UNIVERSITY OF Southampton

University of Southampton Research Repository
ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

http://eprints.soton.ac.uk

University of Southampton

2014

FACULTY OF ENGINEERING AND THE ENVIRONMENT

INSTITUTE OF SOUND AND VIBRATION RESEARCH

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Design of an Array-Based Aid for the Hearing Impaired

by

Marcos Felipe Simón Gálvez

Abstract

University of Southampton

Faculty of Enginnering and the Environment
Institute of Sound and Vibration Research

Doctor of Philosophy

Design of an Array-Based Aid for the Hearing Impaired

by Marcos Felipe Simón Gálvez

The performance of our hearing system degrades with age, causing a reduction in our ability to understand speech. This thesis describes the design of a personal audio system, aimed at enhancing the audio signal from a TV, for hard of hearing listeners. Such system generates a zone of high acoustic pressure in a specific area of a room, whilst minimising the sound radiated to other zones.

In order to obtain a very narrow sound radiation, superdirective techniques are used. Such techniques improve the performance of an array of small dimensions at mid and low frequencies, but they can require the use of a large amount of power and can be very sensitive to errors in the source transfer functions. A practical balance between performance and array effort is achieved in a design that uses phase-shift sources. Two arrays have been built. The first array uses 8 phase-shift sources and provides a high directionality in the free-field. The second array uses 4 phase-shift sources vertically plus 8 horizontally and is more directional in a 3D sense, hence reducing the reverberant field excitation. The design of superdirective filters has been considered in both time and frequency domains. The effect of the reverberant field in the performance of superdirective radiators has also been investigated, by first examining how the radiation pattern is modified, and then by observing how the reverberant field affects the robustness of the control. The performance has been assessed by means of off-line simulations with point source Green functions and measured transfer responses, and by real time measurements in free-field and reverberant environments.

The later chapters of the thesis consider the subjective assessment of arrays as an aid for speech intelligibility. The assessment has first been performed using an objective metric, in particular the speech transmission index (STI). The performance has also been assessed in a subjective study, in which 30 participants have judged the increase in intelligibility. A cochlear model has also been utilised to investigate equalisation methods that minimise masking and provide a better speech intelligibility, particularly for the hearing impaired. Although this has not been tested with behavioural experiments, STI simulations have predicted that it can lead to an intelligibility improvement.

A Felipe y Nieves

Contents

List of Figures

List of Tables

Declaration Of Authorship

I, Marcos Felipe Simón Gálvez, declare that the thesis entitled Design of an Array-Based Aid for the Hearing Impaired and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- part of this work has been published as

2012 Loudspeaker array for Family TV (Marcos F. Simón Gálvez, Stephen J. Elliott, Jordan Cheer), *In ICSV 19, Vilnius, July*, 2012.

2012 A superdirective array of phase shift sources (Marcos F. Simón Gálvez, Stephen J. Elliott, Jordan Cheer), *In The Journal of the Acoustical Society of America*, ASA, volume 132, pp. 746-756, 2012.

2013 Design and Implementation of a Car Cabin Personal Audio System(Jordan Cheer, Stephen J. Elliott, Marcos F. Simón Gálvez), *In J. Audio Eng. Soc*, volume 61, pp. 412-424, 2013.

2013 Hearing Impaired Cochlear Responses (Marcos F. Simón Gálvez, Stephen J. Elliott), *In ICA 2013, Montreal, June*, 2013.

2013 The design of a superdirective array in a room (Marcos F. Simón Gálvez, Stephen J. Elliott), *In AES 52nd conference in sound field control, Guilford, UK, September*, 2013.

2013 Conformación de haz en espacios reverberantes-Superdirective beamforming in rooms (Marcos F. Simón Gálvez, Stephen J. Elliott, Jordan Cheer), *In Congreso Tecniacustica, Valladolid, October*, 2013.

2013 Superdirective personal audio in rooms (Marcos F. Simón Gálvez, Stephen J. Elliott, Arthur Marker, Jordan Cheer), *In Reproduced Sound 2013, Manchester, October*, 2013.

2013 Technical Memorandum 1000: A lumped parameter model of the organ of Corti (Marcos F. Simón Gálvez, Stephen J. Elliott), ISVR, University of Southampton, 2013.

2014 The effect of reverberation on personal audio devices(Marcos F. Simón Gálvez, Stephen J. Elliott, Jordan Cheer), *In The Journal of the Acoustical Society of America*, ASA, volume 135, pp. 2654-2663, 2014.

2014 Personal audio loudspeaker array as a complementary TV sound system for the hard of hearing (Marcos F. Simón Gálvez, Stephen J. Elliott, Jordan Cheer), *In IEICE Transactions*, volume E97,(9), 2014.

2014 Superdirective time domain beamforming for personal audio applications (Marcos F. Simón Gálvez, Stephen J. Elliott, Jordan Cheer), *In EAA Forum Acusticum 2014 Conference, Krakow, Poland*, 2014.

2014 Using listening tests to demonstrate the subjective performance of a superdirective TV loudspeaker array (Stephen J. Elliott Arthur Marker, Marcos F. Simón Gálvez), *In EAA Forum Acusticum 2014 Conference, Krakow, Poland*, 2014.

Signed: .....

Date

Preface

This book contains the scientific findings which I have overcome during the 3 years of my PhD at the ISVR. Apart from the academic side and all the experience I have gained and learned during my stay in the signal processing and control group, numerous events have passed through my life. Being in an international university environment, one has the chance to meet numerous people. The interaction wich such individuals during my stay in Southampton has resulted in a lot of experiences and circumstances. Such facts have shaped my personality and the way I look at things, somehow, and I am aware that they are reflected in this work, despite its scientific character.

Apart from myself, this book has a long list of people that in one way or another have contributed their grain of sand:

- Stephen J. Elliott. The biggest thanks go to you. Thanks very much for having considered me, first for my MSc project, and then, as a PhD candidate. Thanks to all the ideas, suggestions, advices and corrections that you have given me during this time, and also for the patience you have had with my English.

- Jordan Cheer. Thanks also for the patience you have had with my English, and also for all the time we have spent discussing aspects of personal audio systems, apart from helping and teaching me to perform acoustic measurements.

- Arthur Marker. Thanks for your help when building the 4×8 array, when measuring its performance, and for the evaluation of the speech intelligibility improvement.

- Guanjiang Ni. A big thanks goes to you for the chats about fluid coupling in the cochlea.

- Daniel Fernandez Comesaña. Thanks for letting me use the Microflown, and for once, back in 2011, teaching me how to plot a directivity in Matlab.

- Cierrabares and related: Old Fox, Kostinha, Juanito, Goshad, Maltese Falcon, Pablo, César, Anie, Ruben, Sara, Nalleli, Txo, Joan, Hermano, Irene, Miguel, Simbi, Mo and Julia. Thanks for all the experiences we have shared during this time and for making my stay in Soton nicer.

- To the ones I have met/shared experiences with and are not here any more. Thanks for all the good things you have taught me in life.

- Felipe, Nieves and rest of the family. Thanks very much for financing my MSc studies when I came to England. Thanks papa for all the things that I have learned from you, and thanks mama for always having taken care of me.

- Marina. Thanks for you advice in many aspects, for backing up my grammar... and for being a source of inspiration and learning.

To all of you, thanks for everything, you are also part of this book, and I am sure I will remember it every time I look at it.

Acronyms

ACM acoustic contrast maximisation.

AE array effort.

AI articulation index.

AN auditory nerve.

ANOVA analysis of variance.

BKB Bamford-Kowal-Bench.

BM basilar membrane.

CA cochlear amplifier.

CF characteristic frequency.

CHL cochlear hearing loss.

CVC consonant vowel consonant.

DI directivity index.

DPOE distortion product otoacustic emissions.

EP endocochlear potentials.

ERB equivalent rectangular bandwidth.

FE finite elements.

FFT fast Fourier transform.

GUI guided user interface.

HCL hearing comfort levels.

HL hearing loss.

HRTF head related transfer functions.

IHC inner hair cell.

IR impulse response.

LSIF least squares inverse filters.

LTASS long-term average spectrum of speech.

MC mixed control.

MET mechano-electrical transduction.

MTF modulation transfer function.

OC organ of Corti.

OHC outer hair cell.

PA public address.

PB phonetically balanced.

PML perfectly matched layer.

PS phase-shift.

PTA Pure tone average.

RL reticular lamina.

SDOF single degree of freedom system.

SII speech intelligibility index.

SM scala media.

SNR signal-to-noise ratio.

SPL sound pressure level.

SRT speech reception threshold.

STI speech transmission index.

TM tectorial membrane.

WFS wave field synthesis.

Nomenclature

P_{ED} Pressure in the ear drum for the two port network outer-middle ear model

α_k, β_k Octave band weighting and redundancy factors, used in the calculation of the STI

β Regularization parameter

$\Delta\mathbf{Z}$ Uncertain component of a reverberant transfer function

Δ_r Relative observation difference from a control point between the original source and a replicated source

$\Delta_{\mathbf{Z}}$ Space-averaged uncertainties of a transfer function

Δ_{RL} Shear displacement of the RL

Δ_{TM} Shear displacement of the TM

γ Gain of the cochlear amplifier

λ Lagrange multiplier used in the acoustic contrast maximization solution. λ denotes as well an eigenvalue

$\hat{\mathbf{P}}_B$ Vector of bright zone target pressures used in the mixed control solution technique

$\mathbf{A}$ Hypothetical matrix used in examples.

