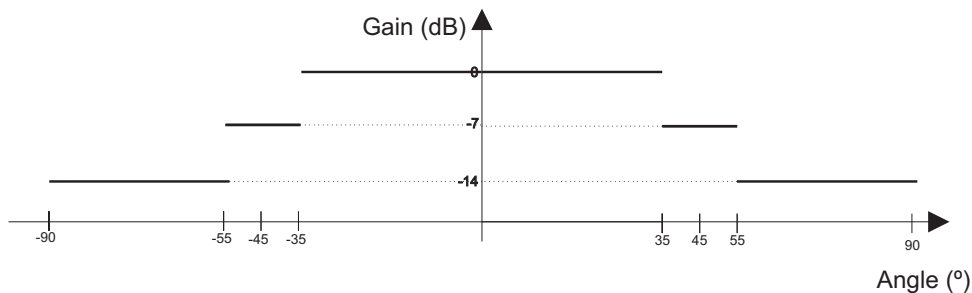


Figure 3.8: Attenuation function for the case of a masker located at 90° .

[KollmeierEtAl93-1] also experienced certain artifacts along the design of its algorithm. One of the pointed parameters were indeed the attenuation function as well as its maximum value. The shape of such function was modified in several ways and the naturalness subjectively evaluated. Nonetheless, it was not possible to perceive consistent differences. In any case, it was decided that keeping the half-way attenuation was a better option than a progressive curve. Since the consequent multiband filter can only have 3 possible changes, less number of quick fluctuations can be produced.

Filtering

Once the attenuation values are obtained for each band, the multiband filter the full window has to pass through can be designed. FIR filters are most suitable for this purpose since they provide linear phase. If IIR filters had been used, the phase of a certain frequencies would change between windows. This phenomenon could yield in annoying artifacts in the reconstructed signal. The filter order was set to 512, high enough to suitably approach the specifications.

As it has been explained before, the signal is low-pass filtered at 10 kHz, so the response of the filter above this frequency is not important in terms of the gain applied to the signal. However, with view to facilitate the filter design, up to the Nyquist frequency the magnitude equals always the value given for the last band, i.e. 8 kHz.

Listening Test

In this chapter the different processes followed to design the listening test employed in this research are shown. The listening test will help to understand if there exist improvements, deteriorations or indifferences between the unaided situation and the proposed approaches.

The algorithms assessment was performed off-line, so the subjects were presented already processed signals. For this purpose several scenarios are synthesized from a number of binaural recordings. These were obtained by means of a manikin, which simulates an averaged human head and torso.

4.1 Introduction

The idea of this research is to try to help people with monaural hearing, thus is mandatory to study which is the actual speech intelligibility under monaural unaided conditions. At the same time, the binaural condition was tested and used as control, being valid to show the actual intelligibility problem in monaurally impaired people.

No use of monaural listeners was contemplated due to the difficulty to find them. Instead healthy hearing subjects were used in this experiment (see 4.2.3). The idea is to investigate the concept in healthy binaural hearing people, and if the system yields any benefit it could be corroborated with real monaural listeners. The monaural listening condition of the selected subjects was simulated by silencing one of their inputs (one channel of the headphones), simulating the correspondent ear as impaired.

4.1.1 Situations

Different ways of presenting the recorded signals were employed. Four configurations can be found: binaural, monaural unaided and monaural aided, with both Algorithm A and Algorithm B. Following sections are entitled to present the commented situations.

Binaural

Both channels (left and right) without manipulation are presented to the listener in this unaided configuration. The sound coming to each of the ears resembles the listening

experience that would occur in a real environment. A scheme of the configuration is shown in figure 4.1.

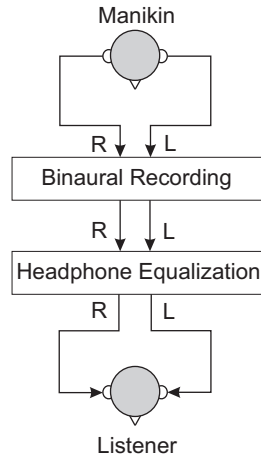


Figure 4.1: Binaural Situation.

Monaural unaided

Only one channel, left or right, (in this study left was chosen for convenience) is sent to the listener through headphones. The unimpaired ear determines the channel to be sent, while the other channel is silenced, simulating this last ear as monaurally impaired. Likewise the binaural situation, no manipulation of the acquired signals beside of the silencing of one of the channels is performed in this configuration. Figure 4.2 depicts the unaided monaural configuration.

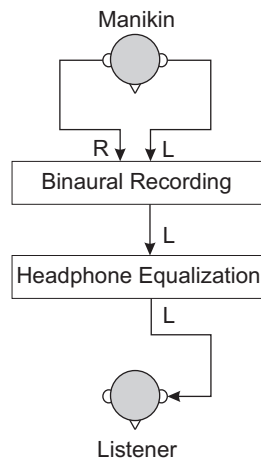


Figure 4.2: Monaural Unaided Situation.

Monaural aided

Same configuration as before, but applying an aiding algorithm. Two different approaches have been designed in this research, as they were explained in chapter 3. The operation of this situation is explained in figure 4.3.

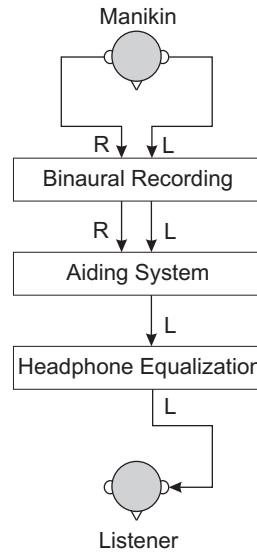


Figure 4.3: Monaural Aided Situation.

4.1.2 Scenarios

The design of the scenarios was made with focus on the exploitation of the developed algorithms, so as to evaluate those into real and adverse situations. For these reasons the scenarios shown in this section try to resemble problematic situations in reality. Three different groups were designed, depending on the number of maskers present in the scenario. Thus, single, double and multiple masking were designed, each containing three different scenarios.

Different output levels were chosen to compensate for the difficulty variability depending on the group. These levels were calculated throughout a pilot test aiming to yield a 50% of correct answers per group so as to evaluate them fairly (see Appendix D for more details).

Note that no recordings were made in the range between $+45^\circ$ and $+90^\circ$ as the right ear was selected as impaired by default. Incident sounds from this range were not considered crucial, as the impairment of the mentioned ear does not affect the overall intelligibility excessively due to the impairment itself. Therefore it was decided to focus all the attention on the angles ranging from -90° to $+45^\circ$, as those were considered the most valuable for this research. See figure 4.4 for a better understanding of the studied versus

non-studied areas.

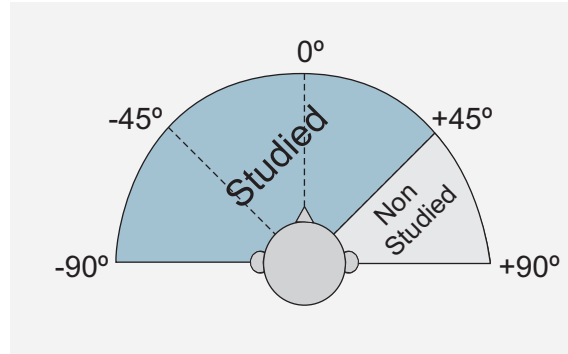


Figure 4.4: Studied range.

Note that blue and red circles in the following sections indicate target and masker(s) respectively, and the arrow shows the direction of speech. Every masker was always located two meters away from the subject, likewise the target.

Single Masking

In the first group three different scenario combinations are shown where only a single masker is used as interferer. In all the situations the speakers are facing the listeners as it was desired to study the most adverse situations for speech intelligibility performance. In figure 4.5 the designed three scenarios can be seen.

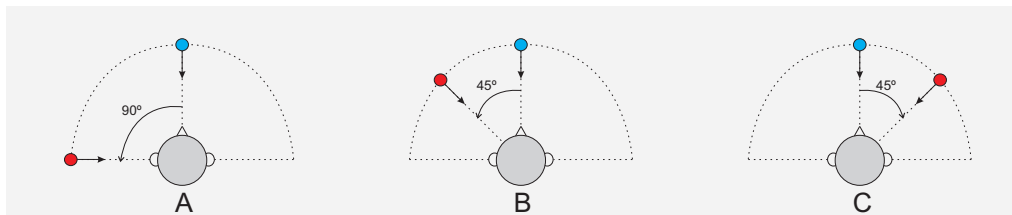


Figure 4.5: Scenarios with a single masker.

Double Masking

Double masking refers to the existence of two different interferers together with the target. The first scenario, A, follows the same principle of evaluating the most adverse situation, while scenarios B and C are based on a more realistic setup configuration for the case of two speakers facing each other and interfering with the talker. Figure 4.6 depicts the proposed scenarios for this group.

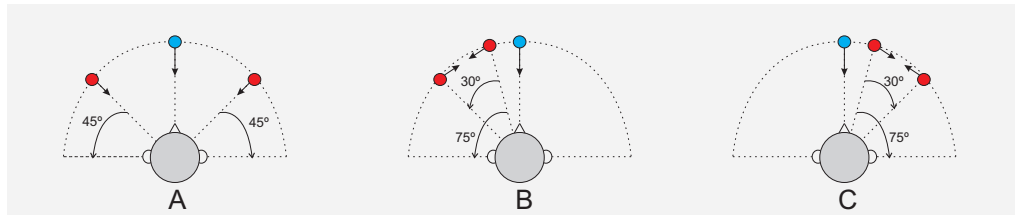


Figure 4.6: Scenarios with double maskers.

Multiple Masking

Last group illustrates the case of four different maskers speaking together with the target. The layouts are based on realistic situations where 5 people (4 maskers plus a target) coincide in different situations. Figure 4.7 shows the designed scenarios for this last group.

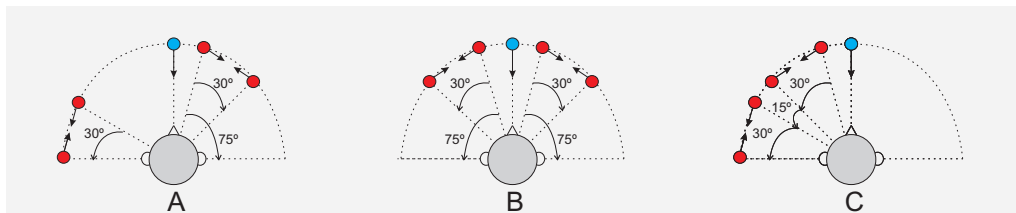


Figure 4.7: Scenarios with multiple maskers.

4.1.3 Means of reproduction

An important issue in the chain of reproduction is the way the last is actually delivered to the listener's ears. According to [HawleyEtAl99] the ability of listeners to extract vital information about the target source can be perturbed in natural environments. By using headphones to reproduce test signals, complete control of the listening experience is achieved. Besides, it facilitates the binaural reproduction of the signals for achieving a complete three dimensional hearing experience, and this issue is easily achieved by means of headphones [Møller92]. An inverse filter compensating for the headphones response (see Appendix B) is also applied, as it provides with relevant binaural and monaural cues to the listener.

4.2 Methods

Once the global aspects of the listening test have been introduced a more detailed study of the methods used to perform the last are presented. Issues such as the utilized intelligibility evaluation method, balancing process of the test, setup configuration, selected

subjects, response interface, data gathering and the test routine are presented below.

4.2.1 Intelligibility evaluation method

Speech intelligibility can be evaluated following different principles. A way to do so are methods based on the study of speech phonemes, such as the Modified Rhyme Test (MRT) [BoliaEtA100].

This method is used as a comparative evaluation of speech intelligibility of single initial and final consonants. It makes use of 50 different groups of six monosyllabic English words (there is also a German version) words rhyming or sounding similar, as shown in table 4.1.

1 st	went	sent	sent	dent	tent	rent
2 nd	hold	cold	told	fold	sold	gold
3 rd	pat	pad	pan	path	pass	pack
4 th	lane	lay	late	lake	lace	lame
5 th	kit	bit	fit	hot	wit	sit
n^{th}
50 th	must	bust	gust	rust	dust	just

Table 4.1: Modified Rhyme Test [Meyer09]

These monosyllables are built as a consonant-vowel-consonant sequence, where the differentiation among the first or last consonant sound is the base of the six phonetically different words. Together with a carrier sentence, these words are presented to the subject, whose task is to identify which of the six similar words has been the one spoken by the talker. Different ways to analyze the results can be found, such as evaluating the number of correct answers, the numbers of incorrect answers, or even the frequencies of particular confusions of consonant sounds [Meyer09].

Another type of method also used for measuring speech intelligibility is the one entitled as Coordinate Measure Response (CMR), which was firstly presented by T. J. Moore in 1981. This method was developed to assess speech intelligibility under multitalker communication environments, such as military environments [BoliaEtA100]. Different phrases are utilized to evaluate this approach, and they all consist of a call sign and a color-number combination. Embedding these three factors in a carrier sentence the phrase to present to the subject is created. See an example in the following line:

*"Ready **Baron**, go to **Blue One** now",*

being *Baron* the call sign and *Blue One* the color-number combination.

The duty of the subject is to listen carefully to both the call sign and the color-number combination. Measures such as the correct detections of calls signs and/or color-number combinations, as well as their reaction time can be evaluated with this method. The result of this test would yield in the ability of a listener to actively focus all attention into a single source while other competing sources are present [BoliaEtA100].

CRM it is not considered as a replacement for phonetically balanced speech intelligibility measures due to its limited vocabulary, but still possesses advantages over these type of measures for certain situations.

According to [Moore81], who compared both the CRM and the MRT in a variety of jamming conditions, it was stated that MRT is more sensitive to interfering noise than CRM, but at the same time concluded that a high correlation in terms of the overall performance was achieved among both methods. In [Brungart01] it is concluded that the CRM can be attractive for detecting small intelligibility changes in noisy environments, as well as the intrinsic nature of the method portability into other languages. This portability provides a rough measure of intelligibility with no need of deriving phonemes depending on the language to be used for the test. The simplicity for gathering the signals and evaluating the results is also considered a major advantage by [Brungart01].

Resuming, CRM is not a comprehensive measure of speech intelligibility as the MRT is, but it becomes suitable if the purpose is a rapid and reliable measure of intelligibility [Brungart01]. Because of this it was decided to make use of the CRM to evaluate the performance of the current research.

Coordinate Response Measure Design

According to the principles about the CRM explained in section 4.2.1 it was decided to use a combination of 4 numbers (1, 2, 3 and 4) and 4 colors (blue, red, green and yellow), yielding in 16 different sentences differing in the color-number combination. A common call sign ("*Dragonfly*") was used in order to focus all the attention into the following color-number combination. An example of the phrase to be recorded can be found next:

*"Ready **Dragonfly**, go to **Yellow Two** now",*

where *Dragonfly* is the call sign and *Yellow Two* the color-number combination. Table 4.2 shows the possible combinations based on the designed configuration.

The process of recording speech samples was performed in the Multi-Channel listening room, with a reverberation time of around 0.3 seconds. In the case of the target, a 24 years old male speaker, he was told to place himself two meters away from the manikin at 0° azimuth facing the latest. The sentences were presented to the speaker beforehand and time to prepare them was given. A sentence recorded by the experimenters was

Color / Number	1	2	3	4
Blue	C1	C2	C3	C4
Red	C5	C6	C7	C8
Green	C9	C10	C11	C12
Yellow	C13	C14	C15	C16

Table 4.2: Configuration of the designed CRM.

presented prior to the recording in order to guide the speaker in speed and content.

Same process was followed to record phrases to be spoken by the maskers, but this time different chapters of the book written by Nick Hornby and entitled as "*High Fidelity*" were used as interfering speech signals, the same way as [KollmeierEtAl93-1] did with another title. Different speakers, six in total (4 male and 2 female speakers ranging between 23 and 24 years old), were recorded one by one. The different scenarios (see section 4.1.2) were synthesized by adding these single recordings.

Find the list of the used devices to perform the recordings in table 4.3.

DEVICE	Manufacturer	MODEL	SERIAL NUMBER
Manikin	AAU	Valdemar Sejr	aau2150-03
Left microphone	Gras	40AD	aau56521
Right microphone	Gras	40AD	aau56520
Phantom source	Neumann	BS48i-2	aau2018-00
Measuring system	01dB	Harmonie	aau56524
Laptop	Siemens	E-series Lifebook	aau60921

Table 4.3: List of devices used for the recordings of scenarios.

In the post-processing stage, an approximation of the same root mean square value was set for all the recorded signals, following the principle that the contribution of the target and masker to both ears had to be the same. After it, silence was added to the recorded samples so as to get rid of the noise floor which could be accumulated when adding different recording to compose a scenario. According to the designed scenarios, the different recorded signals were synthesized and synchronized for the later presentation in a listening test. All the post-processing was done in the technical computing software *Matlab R2007a*, and all the functions used for this purpose can be found in the enclosed CD-ROM.

4.2.2 Balancing

When designing a listening test it is important to balance all the factors involving the test, so as to draw a fair conclusion out of them. In this case study it is wanted to investigate if there are differences between the proposed algorithms and the unaided situation,

pure monaural situation, as well as a comparison with the binaural situation.

It is highly important to balance the order of the presentations of the different situations (later called sessions) across subjects. For this purpose the design is based on the latin squares principle, which is an n by n table fulfilled with n different symbols. The theory of this application lies on the fact that each symbol appears just once in each row and column. This layout makes latin squares appropriate for the design of experiments with no need of evaluating all the existing combinations, which would lead to a large number of combinatory possibilities. Table 4.4 shows an example of the balancing of different situations.

Monaural	Algorithm A	Algorithm B
Algorithm A	Algorithm B	Monaural
Algorithm B	Monaural	Algorithm A

Table 4.4: Test type balancing.

Color-number combinations also have to be balanced so as to give the same weight to every combination (16 in total considering four colors and four numbers), where three repetitions in order to give consistency of each scenario are performed. Table 4.5 shows the proposed order for this balancing.

Scenario	1 st Repetition	2 nd Repetition	3 rd Repetition
Single Masking A (SA)	Blue 4	Yellow 2	Red 3
Single Masking B (SB)	Yellow 3	Red 1	Green 4
Single Masking C (SC)	Red 4	Green 3	Blue 2
Double Masking A (DA)	Green 2	Blue 1	Yellow 4
Double Masking B (DB)	Blue 2	Yellow 3	Green 1
Double Masking C (DC)	Yellow 1	Green 4	Red 2
Multiple Masking A (MA)	Red 3	Blue 1	Green 2
Multiple Masking B (MB)	Blue 2	Red 4	Yellow 1
Multiple Masking C (MC)	Yellow 4	Green 1	Blue 3

Table 4.5: Color-number balancing.

It is also desired that the presentation order of the scenarios does not affect the final outcome of the results. This balancing is also based on the latin squares principle, so every scenario is only present once in the same row and column. A design containing a different color-number combinations per scenario was designed, as shown in table 4.6. According to all these configurations a final table 4.7 englobing all the issues presented before is shown.

α	SC	MA	DA	SA	SB	DB	MC	DC	MB
β	MC	DA	SA	MA	MB	SB	DC	SC	DB
γ	DA	MB	DB	SB	SC	DC	SA	MA	MC
δ	DB	MC	DC	SC	DA	MA	SB	MB	SA
ϵ	SB	DC	SC	MC	SA	DA	MB	DB	MA
ζ	MA	SB	MB	DB	DC	MC	DA	SA	SC
θ	MB	SC	MC	DC	MA	SA	DB	SB	DA
π	SA	DB	SB	MB	MC	SC	MA	DA	DC
σ	DC	SA	MA	DA	DB	MB	SC	MC	SB

Table 4.6: Test type balancing.

Subject nr.	Session 1	Session 2	Session 3	Session 4	Session 5
1	Famil.	Mono (α)	AlgA (θ)	AlgB (δ)	Bin (π)
2	Famil.	AlgA (β)	AlgB (π)	Mono (ϵ)	Bin (σ)
3	Famil.	AlgB (γ)	Mono (σ)	AlgA (ζ)	Bin (α)
4	Famil.	Mono (δ)	AlgA (α)	AlgB (θ)	Bin (β)
5	Famil.	AlgA (ϵ)	AlgB (β)	Mono (π)	Bin (γ)
6	Famil.	AlgB (ζ)	Mono (γ)	AlgA (σ)	Bin (δ)
7	Famil.	Mono (θ)	AlgA (δ)	AlgB (α)	Bin (ϵ)
8	Famil.	AlgA (π)	AlgB (ϵ)	Mono (β)	Bin (ζ)
9	Famil.	AlgB (σ)	Mono (ζ)	AlgA (γ)	Bin (θ)

Table 4.7: Total balance.

Session 1, or Familiarization session is the same for all the subjects, as it was believed that they should all have the exact same training to face the listening test fairly. Note that the binaural session is always performed last because of subjective behaviors that may interfere with the final results, as the subject's mood may go down if he or she would feel that their monaural performance is not good enough compared the binaural one. Besides, the binaural session will only be used for corroborating the listening test.