$\mathbf{f}$ Vector of effective diaphragm forces, f_i, of a 4×8 source

$\mathbf{G}$ Matrix form of the impulse response from each source to each control point of the control zone used by the least squares method

$\mathbf{G}_B$ Matrix form of the impulse response from each source to each control point of the bright zone

$\mathbf{g}'_B$ Target pressure impulse response for an endfire array

$\mathbf{G}_D$ Matrix form of the impulse response from each source to each control point of the dark zone

$\mathbf{G}_{n_B m}$ Matrix form of the impulse response from the source m to the bright zone control microphone n_B

$\mathbf{G}_{n_D m}$ Matrix form of the impulse response from the source m to the bright zone control microphone n_D

$\mathbf{I}$ Identity matrix

$\mathbf{p}_B$ Bright zone pressures vector

$\mathbf{p}_D$ Dark zone pressures vector

$\mathbf{q}$ Vector of array volume velocities

$\mathbf{q}_{GF}$ Optimal array volume velocities vector based on point source Green functions

$\mathbf{q}_R$ Optimal array volume velocities vector based on replicated transfer functions

$\mathbf{S}$ Matrix of cross correlations inside a modal field

$\mathbf{T}_{ECST}$ Two-port network transfer impedance between the ear canal and the stapes

$\mathbf{u}$ Vector of diaphragm linear velocities, u_i, of a 4×8 source

$\mathbf{w}$ Vector of optimal filters optimised in the time domain

$\mathbf{y}_B$ Matrix of sampled outputs of the bright zone microphones y_{Bn_B}

$\mathbf{y}_D$ Matrix of sampled outputs of the bright zone microphones y_{Dn_D}

$\mathbf{y}_T$ Target pressure used by the least squares inverse method in the time domain

$\mathbf{Y}_{BM}$ Diagonal mobility matrix of the cochlear partition

$\mathbf{Z}_B$ Matrix of transfer functions of the bright zone

$\mathbf{Z}_D$ Matrix of transfer functions of the dark zone. It represents as well the transfer impedance of a PS source driver

$\mathbf{Z}_{B,M}$ Matrix of transfer functions of the bright zone based on measured responses

$\mathbf{Z}_{B_N}$ Matrix of transfer functions of the bright zone contaminated with noise

$\mathbf{Z}_{B_R}$ Matrix of transfer functions of the bright zone in where all its columns are identical (replicated matrix)

$\mathbf{Z}_{D,M}$ Matrix of transfer functions of the dark zone based on measured responses

$\mathbf{Z}_{D_N}$ Matrix of transfer functions of the dark zone contaminated with noise

$\mathbf{Z}_{FC}$ Fluid coupling matrix of the cochlear partition

$\mathbf{Z}_{M,Tot}$ Matrix of total mechanical impedances of 4×8 source

ω Angular frequency

$\overline{I}_k$ Mean acoustic intensity on the octave band k used in the calculation of the STI

$\overline{I}_{NOISE}$ Intensity of an interfering noise introduced by a certain transmission channel

$\overline{I}_{TEST}$ Intensity of a test signal for the calculation of the STI

ρ_0 Ambient density of air (1.21 kg.m^{-3})

τ Time delay

$\widehat{\alpha}$ Average absorption coefficients of a room walls

Ξ Amplitude and phase equalization to apply to a replicated matrix

ξ Parameter used to select between pressure matching or contrast maximisation in the combined solution technique

^{H} Hermitian transpose

A_{ST} Area of the stapes footplate

Bl Force factor of a moving-coil source

C Acoustic contrast performance ratio and cost function

c Constant pressure value used when finding the solution of the acoustic contrast maximization

c_0 Speed of sound (343 ms^{-1})

C_B Acoustical compliance of cabinet volume

D Peak value of a moving-coil source mechanical impedance

d Inter-element distance. Distance in m between two array sources. Can be as well the equivalent path length between front and rear responses of a phase-shift source

f_D Force driving the diaphragm of a moving-coil source

f_m Modulation frequency, used in the calculation of the STI

f_N Resonant frequency of a moving-coil driver

$f_{1,-3dB}$, $f_{2,-3dB}$ Half power frequencies

f_{AL} Spatial aliasing frequency

f_{Notch} Frequency at which the phase-shift sources lacks directionality

FPG Forward Pressure Gain

g Shear gain

G_{BA} The receptance of the BM with respect to the acoustic driving pressure, p_A

G_{BO} The receptance of the BM with respect to the active pressure, p_{OHC}

G_{SA} The receptance of the relative displacement between TM and RL with respect to the acoustic driving pressure, p_A

G_{SO} The receptance of the relative displacement between TM and RL with respect to the active pressure, p_{OHC}

h_k Squared impulse response for the octave band k, used in the calculation of the STI with the indirect method

I Electrical current flowing through a moving-coil source

$I_{am,k}$ Masking intensity induced in the octave band k by the octave band $k-1$ used in the calculation of the STI

$I_{rs,k}$ Intensity of the absolute threshold of hearing, used in the calculation of the STI

K_D Stiffness of a moving-coil source diaphragm

l Phase-shift source diaphragm to port distance

L_E Inductance of a loudspeaker coil

L_M SPL of a masker tone

M Number of control sources of an array

M_C Acoustic mass of the air load in the rear side of a diaphragm

M_D Mass of a moving-coil source diaphragm

m_k Modulation index on the octave band k used in the calculation of the STI

m_{BM} Mass of the BM

m_{RL} Mass of the RL

m_{TM} Mass of the TM

N_B Number of control points of the bright zone

N_D Number of control points of the dark zone

N_H Number of horizontal divisions when sampling in a circular geometry

N_T Total number of control points

N_V Number of azimuthal divisions when sampling in a circular geometry

$OFPG$ Overall Forward Pressure Gain

p Acoustic pressure

p_B Acoustic pressure for a single bright zone control point for the acoustic contrast maximization

p_B Reverberant pressure

p_D Acoustic pressure duet to the phase-shift source diaphragm

p_P Acoustic pressure due to the phase-shift source port

p_{AX} On-axis radiated pressure

p_A Acoustic pressure generated by the travelling wave

$p_{D,Array}$ Average pressure created by the array in the dark zone

$p_{D,TV}$ Average pressure created by the TV in the dark zone

P_{EC} Pressure in the ear cana for the two port network outer-middle ear model

p_{OHC} Pressure generated by the OHC

p_{SR} Pressure on the TM due to the RL

P_{ST} Pressure at the stapes for the two port network outer-middle ear model

p_{ST} Pressure on the RL due to the TM

Psi Factor which controls the directivity of a first-order source

Q Directivity factor

q_D Volume velocity of the phase-shift source diaphragm

q_P Volume velocity of the phase-shift source port

Q_{EC}, Q_{ST} Volume velocities in the ear canal and at the ear drum of the two port network outer-middle ear model

Q_{ERB} Quality factor of an equivalent rectangular bandwidth filter

R Acoustical resistance of the earmuff used to simulate an earphone. It may also refer to the correlation coefficient.

r Radius

R_D Damping of a moving-coil source diaphragm

r_D Distance between the phase-shift source diaphragm and a receiver position

R_E Moving-coil source electrical nominal resistance

R_M Moving-coil source mechanical impedance at, Z_M at the resonance pulsation

R_P Acoustical resistance of cabinet back port

r_P Distance between the phase-shift source port and a receiver position

S Surface of a room

s $j\omega$

S_D Surface of a source diaphragm. It equals the mechano-acoustical transduction factor

SNR_k Signal to noise ratio in the octave band k, used in the calculation of the STI with the indirect method

T Equivalent thickness of the BM

T Reverberation time

t Time

T_Q Absolute hearing threshold

$T_{BM/RL}$ Displacement of the BM as a function of that of the RL

$T_{RL/(BM+TM)}$ Displacement of the RL as a function of that of the BM and TM

$T_{TM/RL}$ Displacement of the TM as a function of that of the RL

TI Transmission Index, used in the calculation of the STI

u_{ST} Input excitation at the stapes

V Volume

v_{HL} Peak BM velocity of a hearing impaired cochlea

W Acoustic power

w_{BM} Bending displacement of the BM

w_{RL} Bending displacement of the RL

w_{TM} Bending displacement of the TM

Z_0 Free-field component of a reverberant transfer function, equal to Z_F

Z_E Electrical impedance of a moving-coil source

Z_F Free-field component of a reverberant transfer function, equal to Z_0

Z_P Transfer impedance of a phase-shift port

Z_r Reverberant component of a reverberant transfer function

$Z_{M,Tot}$ Total mechanical impedance of moving-coil loudspeaker

Z_{MC} Mechanical impedance of loudspeaker cavity

Z_{MD} Mechanical impedance of a moving-coil loudspeaker due to the diaphragm

Z_{ST} Input impedance at the stapes

Chapter 1

Introduction

1.1 Hearing impaired people and the need for hearing enhancement

Hearing impairment causes various difficulties in speech communication. People with moderate hearing loss can often clearly understand speech if only one person is talking, but they have difficulties if more than one person is talking at a time, or when background noise and/or reverberation are present [1]. Previous research has indicated that for mild hearing loss, audibility, i.e. the lack of volume, is the most important factor for understanding speech [2]. For severe losses, the poor discrimination of supratreshold[1] stimuli, due to the loss of frequency selectivity, is also of great importance [2]. This is because, apart from the reduction in audibility, which can be avoided with amplification, the widening of the auditory filters due to hearing loss leads to a smearing of the internal representation of the spectral shape of the signal [3]. Hearing impairment also involves a reduction in temporal resolution. This has been observed in older adults, which have shown to have problems in detecting a gap in a speech cue when the gap is short [4]. It has also been observed in animal studies that hearing loss impairs neural phase-locking[2].

One particular circumstance when hearing impaired people suffer from such problems with speech perception is when watching a TV program. In this situation the impaired viewers tend to require a higher reproduction volume, which results in a problem when listeners of different ages are watching TV at the same time. In an attempt to overcome such issues, wearable hearing aids or headphones have

[1] Audible.