4.2.3 Subjects

All the subjects used for the experiment were considered healthy hearing people. An audiometry prior to the listening test was performed to every one of them and none of the volunteers participating in the listening test had to be discarded. The threshold level to discard subjects was set to 20 dB HL or differences of more than 10 dB between ears for the studied frequency bands [250 Hz - 8 kHz].

9 subjects were used to assure a fair balancing and also because it was considered a fair number of subjects considering the time limitations of the present research.

All subjects were students of Aalborg University, and the age of this population ranged between 22 and 28 years old, where 8 males and a single woman participated. English was not the mother tongue of any of the participants, but all of them spoke it fluently.

4.2.4 Setup

The setup prepared for the listening is also of great importance. The environment where the subjects should perform the test has to be as silenced and as controlled as possible in case any problem occurs. For this reason Cabin A in Aalborg University Acoustics facilities was used to perform the test, which is connected with Control Room K for monitoring and controlling of every event occurring in Cabin A.

The subject performing the listening test is alone in Cabin A and uses a screen controlled by a mouse to respond to the stimuli presented through headphones. The information displayed in the screen is processed by a computer in Control Room K which is interconnected with the screen in the cabin through an extender. This way the experimenter in the control room will also be able to see all the actions performed by the subject. Intercommunicator devices are also placed in both room to provide communication if needed, such as the starting time after every session. Figure 4.8 shows the layout of Cabin A for the designed experiment.

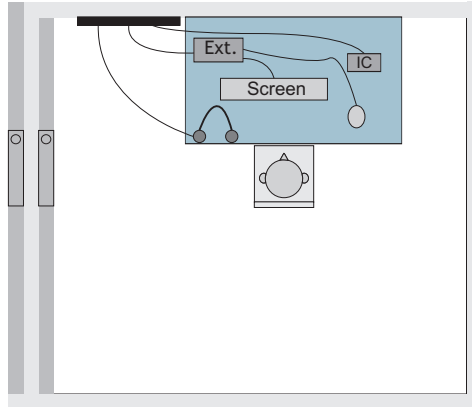


Figure 4.8: CabinA

In control room K all the processes are controlled and monitored by the experimenter. The audio signal from the computer is passed through a power amplifier so as to increase the maximum level to present to the subject as the computer's level was not powerful enough to play the desired sound level. Find the layout of this room in figure 4.9.

Find all the information of the utilized devices in table 4.8.

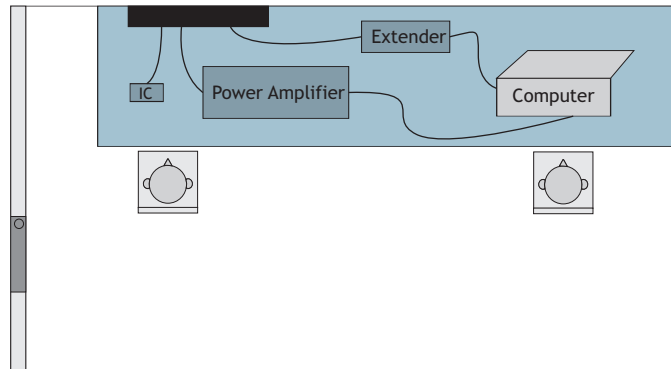


Figure 4.9: ControlRoomK

DEVICE	Manufacturer	MODEL	SERIAL NUMBER
Headphones	Beyerdynamic	DT990	aau2036-12
Amplifier	Sony	STR-DB790	aau56555-00
Laptop	Apple	MacBook	–
Screen	LG	Flatron L1710S	aau57416
Extender	KVM	Danbit	aau215300
InterCom	AAU	Boyer	aau2156-02

Table 4.8: List of devices used for performing the listening test.

4.2.5 Response interface and data collection

A user friendly graphical interface (see captions in Appendix C) built in Matlab environment was used to present the listening test and to collect the data obtained from it for the participating subjects. The presented stimuli were stored in the utilized computer (see table 4.8) and they were played back according to the order stated in 4.2.2. The collected files (subjects answers) were .mat files (matrices) saved in the computer after every session for a later processing of them (see chapter 5).

4.2.6 Test routine

In this section the process that was followed for every one of the participants in the listening test is explained.

First process was to perform an audiometry so as to make sure that the person was eligible for this test. Once this stage was overtaken a set of instructions was given to the subjects, first in a written form and later orally if questions arose. After this stage the subject was informed about how to communicate with the testers in case any problem during the test occurred, by means of an inter communicator placed at both the listening cabin and the control room where the process was monitored. At this point the subject

was ready to perform the test.

The first session is a familiarization session which helps the subject to understand what is the test about, get used to monaural audition and how the answers have to be submitted in the graphical user interface. Different monaural samples (processed and unprocessed) were shown in this stage. This session was the only one where feedback (correct and wrong) about his or her answers was given, as it was desired to teach the participant how the test worked.

Sessions 2 to 5 are performed without any feedback as it was believed that showing the answers at these stages might be counterproductive if the subject's performance was not very good. After all it is very important that the participant focused just on answering what he or she heard, and not be influenced by their own answers. Breaks of 5 minutes between sessions were suggested to the subjects in order not to fatigue them. Every session lasted around 10 minutes (familiarization was 5 minutes), leading to around an hour and 15 minutes of listening test including breaks and audiometry test.

In every break the subject was asked to rate (from 1 to 10) the session's naturalness so as to gain some feedback about the different situations sound quality. With this questionnaire it was desired to obtain an idea about how close to a real situation the presented samples were, as not only the performance of the proposed approaches is desired to evaluate, but also the resemblance with natural sounds. This factor is a very important issue to take into account, as an appropriate operation of a developed approach should be a balance among effectiveness and closeness to reality. This issue is studied in 5.2.3.

Results

In this chapter a number of figures shows the score in percentages as a function of situations, scenarios, group of scenarios and subjects. After the data presentation, analysis of variance (ANOVA) is performed so as to assess whether the proposed algorithms do improve the intelligibility. Likewise, the interaction between the algorithms and the group of scenarios is studied.

5.1 Data presentation

For a detailed analysis, Appendix A contains a table with all the percentages of correct answers by the subjects, as function of the scenarios and situations.

5.1.1 Subjects

The first plot of this section, figure 5.1, indicates the percentage of correct answers per subject split into numbers and colors. Likewise, results are divided depending on the situation. As expected, when presenting a binaural sample to the subjects, their performance clearly improves. Derived from that plot, figure 5.2 shows the global performance of the subjects, once color and number of correct answers percentages are averaged. Both figures do not indicate a strong difference among subjects. Nevertheless, this fact is thoroughly evaluated in section 5.2.4.

5.1.2 Scenarios

The dependence of the results of the designed scenarios is to be evaluated in this section. Before starting to conclude about the different scenarios figure 5.3 is presented, containing the color, number and average scores for the different situations per scenario.

Comparing the results of the scenario Single Masking A with the other scenarios with a unique masker it can be seen that this scenario yields worst results than the rest. However, the positioning of the single masker for this scenario configuration was at -90° , which is assumed as a favorable position as the masker is placed at the furthest angle from the listener (see figure 2.9). This issue may be the result of the target and masker audio signals synchronization, where both speeches seemed to coincide in excess at the target words.

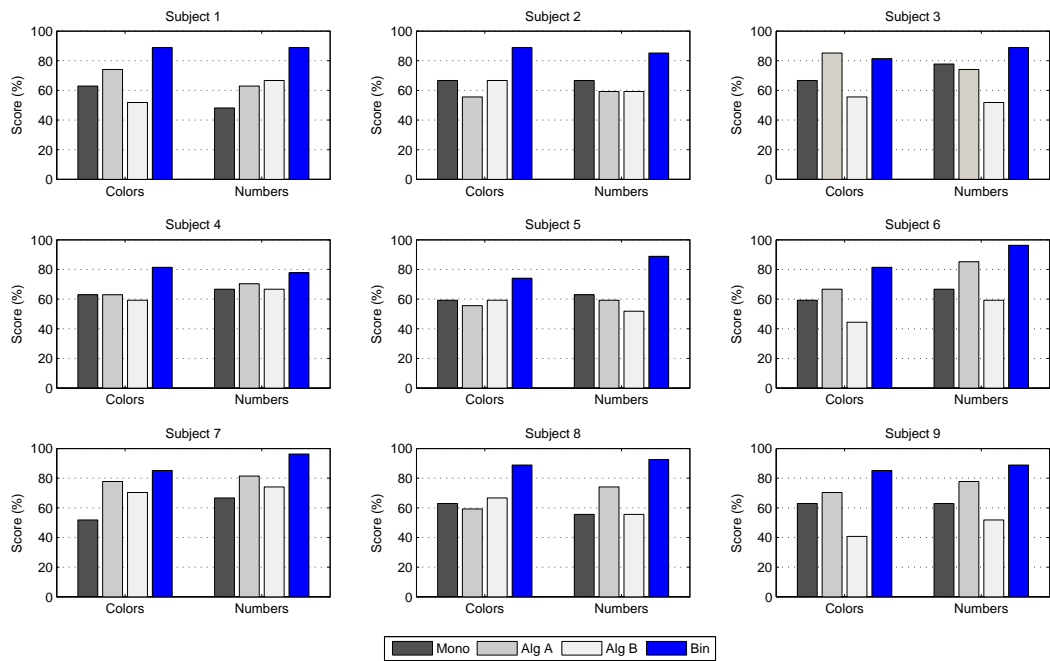


Figure 5.1: Color and number correct answers (in %) per subject for every situation.

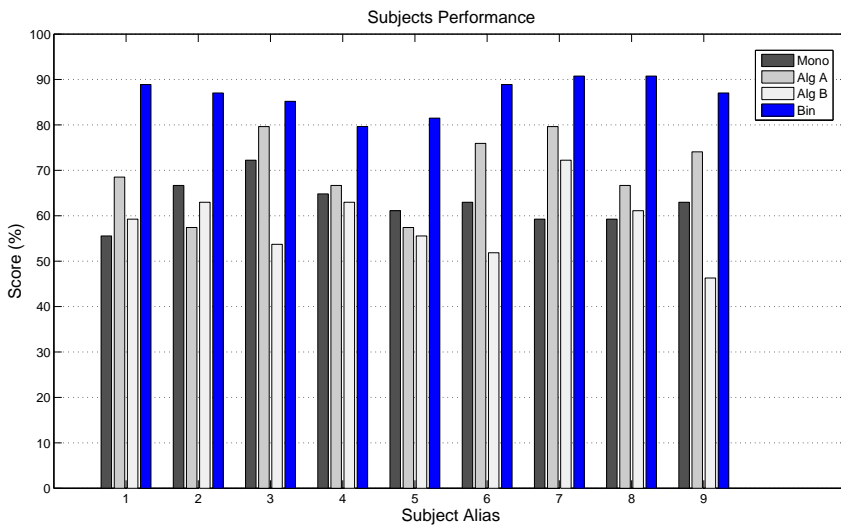


Figure 5.2: Total correct answers (in %) per subject for every situation.