[2] In this case phase-locking refers to the synchronism between the firing activity in the auditory nerve and the periodicity of a waveform in speech [5, 6]

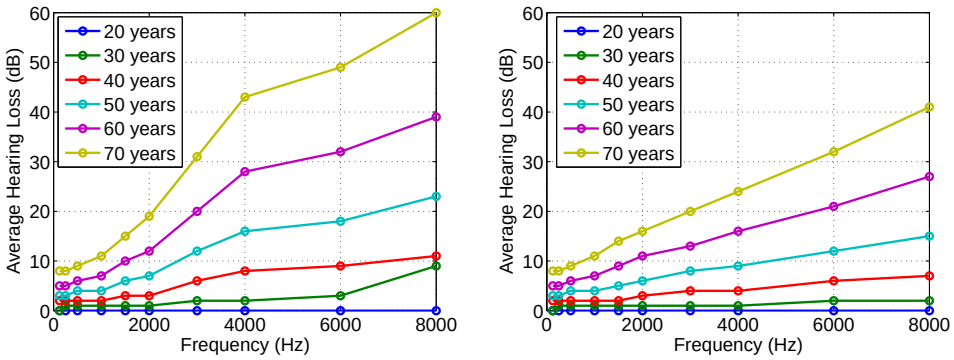
been used, but these have their own problems. The use of a personal audio device can potentially create a localised area of enhanced TV audio, which may give similar results to using headphones, but without the need to wear them.

1.1.1 Presbyacusis

Although hearing loss can be caused in many ways, special attention in this thesis is given to this which accompanies age, also known as *presbyacusis*. The term "presbyacusis" is used to describe the decline of hearing associated with various type of auditory system dysfunction (peripheral and/or central) that go with ageing and cannot be accounted for by extraordinary ototraumatic, genetic, or pathological conditions [7]. Presbycusis implies deficits not only in absolute thresholds but also in auditory perception as well [8]. It is estimated that about 30-35 percent of adults age 65 and older have a hearing loss; this figure is estimated to be 40-50 percent for people aged 75 and older [9].

The hearing loss associated with presbyacusis is usually greater for high-pitched sounds. For example, it may be difficult for someone to hear the nearby chirping of a bird or the ringing of a telephone; nonetheless, the same person may be able to hear clearly the low pitched sound of a truck rumbling down the street.

There are many causes of presbyacusis. Most commonly it arises from changes in the inner ear of a person as he or she ages, but presbyacusis can also result from changes in the middle ear or from complex changes along the nerve pathways leading to the brain. Hearing loss is also thought to be increased by exposure to noise and high level sounds, as measurements in noise free populations have shown to have smaller high frequency losses [10]. Presbyacusis most often occurs in both ears, affecting them equally. Because the process of loss is gradual, people who have presbyacusis may not realize that their hearing is diminishing [11]. The standard ISO 7029 "Statistical distribution of hearing thresholds as a function of age" [12] gives the estimated hearing loss which occurs in a given percentage of the population. The hearing loss predicted to occur in the 50% of the population is plotted in Figure 1.1, against frequency for various ages. Starting at around the age of 30, the hearing sensitivity starts to decline with age, particularly in high frequency. The standard ISO 7029 gives a distribution of hearing loss which is greater in men than in women, perhaps reflecting the larger noise exposure of the men in this study, who grew up without noise regulations.



(a) Men **(b)** Women

Figure 1.1: Hearing losses as a function of frequency expected to be found in 50% of the population at different ages.

1.1.2 Understanding the TV program for the hearing impaired

The problem

As discussed in Section 1.1.1, elder people tend to have reduced hearing capabilities, which impairs their speech communication. This is a problem, when watching the TV, as it gives rise to difficulties in understanding the program information[3].

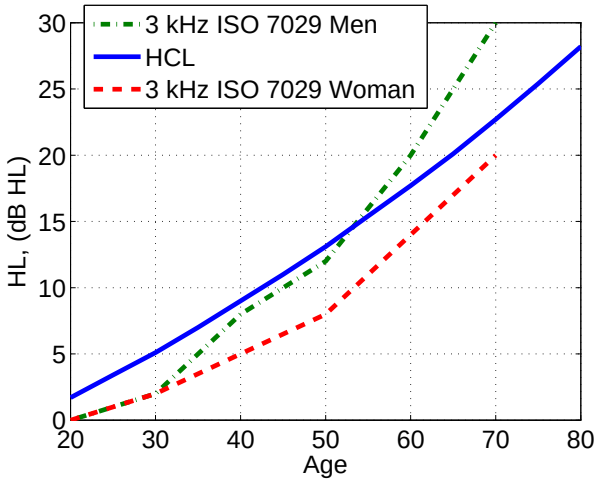


Figure 1.2: Comparison of the distribution of gain with respect to a 12 year old listener needed to satisfy the hearing comfort levels (HCL), as given in [13], and the prediction of hearing loss with age at 3 kHz as given by the standard ISO 7029 [12].

Table 1.1 hearing comfort levels (HCL) as a function of age, as given by Stanley

[3]It has recently been found that 71 % of the adults in the UK can not hear properly what it is said on the TV [9].

Table 1.1: Hearing comfort levels (HCL) of speech as a function of age, according to [13], using a random sampling of 799 subjects.

Age (years)	Hearing comfort level, (dB SPL)
15	53.5
20	55.2
25	56.9
30	58.6
35	60.5
40	62.5
45	64.5
50	66.6
55	68.9
60	71.2
65	73.6
70	76.2
75	78.9
80	81.7
85	84.6
90	87.6

[13]. This is also shown in Fig. 1.2, where it is seen how the gain needed at different ages to satisfy the HCL increases with a very similar trend to the prediction of HL with age, as given by the standard ISO 7029 [12]. Hence, elder TV viewers require higher listener levels compared with healthy young listeners. Such an increase in listening levels is a source of complaint for the members of the household with good hearing, which appears to be common [14], and is estimated to occur in 10 % of households [15]. This is, however, also a problem for the presbycutic listeners themselves. Because they suffer from recruitment, a tendency to select high level volume settings for normal programs makes elder listeners likely to be more bothered by sudden level increases, such as those common in commercial breaks [16].

Older listeners also complaint about the speed of the speech being too fast, lacking clarity, or being swamped by additional sounds. A study performed by the BBC [17] showed that hearing-impaired listeners rated speech more intelligible if background audio level was reduced by 6 dB. A more recent study showed that percussive background music and music with similar frequency content of that of speech are very negative for intelligibility [18].

1.1.3 Current systems

Amplification devices or procedures intended to amplify the audio of the TV to improve speech intelligibility are already available in the market. Such systems are mainly based in headphones, or on additional secondary speakers, which are placed close to the locations of the hearing impaired listeners, so that the audio program is presented at a louder level.

Personal Television Listening Devices

These devices are based in wireless earphones and headset that provides a personal audio application for the interested individuals. Audio brands as Sennheiser have a part of its products catalogue dedicated to these devices [19]. High end applications that can be connected to the TV and provide processing in order to give a good and clear image of the audio can be found at the manufacturer's web page. Fig. 1.3 shows an example of the use of these gadgets. Although providing a good audio quality, the main drawback of these devices is that its use after some time can be tiring to wear and can also be socially isolating.

Figure 1.3: *Sennheiser DirectEar Set IS 410.* Reproduced from [19].

High clarity sound systems

Some manufacturers make specially designed home theatre systems which provide a proper equalisation in order to make hearing simpler to hearing impaired people. One manufacturer is ZVOX. Information about these systems can be found in the manufacturer web page ZVOX. They are based on monaural loudspeakers that tend to "anchor" the sound, creating clear, robust audio that is easier to hear and are effective in reproducing speech. The physical appearance of this device is shown in Fig. 1.4. Such devices are typically designed to be placed closed to the TV, and since they are not spatially selective, they also affect the sound heard by other normal hearing listeners.

Figure 1.4: *ZVOX 325.* Reproduced from [20].

TV EARS and audiofox

TV EARS is a manufacturer specialized in products for hearing impaired TV viewers [21]. This manufacturer has a wireless portable loudspeaker that can be placed near the hearing impaired viewers in order to provide them with a clearer and reinforced sound, see Fig. 1.5. A similar product is made by Audio Fox [22]. It is based on two small loudspeakers placed close to the head of the TV spectators. These two products provide a higher level of the program SPL near the spectators, but are not designed to do an effective control of the sound field around the space.

Figure 1.5: *TV Ears 2.3 Wireless speaker system.* Reproduced from [21].

Figure 1.6: *Audio Fox wireless TV listening device.* Reproduced from [22].

Widex TV-DEX

The danish manufacturer of hearing aids Widex has proposed an approach which directly sends the TV program to the hearing aid (a personal hearing aid that introduces the audio directly to the ear canal). It is called *TV-DEX* [23] and it allows the disconnection of the hearing aid microphone and the ability to listen only to the audio of the TV. This is a good approach, but needs the TV viewer to wear a hearing aid.

Figure 1.7: *Widex TV-DEX transmitter.* Reproduced from [23].

Captions

Captions are words displayed on a television screen that describe the audio or sound portion of a program. Captions allow viewers who are deaf or hard of hearing to follow the dialogue and the action of a program simultaneously. They can also provide information about who is speaking or about sound effects that may be important to understand the whole context. Captions are created from the transcript of a program. A captioner separates the dialogue into captions and makes sure the words appear in sync with the audio they describe. Specially designed computer software program encodes the captioning information and combines it with the audio and video to create a new master tape or digital file of the program. This is the most basic aid to the hearing [24], and although it is not an acoustic solution to the personal audio, it is important to include it here.

Additional soundtrack

This approach was proposed by the BBC. Based on previous studies of hearing balances [17], the BBC has proposed to provide an additional soundtrack with an audio balance suitable for hearing impaired people which should be made possible via IP broadcast delivery [25].

Speech processing algorithms

These algorithms use signal processing to amplify the speech information above the background noise level. A review of some of them has been performed recently by Rumsey [26]. One example of these algorithms is that developed by Müsch [16, 27]. A block diagram of this algorithm can be seen in Fig. 1.8. This algorithm makes use of the central channel of a 5.1 audio system, which is used mostly for speech information. The distinction between speech and not speech information is performed by a speech detection algorithm [28]. Once the speech is detected, the algorithm judges its loudness compared to the background noise, and if needed, reduces the gain of the additional channels, L, R, ls and rs, which do not conduct speech. To determine where the non-audio channels are decreased, the algorithm uses the speech intelligibility index (SII) [29].

A similar algorithm is that developed by Uhle [30], where a pattern recognition system is used to detect when there is speech present in the audio signal.

1.1.4 Amplification

A consistent finding in studies of ageing and hearing loss, is that the loss of audibility is the biggest contributor to the reduction of speech perception [31]. Such loss of audibility is caused by the difficulty in hearing high-frequency speech sounds (due to the high-frequency hearing loss typical of presbyacusis), and by masking of speech by music and effects or the same speech. As seen along this thesis, both effects can be reduced by using a suitable amplification.

There are many clinically proven amplification techniques that determine the degree of gain best suited for people with hearing loss. These can be linear procedures, such as the *half-gain rule*, described by Lybarger [32], or the NAL [33]. There are other more modern and sophisticated, non-linear procedures, as the NAL-NL1 [34] and NAL-NL2 [35], and many others more that are not listed here. These last two perform an analysis in third octave bands of the spectrum and analyse which of them can be amplified more effectively to improve speech

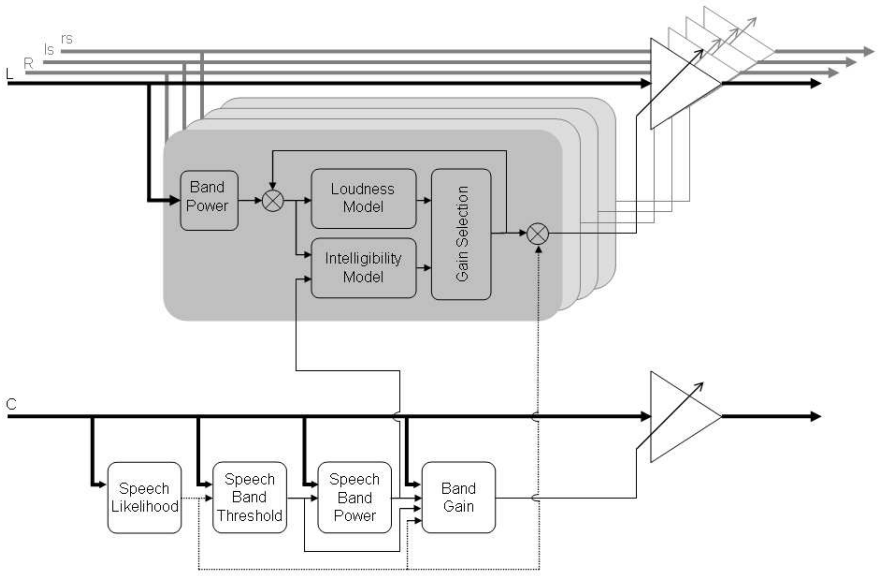


Figure 1.8: Block diagrams of Müsch's speech enhancement algorithm. The thick lines represent 256-point MDCT transform coefficients of audio signals. The four top signals represent non-audio channels, L, R, ls and rs, and the bottom line, C, represents the speech channel. The thin lines represents the n-dimensional control signals. The dotted line represent the control signal of the likelihood of speech presence. Reproduced with permission from [16].

communication, denoted as *effective audibility* [36].

Instead of wearing a hearing aid, suitable amplification may be obtained by creating a localised enhanced volume area around the listener, which is obtained using a personal audio device.

1.2 Personal Audio

The problem of generating an enhanced TV sound for a particular listener can be considered to be one of personal audio. Personal audio involves the generation of listening spaces which are restricted to a specific region. Such systems can be used to increase the privacy of a listener, allow multiple programs to be reproduced in a common space, reduce annoyance to nearby listeners and improve speech intelligibility. Personal audio works by creating constructive interference of the sound in the zones where the pressure is to be maximised, and destructive interference in the places where the sound pressure is to be reduced. Such systems are optimised by solving an inverse problem, in which by knowing the transfer functions of a set of loudspeakers to the various control zones, the driving voltages needed to maximise that interference pattern are found.

Other techniques for the control of sound fields are binaural reproduction over loudspeakers [37, 38] and wave field synthesis (WFS) [39]. Binaural rendering operates by producing a sound field at the listener's ear which is ideally equal to the sound field that will occur in a real environment. Reproduction using WFS aims to reproduce wave fronts using arrays with a large number of sources. This technique is based in Huygens' principle, which states that any point of a wave can be considered as a secondary source. The use of WFS produces a soundfield through which the listener can move whilst still perceiving the intended spatial sound field, whilst that for the case of the binaural reproduction the reproduction is confined to a small listening area, which is a significant advantage for WFS. This advantage is, however, obtained at the expense of using a large number of sources, whilst binaural sound reproduction can be obtained with just 2 sources. Nevertheless, none of these techniques are suitable for the application considered here. Although the rendering of a soundfield equal to that found in a natural listening scenario may be good for speech intelligibility, it does not contribute to the creation of a very narrow radiation pattern that can enhance the sound for particular listeners, and hence, is out of the scope of this thesis.

1.2.1 Early work on personal audio

An early attempt to create a personalised sound reproduction was described by Druyvesteyn and Garas [40]. The authors introduced a system that used three different techniques to create narrow reproduction zones; active control of sound at low frequencies, arrays of loudspeakers at medium frequencies and the directivity of the individual sources at high frequencies. The system aims to reproduce two audio programs in the same environment with minimal disturbance to each other within the same room. The loudspeaker array used a delay and sum configuration, and was able to work in the 1 kHz to 4 kHz band. According to subjective experiments, an effective personal audio strategy must achieve a ratio of levels in the various zones at least greater than 11 dB but preferably of 20 dB [41]. The combined system was able of achieving about 20 dB SPL of difference between the two programs between 250 Hz and 4 kHz. The authors stated that with the use of the array an improvement of 0.15 in the STI [42] was obtained, which corresponds to an improvement of the speech intelligibility from "fair" to "good" or from "good" to "excellent".

The performance of the array was limited at higher frequencies due to the finite separation between loudspeakers and at lower frequencies due to the aperture of the array. More recent advances in personal audio reproduction tend towards

the use of superdirective, or optimal, beamforming techniques. Such techniques increase the directivity of an array for a given number and spacing of array elements [43], and therefore do not rely on the combination of approaches as considered in [40].

Superdirective beamforming

Superdirective, or optimal beamforming has been studied extensively in the field of sensor arrays [44] and in the field of high gain radio frequency antennas [45]. In the context of personal audio reproduction, a number of different strategies have been presented which lead to superdirective systems, such as brightness maximisation [46], acoustic contrast maximisation [46], energy difference maximisation [47], sound power minimisation [48], velocity control reproduction [49], least squares inverse filters [50] and mixed control techniques [51]. From all these techniques, the most widely employed and investigated method have been the acoustic contrast maximisation [46] and the least squares inverse filters [50].

The acoustic contrast maximisation, proposed by Choi and Kim [46] optimises the source strengths of an array of sources to maximise the ratio of mean square acoustic pressure between a bright zone, where the acoustic potential energy is to be maximum, and a dark zone, where the acoustic potential energy is to be minimum. Such optimisation assures that the array minimises the sound radiated into the dark zone whilst maintaining a certain pressure level in the bright zone. An example of this problem is illustrated in Fig. 1.9.

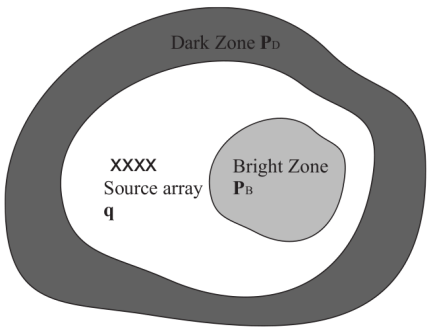


Figure 1.9: Example of two-dimensional control geometry for an acoustic contrast problem. The dark zone represents the area where the sound intensity is to be minimised with the constraint that a certain intensity is present in the bright zone. Reproduced from [52].

The least squares filters find a set of source strengths which minimise the error between a desired sound field and an existing sound field. A block diagram of a least squares system is shown in Fig. 1.10, where it can be observed that the

error is the difference between a set of reproduced signals and a set of desired signals. The particularity of the least squares with respect to the acoustic contrast maximisation is that the target sound field allows the control of the amplitude and phase of the reproduced signal, which gives a better audio quality [53].

The mathematical formulation of both the acoustic contrast maximisation and least squares inverse filters will be considered in Chapter 3 and Appendix 4, but both involve the inversion of matrices that may not be well conditioned and can thus require regularisation.

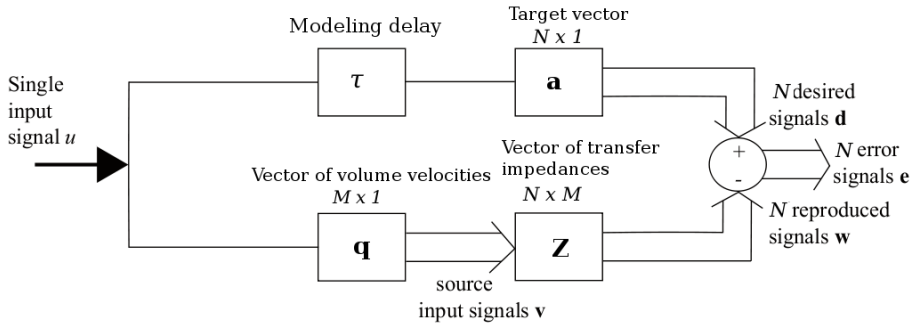


Figure 1.10: Block diagram of the reproduction system used in the least squares approach. Reproduced from [50].

1.2.2 The use of superdirective beamforming on personal audio applications

The control of a soundfield using acoustic contrast maximisation has been previously applied to a practical problem by Elliott and Jones [48]. The authors considered an application in where two listeners are sat adjacent to one another in an aircraft or vehicle type scenario listening to different audio programs. The control system presented by the authors is shown to achieve a good performance at low frequencies, however, at frequencies where the separation between the primary and secondary sources become comparable to the acoustic wavelength, the dark zone becomes localised around the error microphones. The same authors also describe a system in which the dark zone was extended to account for more listeners, as shown in Fig. 1.11, and hence represents a more practical reproduction situation, as for example in an aeroplane or in a car [54].