The other two scenarios show higher scores, indicating that the use of different masking signals can highly affect the number of correct answers, as configuration B for instance is considered more adverse due to the proximity of the masker to the listener. The most

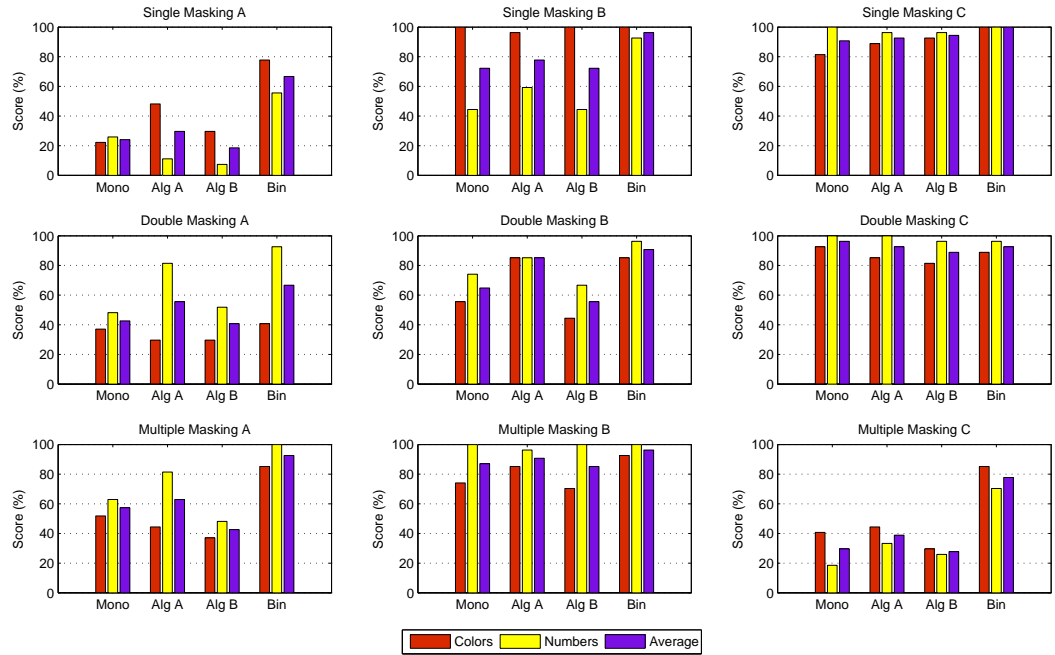


Figure 5.3: Scenarios.

favorable case, scenario containing Single Masking C, shows the higher scores among the studied cases. This is supposedly due to the fact that the head shadow has a positive effect for this type configurations where the masker is placed in the side of the impaired ear.

In the case where two interferers were employed more expected results were obtained. Configuration A in this group of scenarios yields the lowest scores. This result might partly be due to the maskers speech direction, as it was the only case inside this group where the maskers were facing the listener. In the other two configurations the layout of the maskers speech directions was designed following the principle of realistic configurations. This issue affects in a smaller degree the overall listening experience of the listener, as results of Double Masking B and C show. Same way as in the configuration with a single interfering source, in the case where two maskers in the impaired side were employed the highest score was obtained.

Last group of scenarios, four maskers together with the target, shows the same problem stated before of the signals synchronization, as configuration A was expected to score higher than B. Case C in this group performs as expected as the layout containing the four maskers in the unimpaired ear was considered as the most adverse situation among the ones presented in this group.

Single Masking C, Double Masking C and Multiple Masking B present a close to 100%

of correctly scored answers. This is unfavourable since a hypothetical improvement of the algorithms can hardly be reflected. Nonetheless, this fact is the result of a compromise to set the correct answers per group around 50% (see Appendix D), since it was decided to keep the same SNR for each group.

Summarizing, a more thorough study of the mentioned signal synchronization could yield more reasonable results. In any case, these unwanted effects on the results (i.e., scenarios where the color or number were too masked or unmasked) are not too harmful for extracting conclusions. What is actually important is to extract the relative changes among the situations, which were equally affected by these phenomena.

5.1.3 Group of Scenarios

Figure 5.4 depicts the results classified into the three different groups.

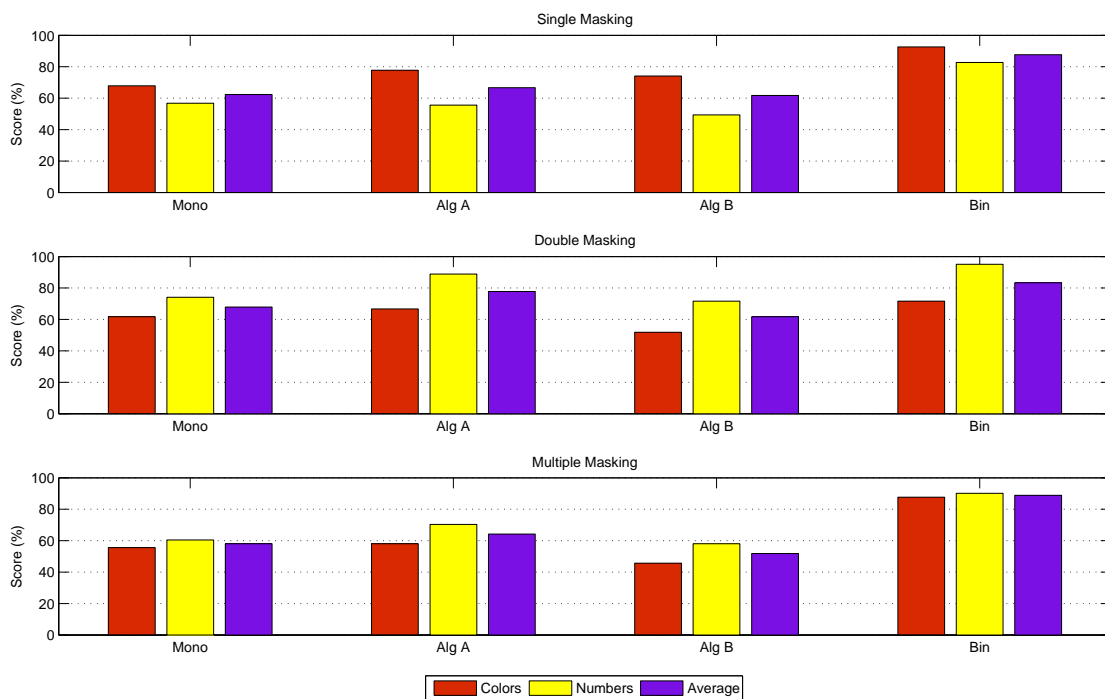


Figure 5.4: Group of scenarios.

Direct comparisons among the different groups cannot be drawn as each of them was presented with different SNR, as it was desired to obtain approximately 50% of correct answers, as explained in Appendix D. Nevertheless, no great differences among the groups can be seen, thus yielding a similar level of difficulty.

5.1.4 Situations

The different situations were presented to the subject one after each other, thus being possible to ask for an assessment of their naturalness. Table 5.1 shows the grades, between 0 and 10, given by all the subjects. In figure 5.5 the global performance of each group is presented, as function of the total percentage of correct answers, as well as the corresponding mean naturalness mark.

Subject Alias	Monaural	Algorithm A	Algorithm B	Binaural
1	4	6	7	8
2	8	7	5	9
3	8	8	7	9
4	6	7	8	9
5	7	5	7	8
6	7	6	6	9
7	9	7	3	9
8	7	5	7	8
9	7	8	6	9
Mean	7.0	6.6	6.2	8.7

Table 5.1: Naturalness grades given to each situation. From 0 to 10

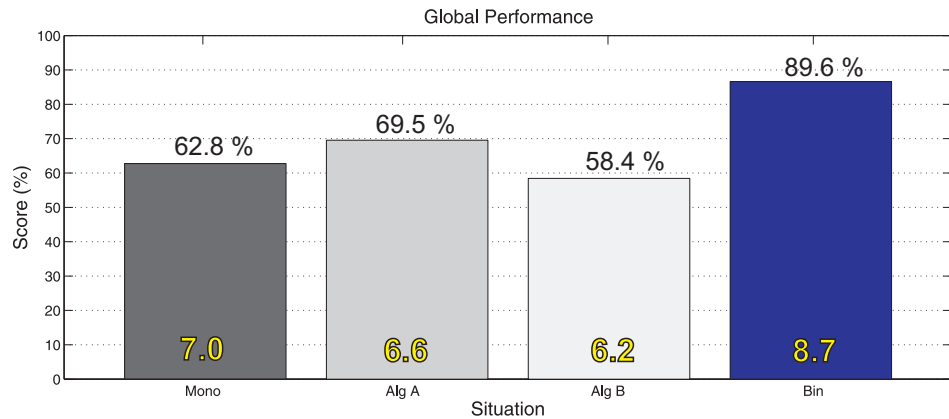


Figure 5.5: Global performance per situation. Mean naturalness given by the subjects presented in yellow inside the bars.

In one hand it can be seen the improvement that Algorithm A (69.5%) achieves compared to the unaided situation (62.8%). On the other hand, Algorithm B (58.4%) shows a slight deterioration. The binaural case study, which was used as control to show the correct

operation of the test yielded a score of 86.6%, showing the importance of binaural hearing for speech intelligibility. Regarding the naturalness of each situation, results follow the expected tendency. Binaural hearing obtains the best mark, opposite to Algorithm B, which loss of naturalness was known by forehand (due to the produced artifacts). Meanwhile, Algorithm A slightly decreases with respect to the Monaural presentations. Nonetheless, the relevance of these results is thoroughly assessed in section 5.2.

5.2 Analysis

In this section a statistical analysis of the before presented data is performed. In the end of it the conclusion of the current study will be drawn.

5.2.1 Interactions

First of all, the possible interaction between situations and scenarios is evaluated. This is, e.g., if Algorithm A performs better for Double Masking than for any other group. In case no interactions are found, situation effect can be assessed. When analysis of variance is applied (see table 5.2) for the three groups (once the Binaural situation is disregarded), there is no sign of any interaction with respect to the algorithm (p-value of 0.1286).