The use of mobile devices such as tablets, portable game consoles and mobile phones is quite extended nowadays. Audio reproduction using the standard loud-

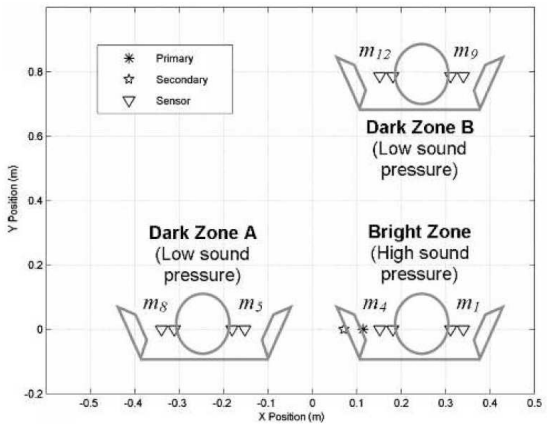


Figure 1.11: Practical personal audio problem with multiple dark zones presented by Elliott and Jones [54]. Reproduced from [54].

speakers of the devices may reduce the privacy of the user and get to annoy other users. A directional radiator may be quite convenient in this case to personalise the audio reproduction. This was first studied by Cheer [55], where the author compared a 2 source endfire array controlled by acoustic contrast maximisation with a phase-shift source. Elliott *et al.* [52] considered the use of a superdirective endfire array of 2, 3 and 4 sources to control the directivity of a mobile phone speaker. The authors observed that such arrangements could largely reduce the radiated sound power, at the expense of using a large amount of electrical power (50 dB for a three source array at low frequency). The authors also studied the possibility of a 2 source endfire with the sources coupled via the cabinet, resulting in a large increase in the efficiency of the array. The use of superdirective arrays for mobile phones was later studied by Cheer *et al.* [56] by means of FE simulations, in where the authors studied the effect that the baffle of the mobile phone has on the soundfield control. Further work by the same authors was later presented in [57]. In this publication the authors showed practical measurements of performance of the baffled device and used a 3D control geometry to concentrate the radiation towards the front of the array, as depicted in Fig. 1.12.

An application of interest is that of restricting the sound radiated by the loudspeakers of a laptop, which may be useful, for example, in an open plan office to improve privacy and reduce the distraction of other users. Such application was presented by Chang *et al.* [59], wherein the authors studied a miniature 9 source line array for personal multimedia reproduction. In order to focus the reproduction, the authors used a control geometry in the near field of the array, in order to obtain control around the head of the laptop user. The control geometry and the proposed personal audio system are depicted in Fig. 1.13. The system obtained a ratio of mean square pressures between bright and dark zones

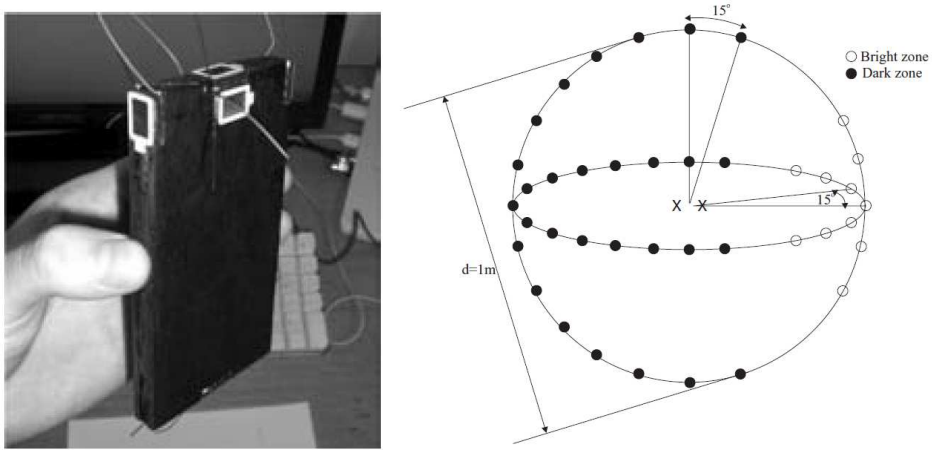


Figure 1.12: The superdirective mobile phone sound system as studied by Cheer *et al.* [57] and the control geometry used to concentrate the radiation towards the user. Reproduced from [58].

of about 20 dB at 1 kHz and about 35 dB at 4 kHz in anechoic conditions. The scattering introduced by the head of the laptop user was also considered in free-field conditions [60], observing that the ratio of pressures between bright and dark zones was reduced by the users head. The authors also studied the possibility of performing 2 channel reproduction with the system [61]. A very similar application using piezoelectric loudspeakers has been presented by Kim *et al* [62].

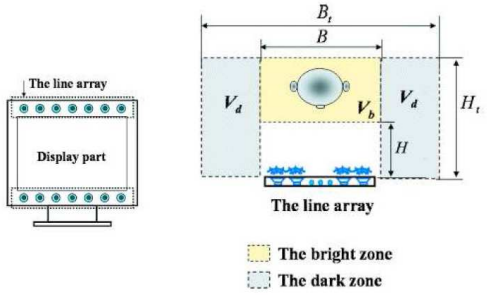


Figure 1.13: Control geometry used by Chang *et al.* [59] to localise sound radiation around the laptop user. Reproduced from [59].

Other potential application for the use of personal audio is the reproduction of individual listening zones inside a car. This is beneficial for a case in which the people sat in the front seats are listening to the radio and the people sat in the rear seats are watching a movie. Such application has been considered by Cheer *et al.* [58] where the authors presented a combined system using the loudspeakers of the car sound system to control low frequencies plus 8 extra phase-shift sources to control the mid-high frequency range. The system showed to perform well when the bright zone was selected to be the front seats of the car, but was limited

when the bright zone was constituted by the rear seats due to the unidirectionality of the phase-shift sources. Further work in this field has tended towards the utilisation of first order arrays to control the mid-high frequency range in order to enhance the performance of the rear bright zone, and to provide a smoother reproduction in the listening zone [63].

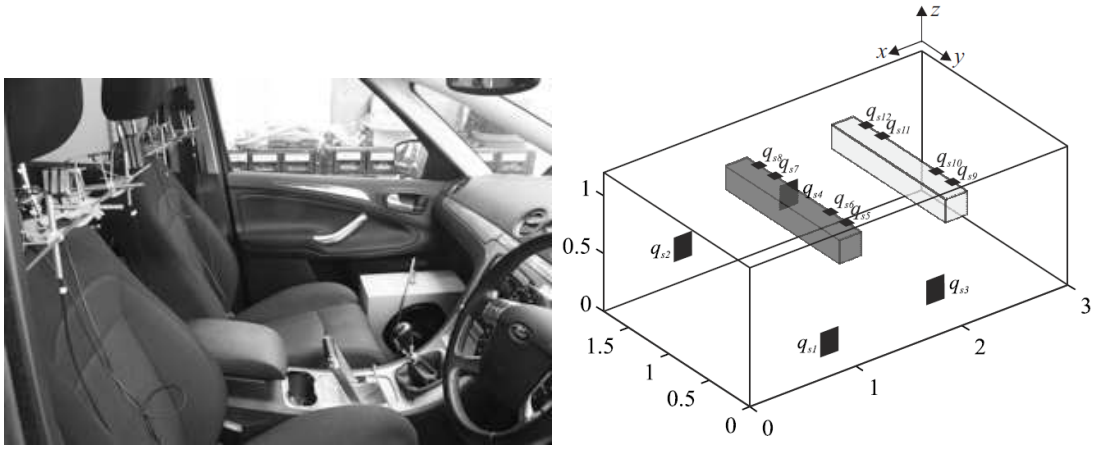


Figure 1.14: Personal audio loudspeaker and microphone set up as studied by Cheer *et al.* [57] and the control geometry used to create the front and rear bright zones. Reproduced from [57].

Optimal beamforming techniques have also been widely used for improving the performance of line arrays at low and mid frequencies. Boone *et al.* [64] studied the use of superdirective beamforming in an endfire line array of 8 sources. The authors studied the trade off between electrical power used by the array and the amount of superdirectivity by using monopole simulations. The array filters were optimised so that the DI was maximised. The authors then modelled the array using finite elements (FE) and finally built a prototype using commercial speakers. In order to control the power radiated by the array the authors introduced a regularisation parameter.

The previously listed applications use the acoustic contrast maximisation technique to create superdirective filters, however, the use of least squares has also been widely studied to generate highly directional arrays. Luzarraga studied the use of least squares for controlling the soundfield radiated by a commercial public address (PA) system [65], using a cylindrical propagation operator for the target field. The author introduced a high pass filter in the target function in order to prevent the array filter boosting excessively the sources at low frequency. Frequency dependent regularisation proportional to the largest singular value at each

frequency was introduced.

Least squares filters have also been used in dual layer miniature arrays for the creation of private listening zones. Shin *et al.* [66] first studied the directional performance that could be obtained using a *classic*[4] control geometry to create a private sound field, established to concentrate the energy in broadside and reduce the energy radiated to the back of the array, as shown in Fig. 1.15. The array used by the authors can be observed in Fig. 1.15, and consists of two rows of 8 sources, wherein the second row of sources is used to reduce the back radiation of the array.

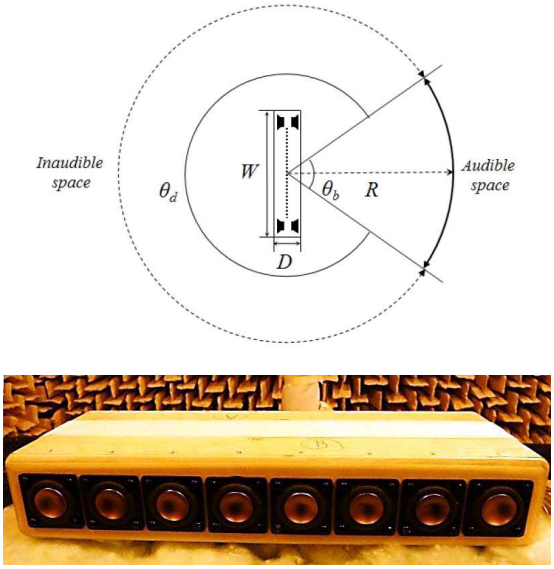


Figure 1.15: Control geometry used for creating a"private sound field" and 16 source line array used by Shin *et al.* [66]. Reproduced from [66].

The filters created via the acoustic contrast maximisation in the previous approaches maximise the ratio of mean square pressure between both control zones, but do not guaranty a smooth reproduction in the listening area. In the other hand, the least squares filters are able of obtaining a better reproduction in the listening area at the expense of having a reduced directional performance. It may be desirable in some circumstances to obtain a control of the sound field that obtains a trade off between both approaches. This has motivated research in sound control techniques which are able of maximising the directivity whilst providing a more flat reproduction in the listening zone. Chang and Jacobsen presented a formulation in which they introduce an extra control parameter in the least squares formulation [67, 51]. By modifying such parameter between 0 and

[4]In this case a classic control geometry refers to that which consists of a bright zone and a dark zone.

1, the obtained filters allow to maximise the acoustic contrast between the listening zones, or to obtain a reproduced pressure as close as possible to the target.

A similar algorithm has been recently presented by Cai *et al.* [68], which via a convex optimisation finds such set of array filters which maximise the contrast ratio in a given control geometry, but still allows to control the phase in the listening zone. Shin *et al.* [69] introduce a similar formulation as that of Chang and Jacobsen [67] wherein the authors introduce an extra control zone, termed *grey* zone. This minimises the mean square error in both bright and dark zones with the error being given less importance in the grey zone, which produces a better directional response with a more precise reproduction of the target field. The amount of minimisation of the error in the grey zone is controlled by an extra parameter. The control geometry used by Shin *et al.* in [69] is shown in Fig. 1.16. Another approach to produce a more accurate reproduction of the soundfield, but at the same time maximising the ratio of mean square pressure between two listening zones, has been presented by Coleman *et al.* [70], in which they studied the performance of such method by using a circular loudspeaker array. The algorithm proposed by the author allows the reproduction of a plane wave in the listening zone in order to enhance the spatial quality of the target audio.

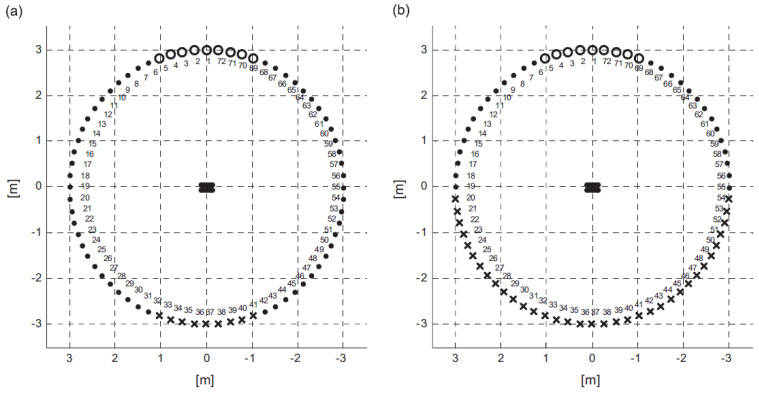


Figure 1.16: Control geometries as used by Shin *et al.* [69], wherein the open circles represent the bright zone, the closed circles the grey zone and the crosses the dark zone. Reproduced from [69].

1.3 Issues associated with personal audio

1.3.1 Superdirective filters in the time domain

The examples of superdirective beamforming presented above perform the optimisation of the array filters in the frequency domain. The optimisation problem is thus performed frequency by frequency, obtaining an optimal filter in the form of a transfer function in the frequency domain. The impulse response of the optimal filters are obtained afterwards by performing an inverse Fourier transform on the frequency response corresponding to each source [57]. The use of least squares filters optimised in the time domain has been previously confined to the equalisation of room impulse response and for binaural reproduction [71, 72], but it has not been applied for personal audio applications. The subjective audio quality obtained by the least squares filters optimised in the time domain in comparison with these being obtained in the frequency domain was studied by Norcross *et al* [73]. The authors show that for filters length smaller than 2048 coefficients, the filters optimised in the time domain provided a better audio quality than those optimised in the frequency domain.

The use of acoustic contrast maximisation with filters optimised in the time domain has been recently presented by Cai *et al.* [74]. The authors used the active control formulation adapted for the acoustic contrast maximisation by Elliott *et al.* [75] in an 8 source line array [74, 76]. Cai *et al.* [74] found that the acoustic contrast maximisation formulated in the time domain was unable to provide a flat frequency response, and they introduced an alternative formulation in order to smooth the response.

1.3.2 Personal audio reproduction in reverberant spaces

The performance of most personal audio applications has generally been studied only in anechoic environments. This is far from the real environment in which the TV reproduction is intended to be performed, which in a typical room is going to be a mixture between a modal and a diffuse field [77]. The amount of research on the performance of directional radiators in non anechoic fields is small. Reproduction in such environments is difficult, due to the rapid variation of the acoustic transfer functions at different positions in a room [78]. Research has tended towards equalisation of the transfer functions of sources over multiple points using least squares inverse filters [72, 71]. In contrast, sound reproduction would require the equalisation to be extended over a wider control region. To this end, some researchers have measured and averaged the transfer functions

over a certain region [79], however, these techniques have shown "not to have enough accuracy" for sound reproduction inside a room. Other studies have created array filters for sound reproduction in reverberant environments by using the basis function of the room [80]. The efficiency of such methods has only been assessed, though, by numerical simulations and not by experimental measurements.

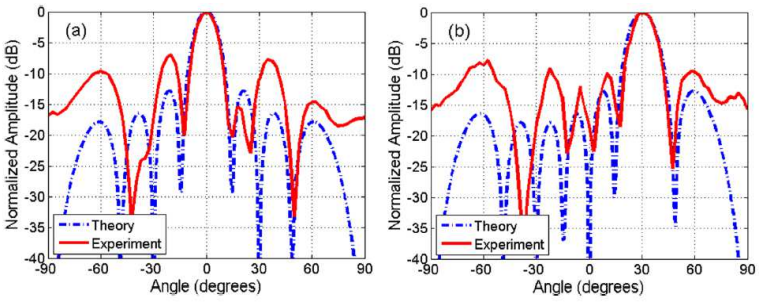


Figure 1.17: Free-field theoretical directivity, (a), and that measured in a reverberant field, (b), as presented by Anderson *et al.* [81]. Results reproduced from [81].

The performance of a single directional acoustical radiator in a reverberant field was studied by Beranek [82], who derived an expression for the total acoustic field depending on the acoustic characteristics of the room and the radiator. A similar formulation was used to obtain the directional characteristics of a source in a reverberant environment by Druyvesteyn and Garas [40]. Measurements of the performance of a line array in reverberant environments have also been presented by Anderson *et al.* [81], showing a reduction of the directional characteristics in comparison with the theoretical free-field value, as shown in Fig. 1.17. Recent work has analysed the performance of different transfer response models for creating the filters of a superdirective array in a reverberant environment [83]. Similar work was presented by Wen *et al.* [84], where the authors show that by measuring the transfer functions of the room the performance was largely increased with respect to a delay and sum array. Yon *et al.* [85, 86] showed that by placing the array in a reverberant environment, the reflections with the room surfaces can extend the aperture of the array and virtually increase the number of sources of the radiator. Thus, reflections on boundaries allow the creation of a virtual array of emitting transducers in every direction, as shown in Fig. 1.18. This research employed the first order reflections of the room to obtain a more accurate reproduction in the listening zone, using existing commercial sound-bars [87] or by using high order sources [88].

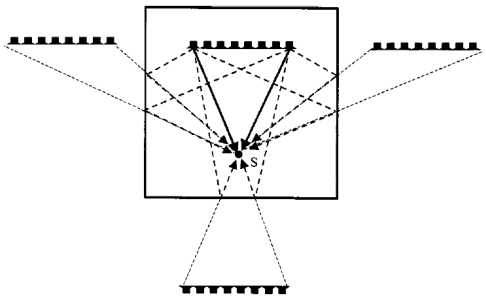


Figure 1.18: Principle of mirror images as presented by Yon *et al.* [86]. Results reproduced from [86].

1.3.3 Energy consumption and robustness of superdirective personal audio systems

Although, in principle, superdirective beamforming techniques are able to significantly increase the directional performance of a line array at low and medium frequencies, this needs a very large amount of electrical power to drive the loudspeakers [53]. For example, the use of superdirective techniques allow the directivity of an 8 source line array to be made frequency independent, but this can need to use 200 dB more of electric power at low frequency than that used at 3 kHz [53]. Apart from the fact that such systems need a huge amount of power, superdirective systems are also very sensitive to errors in the transfer functions of the sources. How sensitive an array is with respect to source transfer function mismatches can be treated in terms of robustness; the more robust an array is the less sensitive it will be to such mismatches.

In superdirective antennas, for example, it has been observed how the practical response is far from the theoretical superdirective value if random errors are introduced in the antenna weightings [89]. The characteristics of these errors and how to obtain robust beamforming in their response has been studied for microphone arrays, assuming that the errors have a certain probability factor [90]. Work on loudspeaker arrays has also studied the effect of errors in source sensitivity and position, and how these affect the ratio of mean square pressures between two spatial zones [91, 92]. Such errors reduce the directional performance in the frequency range where the sources of the array are highly correlated, spatially speaking [93]. Superdirective techniques need the inversion of a matrix to obtain the array filters, and these matrices are ill-conditioned at such frequencies. The robustness of a line array depends also on the number of sources used for the acoustical array. The more sources used, the greater the directional characteristics, but also the less robust the array is, so that a balance between robustness

and performance is needed [94, 95]. Generally, line arrays and high order acoustical radiators are made robust to changes in their transfer responses by applying regularization [64, 91], which also controls the power used to drive the array sources [53]. Recent work by Elliott *et al.* [91] has suggested that the space-averaged uncertainties introduced into the transfer responses of an array by the diffuse field can act analogously to regularization. This has been further studied by performance measurements in a reverberant environment [96], in which it has been shown that mismatches in the array transfer functions of 1 dB or 3 dB do not affect the final array response in a reverberant field.