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
X1	0.1867	1	0.18673	2.33	0.1286
X2	0.3933	2	0.19667	2.46	0.0888
X1*X2	0.0216	2	0.0108	0.14	0.8738
Error	12.4774	156	0.07998		
Total	13.079	161	8		

Table 5.2: 2-way Anova with interactions: involving Situations (X1) and Scenarios (X2).

5.2.2 Situation Effect

At this point the performance of the algorithms is finally assessed. Even though a look at the raw results seems that Algorithm A induces certain improvement in the intelligibility while Algorithm B worsens it, analysis of variance is still needed in order to study the relevance of these changes.

Consequently, 2-way ANOVA is applied, where the independent variables are scenarios and situations. Scenarios instead of groups were selected as independent variables in order to have a more global view of the matter under study. For the assessment of Algorithm A, Binaural and Algorithm B situations are disregarded, as the aim is to evaluate

the unaided situation versus the proposed Algorithm A uniquely. A p-value of 0.0075 is obtained (see table 5.3), clearly below the recommended 5% level of significance [Hicks99]. Therefore, the efficiency of Algorithm A can be ensured.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
X1	0.1867	1	0.18673	7.34	0.0075
X2	9.025	8	1.12813	44.34	0
Error	3.8673	152	0.02544		
Total	13.079	161			

Table 5.3: 2-way Anova involving: Situations Monaural and Algorithm A(X1); and Scenarios (X2).

A similar analysis is performed over Algorithm B (disregarding Binaural and Algorithm A situations), and the resulting p-value appears to be 0.0784 (see table 5.4). Assuming the same level of significance, it can not be fully stated that this solution worsens the intelligibility.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
X1	0.0756	1	0.07562	3.12	0.0794
X2	10.559	8	1.31987	54.44	0
Error	3.6852	152	0.02424		
Total	14.3198	161			

Table 5.4: 2-way Anova involving: Situations Monaural and Algorithm B(X1); and Scenarios (X2).

In the box plot shown in figure 5.6 the variance of the different situations can be evaluated. Algorithm A and specially the Binaural situation present a non normally distributed data. This is due to the fact that the results get closer to the 100% of correct answers, occurring then a saturation effect around at the upper part. Meanwhile, Algorithm B and the Monaural condition show a closer to normal distribution.

A further analysis consisting in evaluating the scenarios per group is analyzed next so as to check if any of the algorithms performs better for any of the scenarios in each group.

First the scenarios with a single interferer are analyzed in two different two-way ANOVA, the first comparing Monaural and Algorithm A situations, and the second doing the same with the Monaural and Algorithm B conditions. Tables 5.5 and 5.6 show that a p-value of 0.2562 and 0.8614 respectively for the approaches under study, suggesting that no significance improvement happens in single masking scenarios with Algorithm A, while no significance deterioration occurs when using Algorithm B for this group of scenarios.

In the evaluation of situations for two maskers the same procedure as before is performed. According to tables 5.7 and 5.8 Algorithm A shows a significant improvement (p-value of

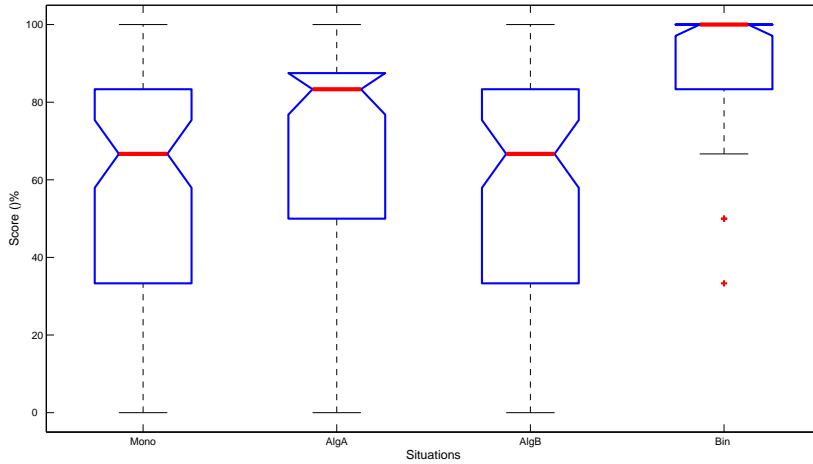


Figure 5.6: Box plot showing the distribution of data among the situations.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
X1	0.02519	1	0.02519	1.32	0.2562
X2	4.07801	2	2.039	106.79	0
Error	0.95464	50	0.01909		
Total	5.05784	53			

Table 5.5: 2-way Anova involving: Situations Monaural and Algorithm A(X1); and Single Masking Scenarios (X2).

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
X1	0.00051	1	0.00051	0.03	0.8614
X2	4.8546	2	2.4273	145.47	0
Error	0.83431	50	0.01669		
Total	5.68942	53			

Table 5.6: 2-way Anova involving: Situations Monaural and Algorithm B(X1); and Single Masking Scenarios (X2).

0.0287), while Algorithm B (p-value of 0.1855) seems not to deteriorate significantly the speech intelligibility under the situation of two simultaneous maskers.

Last group of study is the one with multiple maskers, and tables 5.9 and 5.10 show that no significance difference occur for the improvement of Algorithm A (p-value of 0.2099) neither for the deterioration of Algorithm B (p-value of 0.1808).

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
X1	0.13168	1	0.13168	5.08	0.0287
X2	1.86506	2	0.93253	35.94	0
Error	1.29734	50	0.02595		
Total	3.29408	53			

Table 5.7: 2-way Anova involving: Situations Monaural and Algorithm A(X1); and Double Masking Scenarios (X2).

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
X1	0.05144	1	0.05144	1.18	0.1855
X2	2.39183	2	1.19592	41.91	0
Error	1.42689	50	0.02854		
Total	3.87016	53			

Table 5.8: 2-way Anova involving: Situations Monaural and Algorithm B(X1); and Double Masking Scenarios (X2).

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
X1	0.05143	1	0.05143	1.61	0.2099
X2	2.68809	2	1.34404	42.17	0
Error	1.59357	50	0.03187		
Total	4.33309	53			

Table 5.9: 2-way Anova involving: Situations Monaural and Algorithm A(X1); and Multiple Masking Scenarios (X2).

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
X1	0.05144	1	0.05144	1.84	0.1808
X2	3.03169	2	1.51584	54.29	0
Error	1.396	50	0.02792		
Total	4.47913	53			

Table 5.10: 2-way Anova involving: Situations Monaural and Algorithm B(X1); and Multiple Masking Scenarios (X2).

5.2.3 Naturalness Effect

Whether or not Algorithm A and B did change the naturalness impression of the utilized subjects is evaluated by running two ANOVAs. First, just Monaural and Algorithm A columns are considered, which yields a p-value of 0.4028 (see table 5.11). Therefore Algorithm A does not deteriorate significantly the naturalness impression compared with the Monaural situation.

When the same process is carried out with Algorithm B, the p-value provides a very

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
X1	0.8889	1	0.8889	0.78	0.4028
X2	17.1111	8	2.13889	1.88	0.1957
Error	9.1111	8	1.13889		
Total	27.1111	17			

Table 5.11: 2-way Anova involving: Situations Monaural and Alborithm A (X1); and Subjects (X2).

similar result: 0.4018 (see table 5.12). Thus, it can not be fully ensured that Algorithm B deteriorates the subjects naturalness impression.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
X1	2.7222	1	2.72222	0.78	0.4018
X2	5.7778	8	0.72222	0.21	0.9802
Error	27.7778	8	3.47222		
Total	36.2778	17			

Table 5.12: 2-way Anova involving: Situations Monaural and Alborithm B (X1); and Subjects (X2).

5.2.4 Subject Effect

Additionally, the study of subject effect is also assessed, as it was found interesting to check to what extent the selected population was appropriate. This time a 3-way ANOVA, involving situations, scenarios and subjects, was applied. The binaural situation was not considered as a high number of correct answers were obtained and it was believed that this issue may affect the fairness of the subject effect study. The result rejects the hypothesis of such effect, since the obtained p-value is 0.8722 (see table 5.13). Therefore, subjects did not perform very different from each other, which is an indicator of a suitable sample of individuals.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob > F
X1	0.5082	2	0.25412	3.04	0.0499
X2	0.5	2	0.25	2.99	0.0524
X3	0.3189	8	0.03987	0.48	0.8722
Error	19.2469	230	0.08368		
Total	20.5741	242	8		

Table 5.13: 3-way Anova involving Situations (X1), Scenarios (X2) and Subjects (X3).

Conclusions

The reduced skills of monaural listeners in terms of speech intelligibility was the motivating thread for this project. The lack of two healthy auditory inputs prevents them from a reliable analysis of the binaural cues. The processing of these cues allows what is known as cocktail party effect, i.e., the ability of focusing the attention on a single target within a noisy environment. To accost this problem, solutions based on the computation of interaural differences were searched. Thus, two different proposals were developed and implemented, both based on the computation of the Interaural Time Differences. The so-called Algorithm A consists in a simple approach which takes advantage of the shadow effect of the head, thus selecting the channel with higher SNR at each moment. Algorithm B was inspired on the notes from a previous work though, and it aims to attenuate the noise coming from lateral directions. Nevertheless, even though the main philosophy is the same, the processing varied considerably.

Both proposals were off-line assessed by means of a listening test, in which their efficiency was compared with respect to an unaided situation. So as to facilitate the selection, binaural subjects were used, whose monaural perception was simulated. Surprisingly, results led to a slight deterioration (around 4%) of the intelligibility when the Algorithm B was applied. Nonetheless, thorough statistical analysis of variance did not indicate this fact as sufficiently significant. However, Algorithm A achieved an intelligibility improvement which rounded the 7%. Moreover, the relevance of this result was corroborated by the analysis of variance. Subjects were also asked to rate the naturalness of the speech for each situation. Even though both algorithms obtained slightly worse marks, no significance difference compared to the unaided situation was found. As a control test, binaural presentations were likewise evaluated, which yielded in a clear increase of the percentage of correct answers. This tendency indicates that the listening test was suitably carried out.

Some of the scenarios provided too high scores, quite close to the 100%. This might have resulted in unfavourable results, since the margin of improvement for the algorithms was notably reduced. The reasons consisted on a too unmasked message, issue which was tried to be solved after noticing it in the pilot test, but that in practise could not be fully eradicated. A more thorough preparation of the target and masker signals would then determine whether or not Algorithm A could still improve the achieved performance. On the other extreme, some scenarios seemed to be quite difficult for the participating subjects. However, this fact is not so critic since unaided and aided situations were affected in the same way, and relative changes were those actually studied.