1.3.4 Reducing radiation from the rear of the array

Apart from the steering of the main beam, another issue with such arrays is the attenuation of the back radiation of the array, which would otherwise excite the reverberant field. The directivity of a line array with omnidirectional drivers is symmetrical along the main axis of an array. For the case of a broadside array, for example, the array will radiate as much from the back as from the front, which is undesirable for many applications. In order to reduce the back radiation, directional sources using multiple drivers can be used. The use of such an arrangement has been previously studied by Pefretzschner and Romera [97], who placed an extra source at the back of a line array to reduce the back radiation. A similar concept has been exploited for creating directional sources by Olson [98], where the directional control was achieved with a delay between the front and rear sources. Commercial implementations of these kind of systems can also be found in public address systems, with the aim of reducing feedback [99, 100]. The use of acoustic contrast maximization to control the directivity of a pair of sources has also been studied [52], showing greater performance than when the control is performed with a simple delay. However, although these solutions achieve a good performance, they require multiple sources to work, and delay networks or digital filters to perform the desired control between the front and rear sources.

A simpler and more efficient alternative is the use of phase-shift loudspeakers. Phase-shift loudspeakers use the same working principle as phase-shift directional microphones, for which the sensitivity to pressure gradients is controlled by using resistance loaded ports [101, 55]. The use of a phase-shift enclosure to create a directional response was first introduced by Iding [102], who presented measurements of directional performance. The phase-shift source has also been studied by Olson [98] and by Holmes [103]. By selecting the acoustical compliance of the

loudspeaker's cabinet and the resistance of a rear port, an optimal directional pattern can be achieved. Ideally, such pattern should be tuned to be a hypercardioid source, as this is the first order limaçon which minimises the input to the reverberant field [104]. The bandwidth of utilisation of a phase-shift source is, however, limited by its dimensions, and is maybe because of that, although it is a concept which dates from the 70s, its utilisation has been limited.

1.4 Providing an alternative personalised audio for the hearing impaired

It is clear that the speech intelligibility of a hearing impaired listener will benefit from an increase in volume. It has also been seen that numerous alternatives exist in the market which can amplify the sound to partially recover from the loss in audibility. All these approaches, however, are based on headphones or in secondary speakers placed close to the listeners. The headphones are effective but after a long time can become tiring. The placement of secondary sources close to the listeners may be also effective in increasing audibility but will also increase the overall sound pressure level along the room, which can result too loud for normal hearing listeners [13].

An ideal sound system to enhance the audio of the hearing impaired will be based on the following criteria:

- The system should enhance the audio, which improves the intelligibility of speech, for the hearing impaired listener without the need to wear any kind of hearing aid or extra device.
- The action of the system is restricted to an area of reduced size where the hearing impaired person or people are placed, and must not affect the quality of the audio being perceived by normal hearing listeners outside that area.
- The system should not significantly increase the reverberant sound in the room where it is being used.

Such system requires a very narrow spatial reproduction, which can only be performed by the use of active control techniques or by the use of loudspeaker arrays. The use of active control is out of the scope of the application, as it works better for frequencies below a few hundred Hz [40, 105], and the aim is to increase the

audio of the presbyacutic hearing loss, which are large above 500 Hz.

An example of how an acoustic source array could be used to increase the intelligibility in the hearing impaired is shown in Fig. 1.19. Although an array can be placed closer to a listener, in this case the array is placed below the TV, as a typical sound bar, and creates a beam of sound to a position where a hearing impaired person is placed. The radiation of the array to other directions is minimised, so that the rest of the TV viewers listen to the audio coming from the TV sound system.

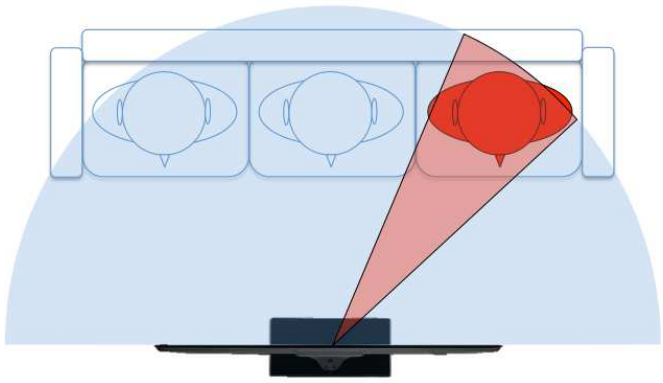


Figure 1.19: Suggested listening situation for the enhancement of speech intelligibility: The normal hearing listeners (in blue) listen to the audio coming from the TV, which propagates omnidirectionally whilst that the hearing impaired (in red) listener benefices from the high frequency boost of the array. Reproduced from [106].

1.5 Structure of the thesis

This thesis mainly focuses in the development of a personal audio system which can be used to enhance the speech intelligibility of hearing impaired TV viewers. From the literature review above it can be seen that although a good deal of research has been done in personal audio reproduction, none of the applications has been concerned with the aims of this thesis. Furthermore, most of the applications are based on a free-field reproduction, which is far from the final scenario of utilisation of such hearing aid device. Therefore in the context of sound reproduction, special attention has been put in designing the array so that the reverberant field is not excited. Extra attention has also been put in the study of how to maximise the performance of an array inside a reverberant field, and how the beamforming is affected by such field. Apart from the directive performance obtained with the array, the perceived quality of the reproduced programme is

also quite important. To this end, objective and subjective quality tests have been carried out with the array.

There is a direct link between the degradation of the human cochlea with age and the reduction of speech intelligibility. A model of the cochlea has thus been investigated to study the ways in which the response is affected with hearing loss. The last part of this thesis studies how this hearing loss affects the audibility of sounds and how better forms of amplification could be provided with the array.

Chapter 2

The study of phase-shift radiators is contained in this chapter. First, simulations using free-field monopoles together with a lumped parameter model of the rear port are used to understand the behaviour of the phase-shift sources along frequency. As these sources experience some limitations, these are further studied by means of finite element simulations. A new proposed model of rear port resistance is designed and then built, which leads to a large broadening of the effective utilisation bandwidth of the sources.

Chapter 3

This chapter introduces the theory of acoustical arrays. First the superdirective methods in the frequency domain are revised, such as acoustic contrast maximisation and least squares filters. The performance that is obtained using such design methods is studied and compared. The two arrays developed along this thesis, the 1×8 array and 4×8, array are presented. A study of the steering capabilities of a phase-shift source array is also introduced.

Chapter 4

The creation of superdirective filters in the time domain is discussed in this chapter. Three techniques are examined: the acoustic contrast maximisation, the least squares inverse filters and the mixed control algorithm [51]. The performance of the filters created in the time domain is then compared with that obtained by the filters created in the frequency domain. The time domain formulation is also used to maximise speech intelligibility inside reverberant environments.

Chapter 5

The effect of reverberation in superdirective acoustic source radiators is studied in this chapter. First, a formulation is introduced which allows to model the

directional performance of a radiator in a diffuse field, by knowing the free-field performance of such radiator and the absorption and dimensions of the room where its performance is studied. The effect of the mismatches introduced by the diffuse field is then studied by means of a theoretical analysis, and by measurements of performance in a room, showing how the diffuse field acts analogously to regularisation towards the calculation of the vector of optimal source strengths.

Chapter 6

Apart from the acoustical measurements performed on the array it is important to assess its subjective effect on improving speech intelligibility in hard of hearing TV viewers. In this chapter such testing is initially carried out with the use of an objective speech intelligibility technique, such as the speech tranmission index (STI). In order to further back up these results, subjective tests are then described, carried out with 30 individuals.

Chapter 7

This chapter introduces a physiologically realisable model of the organ of Corti, which is able of faithfully reproduce the response of the human cochlea in the frequency domain. The model is used to reproduce psychoacoustics tuning curves and the upward thresholds of masking. Hearing loss is also introduced into the model, and its effect on a number of psychoacoustic factors is considered. This model allows the amplification to be found, that minimises the masking in the hearing impaired cochlea and hence maximises speech intelligibility.

Chapter 8

The final chapter presents a summary of the work presented along the whole thesis. Suggestions for further work in certain areas are also given.

1.6 Contribution of the thesis

The main contributions of this thesis are:

1. Study and re-design of phase-shift radiators. The phase-shift radiators introduced here are able of producing a directivity index close to that of a hypercardioid between 500 Hz and 6 kHz.

2. Design and construction of 2 arrays to enhance the speech intelligibility of hearing impaired listeners; the 1×8 array and the 4×8 array. Both

arrays use phase-shift sources, and are able of obtaining a large directional performance in free-field and in reverberant conditions.

3. Development of a formulation to study the performance of acoustical radiators in reverberant environments, and the study of the performance of superdirective arrays in a diffuse field, by means of a theoretical model and by means of in situ measurements. It is also shown by means of simulations and measured results, how even without spatial averaging, the reverberation acts as regularization. This constrains the amount of power used and makes the array robust to source mismatches.

4. Adapting the formulation for the use of superdirective acoustic contrast maximisation, least squares and mixed control superdirective approaches into the time domain, applied for personal audio systems. This formulation allows to reduce the modelling delay needed by the superdirective filters. The performance of this formulation with respect to the traditional frequency domain design methods was studied by free-field monopole simulations and off-line predictions of performance.

5. The speech intelligibility improvement introduced by the array has been shown to be equivalent to a decrease of 4 dB SNR, as assessed by mean of objective techniques such as the speech transmission index (STI) and by behavioural tests carried away with 30 subjects.