Algorithm B was applied in such manner that naturalness was wanted to be preserved, so it was not possible to operate with high attenuation values. Therefore, its implementation ought be revised: if artifacts can be reduced, stronger attenuation to the lateral energy would be possible, thus supposedly leading to an increase of the intelligibility. In any case, the promising results obtained by Algorithm A suggest that is worthy to keep on working on its design and testing its performance over real monaural listeners. Due to its simplicity and consequent low computational load, an eventual real time implementation seems quite attractive and feasible.

Future Work

Further ideas which were considered suitable for future improvements related to the current research are shown next.

- Use of actual monaural subjects to evaluate the promising first approach (Algorithm A).
- Real time implementation of Algorithm A.
- A more thorough setting of the masker(s) and target synchronization, aiming to equally mask every synthesized sample.
- Increase the system sampling frequency so as to increase the resolution for the ITDs computation.
- Combine attenuation with amplification in Algorithm B.
- Assess Algorithm B with higher attenuation, despite to expected loss of naturalness.
- Make use of Interaural Level Differences for the high frequencies.
- Calculate ITDs by means of its phase rather than by cross-correlating signals.
- Adjust the attenuation function of Algorithm B by different ways. For instance, to multiply in time each output of the bankfilter by the corresponding gain factor, then add all the signals to reconstruct the corresponding window.

Bibliography

- [Batteau67] D. W. Batteau, 'The role of the pinnae in human localization', Proc. Roy. Soc., B 168, 1967.
- [Blauert05] J. Blauert, 'Communication Acoustics', Springer, 2005.
- [Blauert97] J. Blauert, 'Spatial Hearing: The Psychoacoustics of Human Sound Localization', The MIT Press, 1997.
- [BoliaEtAl00] R. S. Bolia, 'A speech corpus for multitalker communications research', Journal Society of America, vol.107(2), 2000.
- [Bronkhorst00] A. W. Bronkhorst, 'The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions', Acta Acustica, vol.86, 2000.
- [Brungart01] D. S. Brungart, 'Evaluation of speech intelligibility with the coordinate response measure', Journal Society of America, vol.109(5), 2001.
- [BuserAndImbert92] Pierre Buser, Michel Imbert, 'Audition', The MIT Press, 1992.
- [Butler69] R. A. Butler, 'Monaural and binaural localization of noise burst vertically in the median sagittal plane', Journal of Medical Research, 3, 1969.
- [Cherry53] C. Cherry, 'Some experiments on the recognition of speech, with one and two ears', Journal of Acoustic Society of America, vol.25, 1953.
- [DíezAndChristensen06] Y. Díez Moral, E. T. Christiansen, 'Aiding Monaural Listeners by Utilizing Binaural Cues', Master Science Thesis, Department of Acoustics, Aalborg University, 2006.
- [Dillon01] H. Dillon, 'Hearing Aids', Thieme, 2001.
- [DISR03] The Danish Institute for Social Research, 'When hearing fails: Impact of hearing loss on work, education and personal health', 2003.
- [EricsonEtAl88] H. Ericson, I. Svård, O. Högset, G. Devert, L. Ekström, 'Contralateral routing of signals in unilateral hearing impairment - a better method of fitting', Scandinavian Audiology, vol.17, 1988.
- [FreedmanAndFisher68] S. J. Freedman, H. G. Fisher, 'The role of the pinna in auditory localization', Neuropsychology of Spatially Oriented Behaviour, Dorsey Press, Illinois (USA), 1968.

- [GardnerAndGardner73] M. B. Gardner, R. S. Gardner, 'Problem of localization in the median plane:effect of pinnae cavity occlusion', *Journal of the Acoustical Society of America*, vol.53, 1973.
- [HawleyEtAl99] M. L. Hawley, R. Y. Litovsky, H. S. Colburn, 'Speech intelligibility and localization in a multi-source environment', *Journal of the Acoustical Society of America*, vol.105, 1999.
- [HearIt09] <http://www.hear-it.org/page.dsp?page=430>
- [Hicks99] C. R. Hicks, K. V. Turner Jr., 'Fundamental Concepts in the Design of Experiments', Oxford University Press, New York (USA), 1999.
- [Hirsh71] I. J. Hirsh, 'Masking of speech and auditory localization', *Audiology* 10, 1971.
- [Koehnke94] J. Koehnke, 'Binaural Performance in Listeners With Impaired Hearing: Aided and Unaided Results', University of South Alabama (USA), 1994.
- [KollmeierEtAl93-1] B. Kollmeier, J. Pessig, V. Hohmann, 'Binaural Noise-Reduction Hearing Aid Scheme with Real Time Processing in the Frequency Domain', *Scandinavian Audiology Suppl.*38: 28-38, 1993.
- [KollmeierEtAl93-2] B. Kollmeier, J. Pessig, V. Hohmann, 'Real Time multiband dynamic compression and noise reduction for binaural hearing aids', *Scandinavian Audiology, Journal of Rehabilitation Research and Development*, vol.30vNo. 1, 1993.
- [Kollmeier94] B. Kollmeier, 'Signal Processing for Hearing Aids Employing Binaural Cues', Universität Oldenburg (Germany), 1994.
- [McPartland97] J. L. McPartland, J. F. Culling, D. R. Moore, 'Changes in lateralization and loudness judgements during one week of unilateral ear plugging', *Hearing Research*, Vol. 113, 1997.
- [Meyer09] <http://www.meyersound.com/support/papers/speech/mrt.htm>
- [Middlebrooks99] J. C. Middlebrooks, 'Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency', *Journal of the Acoustical Society of America*, 106, 1999.
- [Mills72] A. W. Mills, 'Auditory localization in Foundations of Modern Auditory Theory', Vol. II, J. V. Tobias, Academic Press, New York, 1972.
- [Moore04] Brian C. J. Moore, 'An Introduction to the Psychology of Hearing', Elsevier Ltd., 2004.
- [Moore81] T. J. Moore, 'Voice communications jamming research', *AGARD Aural Commun. in Aviation*, N81-31449 22-32, 1981.

- [Møller92] H. Møller, 'Fundamentals of binaural technology', Institut for Elektroniske Systemer, Aalborg University, 1992.
- [Møller94] H. Møller, D. Hammershøi, C. B. Jensen, M. F. Sørensen, 'Transfer characteristics of headphones measured in human ears', Audio Engineering Society, Vol. 43, No. 4, 1995.
- [Oppenheim89] A.V. Oppenheim, 'Discrete-time signal processing', Prentice Hall, 1989.
- [Owens93] F.J. Owens, 'Signal Processing of Speech', Macmillan New Electronic Series, 1993.
- [PerssonEtAl01] P. Persson, H. Harder, S. Arlinger, B. Magnuson, 'Speech recognition in background noise: Monaural versus binaural listening conditions in normal-hearing patients', *Otology And Neurology*, vol.22, 2001.
- [PoissantEtAl06] S. F. Poissant, N. A. Whitmal, R. L. Freyman, 'Effect of reverberation and masking on speech intelligibility in cochlear implant simulations', *Journal of the Acoustical Society of America*, vol.119, 2006.
- [Poulsen05] T. Poulsen, 'Acoustic Communication, Hearing and Speech', DTU Lecture Note 31230-05 , 2005.
- [Sammeth99] C. A. Sammeth, 'The role of consonant-vowel amplitude ratio in the recognition of voiceless stop consonants by listeners with hearing impairment', *Journal of Speech, Language, and Hearing Aids*, vol.42, 1999.
- [StevensAndNewman36] S. Stevens, E. Newman, 'The localization of actual sources of sounds', *American Journal of Psychology*, 48, 1936.
- [Titze94] I.R. Titze, 'Principles of Voice Production', Prentice Hall, 1994.
- [Vestergaard04] M. D. Vestergaard, 'Benefit from amplification of high frequencies in hearing impaired: aspects of cochlear dead regions an auditory acclimatization', Technical University of Denmark, 2004.
- [Wenzel93] E. M. Wenzel et al., 'Localization using non-individualized head-related transfer functions', *Journal of the Acoustical Society of America*, 94, 1993.
- [Widex09] <http://www.widex.com>
- [Yost94] W. A. Yost, 'The Cocktail Party Problem: Forty Years Later', Loyola University of Chicago (USA), 1994.

Appendix A

Detailed Results

The following table contains all the subject's percentage of correct answers from the listening test, for the different situations (Mono, Algorithm A, Algorithm B, Binaural) and groups of scenarios (Single, Double and Multiple Masking). Scores in the same row correspond to the same subject. Each square is the result of averaging the three answers each subject gave per scenario, considering the mean between color and number guesses.

		Single Masking			Double Masking			Multiple Masking		
		a	b	c	a	b	c	a	b	c
M	O	0.17	0.83	0.83	0.17	0.83	0.83	0.33	0.83	0.17
		0.00	0.83	1.00	0.50	0.67	1.00	0.83	1.00	0.17
		0.67	0.67	1.00	0.83	0.33	1.00	0.67	1.00	0.33
		0.33	0.67	0.83	0.67	0.50	1.00	0.67	0.83	0.33
		0.17	0.67	1.00	0.17	0.67	1.00	0.83	0.83	0.17
		0.50	0.67	0.83	0.17	0.83	1.00	0.67	0.83	0.17
		0.17	0.67	0.83	0.50	0.67	0.83	0.50	0.67	0.50
		0.00	0.83	1.00	0.33	0.50	1.00	0.50	1.00	0.17
		0.17	0.67	0.83	0.50	0.83	1.00	0.17	0.83	0.67
A		0.33	0.67	1.00	0.50	0.83	0.83	0.50	1.00	0.50
		0.17	0.5	0.67	0.33	1.00	1.00	0.50	1.00	0.00
		0.50	0.83	1.00	0.67	0.83	1.00	0.83	1.00	0.50
		0.00	0.83	0.83	0.50	0.83	1.00	0.67	0.83	0.50
		0.17	0.83	0.83	0.67	0.50	1.00	0.50	0.67	0.00
		0.33	0.83	1.00	0.50	1.00	0.83	0.83	0.83	0.67
		0.50	0.83	1.00	0.67	1.00	0.83	0.83	1.00	0.50
		0.33	0.83	1.00	0.67	0.67	0.83	0.33	0.83	0.50
		0.33	0.83	1.00	0.50	1.00	1.00	0.67	1.00	0.33
B		0.17	0.83	0.83	0.50	0.67	0.83	0.67	0.67	0.17
		0.00	0.67	1.00	0.33	0.67	1.00	0.67	1.00	0.33
		0.33	0.67	0.83	0.50	0.33	1.00	0.33	0.67	0.17
		0.17	0.67	0.83	0.67	0.67	1.00	0.50	0.83	0.33
		0.17	0.67	1.00	0.50	0.50	0.83	0.33	0.83	0.17
		0.17	0.83	1.00	0.17	0.50	0.83	0.17	0.83	0.17
		0.50	0.67	1.00	0.50	0.83	1.00	0.67	1.00	0.33
		0.17	0.83	1.00	0.33	0.67	0.67	0.33	1.00	0.50
		0.00	0.67	1.00	0.17	0.17	0.83	0.17	0.83	0.33
B	I	0.83	0.83	1.00	0.67	1.00	1.00	0.83	1.00	0.83
		0.67	1.00	1.00	1.00	0.83	0.83	1.00	1.00	0.50
		0.67	1.00	1.00	0.50	0.83	1.00	1.00	1.00	0.67
		0.67	0.83	1.00	0.33	1.00	0.67	1.00	1.00	0.67
		0.67	1.00	1.00	0.50	1.00	0.83	0.83	0.83	0.67
		0.67	1.00	1.00	0.83	0.83	1.00	0.83	1.00	0.83
		0.67	1.00	1.00	0.67	0.83	1.00	1.00	1.00	1.00
		0.5	1.00	1.00	0.83	1.00	1.00	0.83	1.00	1.00
		0.67	1.00	1.00	0.67	0.83	1.00	1.00	0.83	0.83

Figure A.1: Every subjects' results per situation and scenario.

PTFs Measurement Report

Every time a listening test is performed by means of headphones, their frequency response must be compensated. Thus, the headphone transfer functions, so-called PTFs (one per channel), are deconvolved with the test signals. This appendix describes how the PTFs were obtained as well as how they were applied in order to avoid their influence.

In [Møller94] a number of headphones are assessed, where the model Beyerdynamic DT990 provided one of the flattest frequency response. Besides, the availability of them at AAU facilities led to its use for the listening tests.

B.1 Procedure

With a view to obtain generic PTFs applicable to any subject, it was decided to use Valdemar, the manikin developed by the Acoustic Department at the AAU, to record binaural responses. The desired impulse responses were obtained by cross-correlating MLS sequences with the recorded microphone outputs. All the process was performed by means of the computer-based measuring system Harmonie. Both channels (L and R) were measured five times, after reposition of the headphones, so as to obtain an averaged response (as performed in [Møller94]). A larger number of measurements could be counterproductive, since the final signal would result too smooth and details would be lost.

B.1.1 Setup

The setup is shown in figure B.1. Harmonie provides a module specially thought for measuring impulse responses: dBFA32. From it, MLS sequences were sent to the headphones. Following the procedure in [Møller94], a SPL about 80-90 dBA was set, reason why an amplifier was interconnected in between. The microphones, calibrated previously to the recordings, were connected to a phantom source after the preamplifiers. The measurements were performed at the Multichannel Listening Room, within the laboratory facilities of the Acoustic Department at the AAU.

The MLS sequence was set to 16th order (which implies $2^{16} - 1$ points). Since Harmonie works with a sample frequency of 51.2 kHz, this configuration leads to a response of 1280 ms, long enough to cover the impulse responses. An averaged response was obtained over 16 sequences. The amplifier was adjusted to provided 85 dBA SPL through the headphones.

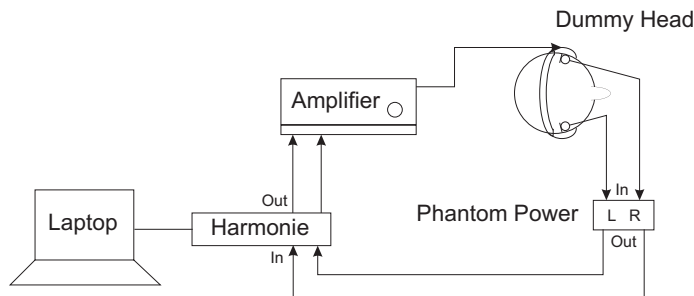


Figure B.1: PTFs measurement setup

B.1.2 Equipment list

Table B.1 lists the devices used along the measurement.

DEVICE	Manufacturer	MODEL	SERIAL NUMBER
Headphones	Beyerdynamic	DT990	aau2036-12
Manikin	AAU	Valdemar Sejr	aau2150-03
Left microphone	Gras	40AD	aau56521
Right microphone	Gras	40AD	aau56520
Phantom source	Neumann	BS48i-2	aau2018-00
Amplifier	Sony	STR-DB790	aau56555-00
Measuring system	01dB	Harmonie	aau56524
Laptop	Siemens	E-series Lifebook	aau60921

Table B.1: List of devices used for the measurement of the PTFs.

B.1.3 Post-processing

By means of Matlab, an average among the five repositions per ear was performed. Next, Linear Predictive Coding (LPC) method was applied to both averages. This method provides the coefficients of the FIR filter to be applied, which can be referred to as the inverse filter of the PTFs. Once this filter is applied the headphones response is compensated.

The full sets of the measured PTFs per ear are plotted in figures B.2 and B.3, whereas the final averages are shown in B.4. In order to check the performance of the inverse filter, figure B.5 shows the results of applying the applied inverse filter. It can be noted that the filters, even though they provide a totally flat output, a notable attenuation was induced. This was easily solved by multiplying by 3 the filters, thus being possible to reproduce at the expected level.

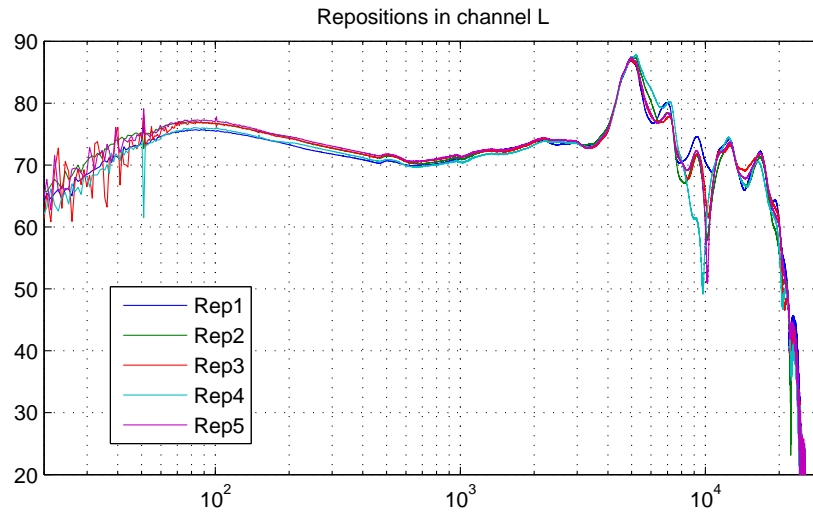


Figure B.2: The five PTFs measured on left ear.

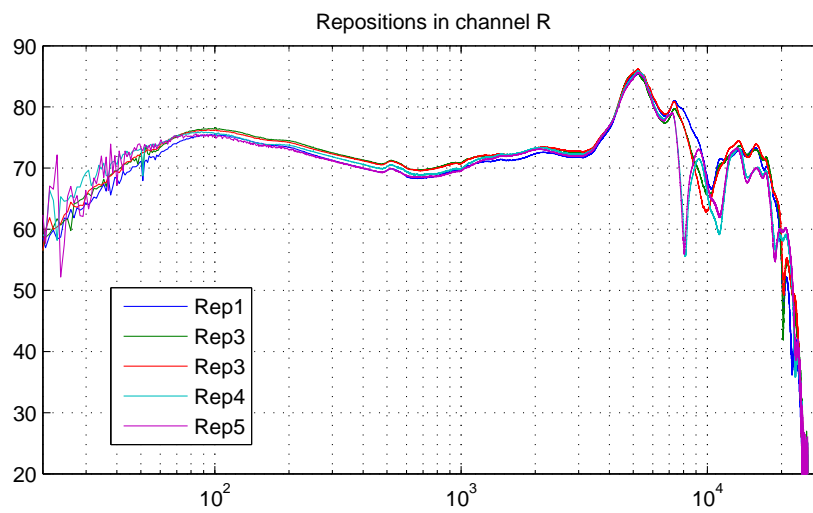


Figure B.3: The five PTFs measured on right ear.

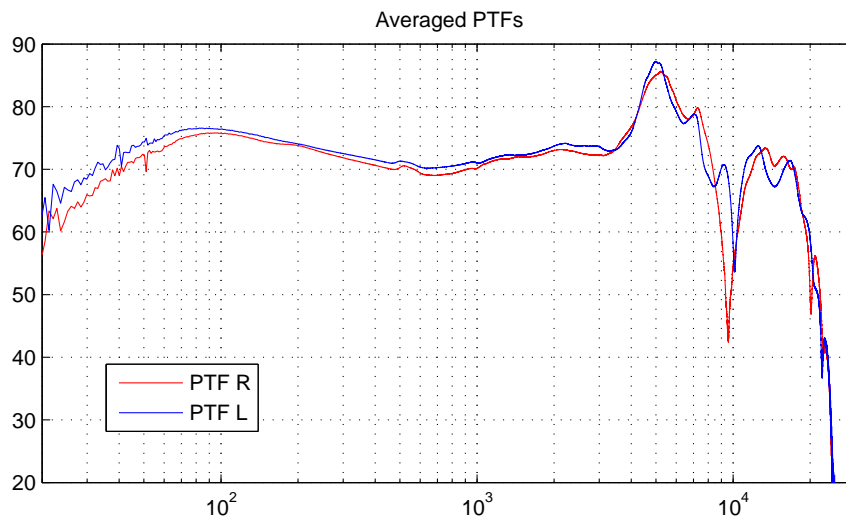


Figure B.4: Averaged PTFs for both channels.

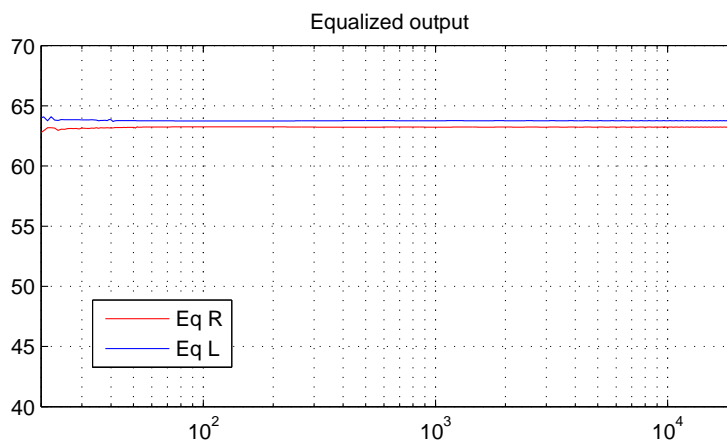


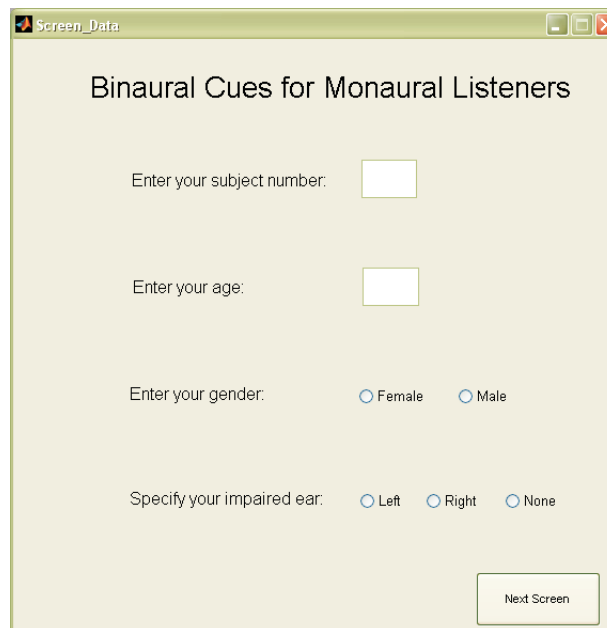
Figure B.5: Output applying the inverse filter.

Graphical User Interface

A user friendly graphical interface developed in *Matlab* environment was created to guide the subjects performing the listening test throughout the last. All their answer were saved as the sessions were performed and later were extracted to analyze the results.

This interface consists in different screens and each of them is shown and explained briefly next.

First screen is the one entitled to gather all the data of the subject participating in the test. An alias which will later determine the order of presentation of the recorded samples, together with the age, gender and most sensitive ear (if any) are questions that have to be answered before passing to the next screen. Find a screen shot if the mentioned screen in figure C.1.



The screenshot shows a window titled "Screen_Data" with a light beige background. The main heading is "Binaural Cues for Monaural Listeners". Below the heading, there are four rows of input fields:

- "Enter your subject number:" followed by a white text input box.
- "Enter your age:" followed by a white text input box.
- "Enter your gender:" followed by two radio buttons labeled "Female" and "Male".
- "Specify your impaired ear:" followed by three radio buttons labeled "Left", "Right", and "None".

A "Next Screen" button is positioned at the bottom right of the form area.

Figure C.1: Data Screen.

Next step for the user was to read the instructions carefully so as to prepare the subject for the upcoming listening test. Figure C.2.

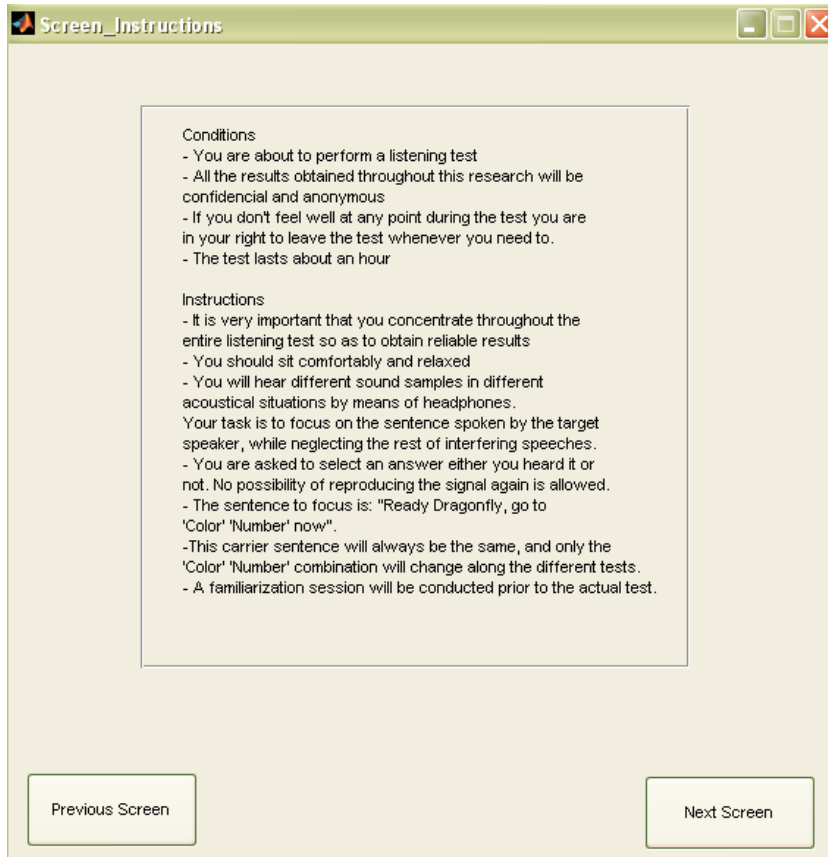


Figure C.2: Instructions Screen.

The first session is a familiarization session (see section 4.2.2) and a screen shot of it when the subject failed an answer is presented in figure C.3.

Last screen shot (figure C.4) is an example of the remaining screens (session 2 to 5), where the subject is asked to submit an answer based on what he or she heard in the previously presented audio sample.

Once all sessions have been presented the data of each session was saved in different *.mat* files for a later evaluation of the results.



Figure C.3: Screen Session 1 (Familiarization Session).

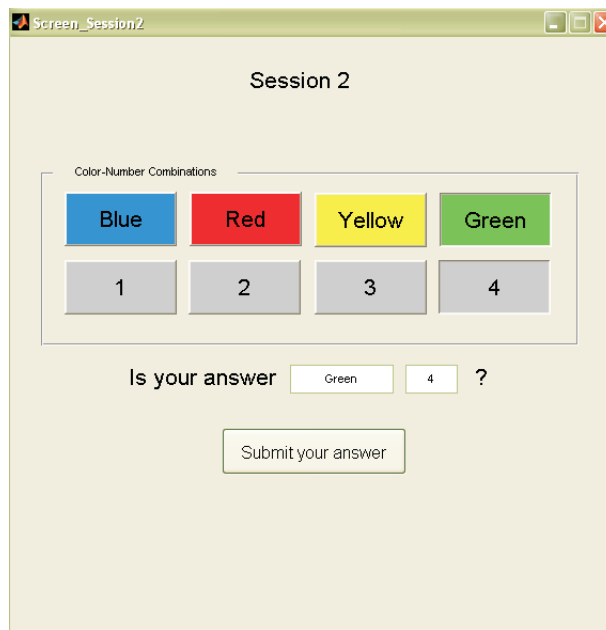


Figure C.4: Screen of Sessions 2 to 5.

Pilot Test

In order to yield with appropriate presentations a pilot test was performed.

The situation to evaluate these presentations was the monaural one, as this will be our reference throughout all the research. In order to see if actual changes between the unaided condition and the aided one occur, a score of 50% was desired per group of scenarios. This way an improvement, deterioration or indifference of the proposed approaches will be feasible and consequently a conclusion out of them will be drawn.

For this case study is of great importance to set a specific SNR per group of scenario. Different SNR per group were chosen due to the variability of guesses depending on the number of maskers. It was wanted to set the same SNR for the scenarios belonging to the same group for an evaluation of the results depending on the grouping.

Two subjects who do not belong to the present project group neither to the population sample selected for the later listening test were utilized to obtain these values. Table D.1 shows the scores of the different subjects for the different groups of scenarios (each with the first SNR approach).

	Single (SNR: -15 dB)	Double (SNR: -12 dB)	Multiple (SNR: -10 dB)
Subject 1	67%	44%	33%
Subject 2	72%	72%	44%
Average	69.5%	58%	38.5%

Table D.1: Pilot test scores.

Both subjects claimed that the difficulty to guess the color-number combination (mainly in single masking) was highly dependent on the interfering signal. In some cases the target's speech was fully masked by the masker speech, and vice versa, when the masker had a break (silence) in his/her speech the target speech was clearly audible and recognizable. Because of this issue it was decided to shift the target's sequence of the affected cases, aiming to similarly mask both words for all the scenarios.

At the same time different SNR values were chosen for every group of scenarios, as it was desired to achieve scores closer to 50%. Single and double masking SNR were increased and multiple masking SNR was decreased so as to approximate these results to scores of

50%. The new results with the mentioned SNR per subject and group of scenarios are presented in table D.2.

	Single (SNR: -20 dB)	Double (SNR: -15 dB)	Multiple (SNR: -8 dB)
Subject 1	61%	44%	44%
Subject 2	67%	67%	56%
Average	64%	55.5%	50%

Table D.2: Pilot test scores.

These results were found appropriate for the upcoming listening test. Find the chosen SNR per group of scenarios next:

- Single Masking: -20 dB SNR.
- Double Masking: -15 dB SNR.
- Multiple Masking: -8 dB SNR.

The choices made in this section were later reaffirmed (mainly for single and multiple masking) according to the obtained results of the monaural condition for the nine subjects used in the listening test, as shown in figure D.1, which shows scores of 62%, 68% and 58% for single, double and multiple masking respectively.

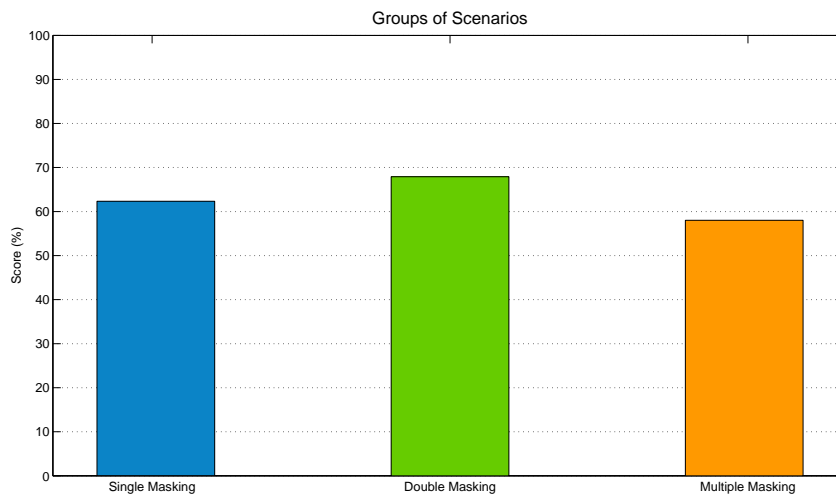


Figure D.1: Group of Scenarios.

These pilot test employed just two subjects, and a greater number of subject would have been desirable to set more precise SNR values for the designed test, as this issue is of

great importance for the overall performance of the test.