6. Development of a physically realisable model of the hearing impaired cochlea. The model has been used to study hearing impairment and a number of psychoacoustic factors, such as frequency selectivity or upward masking. Such psychoacoustic models have suggested alternative ways of equalising the array program to further improve speech intelligibility.

The work performed in this thesis has lead to various publications and conference proceedings. These references are given in the declaration of authorship.

Chapter 2

The design of phase-shift (PS) sources

The spatial properties of the sound control depend on the direction along which the sources of an array are set and on the individual directivities of such sources. For example, a delay and sum line array, where all the sources radiate with the same amplitude and an electronic delay can be set to each source, can be steered to either left or right of the array. Although the main beam of the array can be steered, however, the directivity of the array is symmetrical with respect to its main axis, as it depends on the directionality of the individual sources. In order to create a line array that radiates only towards its front, the radiation coming from the back of the sources has to be reduced. This needs individual sources that are themselves directional, which can be obtained if each source is constituted by an smaller array, or by the use of passive directional radiators as phase-shift (PS) sources.

This chapter describes an investigation into the operation and tuning of such PS sources, with the aim of, firstly, informing how the individual sources of the arrays presented in this PhD thesis have been designed and built and, secondly, tune the individual PS sources so that its directional performance is large in a broad bandwidth. An analytical model is first introduced, which has been used to perform some initial simulations with first order sources. A more accurate model is also introduced, based on a FE simulation. This last model has been used to understand the behaviour of some previously built PS sources, and has facilitated the design and construction of more efficient radiators. The performance obtained with these new sources is then presented. The methodology used to estimate the transfer function and the directivities of the sources is shown in Appendix J.

2.1 The phase-shift source as a directional radiator

Due to the out of phase radiation of the front and rear faces of the driver, an unbaffled moving coil loudspeaker radiates with a dipole-like directivity pattern, which results in very inefficient radiation. To reduce the radiation towards the back of the source and overcome such limitation, the loudspeaker is generally placed in a closed cabinet [107]. However, when the loudspeaker is placed in a closed cabinet, the air volume kept inside the cabinet acts as an additional stiffness to the suspension of the driver. This rises the free-air resonance of the loudspeaker, which is undesirable, as the low frequency response is reduced. The directional characteristics of the source are also modified when the dimensions of a closed back loudspeaker are small compared to the wavelength of radiation, so that the sound propagates with identical energy towards each direction, behaving effectively like an omnidirectional source [108]. For a source placed in a cabinet with a maximum dimension a, it is estimated that the radiation is omnidirectional for the range of frequencies below $ka \leq 2\pi$, which for a cabinet of 35 mm is about 10 kHz. Above this frequency, when the wavelength of radiation is smaller than the source dimensions, the loudspeaker works analogously to a continuous array, and the radiation pattern is narrowed.

The only existing way of creating a more directional source is to obtain a constructive interference at one point and a destructive interference at other points. By this mechanism, a first order, or gradient source, is obtained by the superposition of the sound waves generated by two sources. Directional speakers built as an agrupation of sources have been studied previously. Olson [98] performed a large study in where various arrangements were tested, suggesting its application to column speakers. Similar applications were suggested by Pfretzschner and Romera [97], who added an extra radiator with an electronic delay to reduce the radiation coming from the back of a line array. More recent studies have used also ensembles of two sources to create more directional radiators [48, 52, 109, 57, 51], however, these have introduced the use of superdirective filters, to maximise the directional radiation over the whole frequency range. Although the back radiation of a line array can be reduced in the same way as presented in these examples, it needs twice the amount of drivers for a line array of a given number of sources, plus extra electronic processing.

A similar but more simple and cost effective way of reducing the back radiation

of a source is obtained by placing the driver in a phase-shift (PS) cabinet. A PS cabinet is a special enclosure with an opening at the rear, in which due to the special construction of the cabinet, the rear opening radiates with a certain delay with respect to the driver. Thus, the pressures created by driver and rear opening sum with identical phase at the front of the speaker and with opposite phase at the rear of the source, which obtains a cardioid-like limaçon. A similar effect is used in cardioid microphones, which was introduced by Bauer in a patent of 1941 [110]. The use of a PS network to create a directional loudspeaker was first introduced by Iding [102] and later studied by Holmes [101], who referred to the phase-shift cabinet as "acoustic resistance box".

In a PS source, two acoustic sources are involved: the loudspeaker diaphragm and the vented aperture at the rear of the cabinet. At low frequencies, where the acoustic wavelength is large compared with the dimensions of the cabinet, $ka = 2\pi$, the sound waves generated by the loudspeaker diaphragm and the cabinet rear aperture interfere in the space. If the pressure radiated by the rear aperture is delayed by an appropriate amount of time, a directional response is achieved. However, if the wavelength of radiation is smaller than the cabinet dimensions, cancelling effects appear, which create an undesired directional and frequency response. Ideally, however, if a resistive material that attenuates the radiation at high frequencies is placed in the rear aperture, a higher front radiation can be obtained.

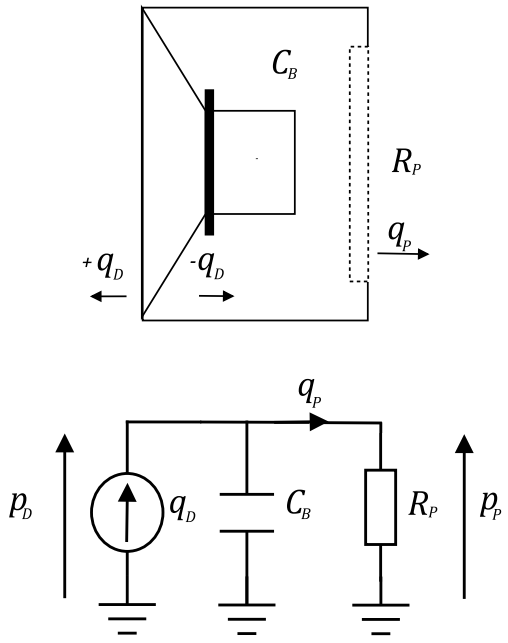


Figure 2.1: Simplified PS loudspeaker arrangement and low-frequency electroacoustic equivalent.

Assuming that the wavelength of radiation is much larger than the cabinet dimensions, a PS loudspeaker may be modelled by an acoustical resistance that is constituted by the rear port of the cabinet, and an acoustical compliance, represented by the volume of the cabinet internal cavity. A sketch of this implementation is shown in Fig. 2.1. At sufficiently low frequency, the phase-shift transmission line can be approximated using an electroacoustic circuit as shown in Fig. 2.1. In this approximation it is assumed that the acoustical reactance of the port of the cabinet is much smaller than its acoustical resistance, R_P. It is also assumed that the mechanical impedance of the cabinet is much larger than that of the source, and so it dominates. The acoustic resistance is then frequency independent, and together with the compliance of the interior of the cabinet introduces a delay that increases linearly with frequency [103].

Considering that $q_D(j\omega) = q_D e^{j\omega t}$, and that for a single frequency this is written for simplicity as q_D, the volume velocity of the driver may be expressed as a function of the force driving the diaphragm, f_D, and its total mechanical impedance, $Z_{M,Tot}$

$$q_D = \frac{f_D}{Z_{M,Tot}} = \frac{f_D}{Z_{MD} + \left(\frac{1}{R_P} + j\omega C_B\right)^{-1}}. \tag{2.1}$$

Assuming the mechanical impedance of the driver to be small compared to the port impedance, the volume velocity at the port can be written as

$$q_P = \frac{\left(\frac{1}{R_P} + j\omega C_B\right)^{-1}}{R_P} q_D = \frac{1}{1 + j\omega R_P C_B} q_D, \tag{2.2}$$

which gives a relation of $|\frac{q_D}{q_P}| = 1$ for frequencies where $\omega < \frac{1}{R_P C_B}$, being the phase between the two pressures established by $\angle \frac{q_D}{q_P} = \tan^{-1}(\omega R_P C_B)$. At low frequency, the phase-shift is controlled by the acoustical resistance of the port, R_P, and the compliance of the cabinet volume, C_B, written as

$$\phi = \omega C_B R_P, \tag{2.3}$$

where C_B and R_P form the time constant of the phase-shift, τ

$$\tau = C_B R_P. \tag{2.4}$$

This constant can also be given in terms of the equivalent path length difference, d

$$d = c_0 C_B R_P \tag{2.5}$$

The selection of these parameters lie on the dimensions of the cabinet and the characteristics of the resistive material. The cabinet compliance is given by [104]

$$C_B = \frac{V}{\rho_0 c_0^2}, \tag{2.6}$$

where V refers to the volume of the enclosure, ρ_0, is the density of air, 1.21 kg.m^{-3} and c_0=344 ms^{-1}, is the speed of sound in the air. By substituting Eq. 2.6 into Eq. 2.5 the equivalent path length may be written as

$$d = \frac{VR_P}{\rho_0 c_0}. \tag{2.7}$$

If the acoustical parameters are chosen adequately, it can be assumed that the phase-shift network does not change the magnitude of the pressure radiated by the rear of the system, and therefore that the magnitude of diaphragm and port pressures are equal [55]. Under this assumption the complex volume velocity of the port can be written as

$$q_P = -q_D e^{-jkd}, \tag{2.8}$$

where the phase inversion that the port introduces is denoted by $-q$, being e^{-jkd} the frequency dependent delay due to the phase-shift. A simple model using monopoles can now be derived. For the case of the diaphragm pressure

$$p_D = j\omega\rho_0 c_0 q_D \frac{e^{-jkr_D}}{4\pi r_D}, \tag{2.9}$$

where ω represents the angular pulsation frequency, q_D the diaphragm volume velocity, and r_D the distance between the diaphragm and the receiver position. The port pressure is given by

$$p_P = j\omega\rho_0 c_0 q_P \frac{e^{-jkr_P}}{4\pi r_P}, \tag{2.10}$$

being q_P the port volume velocity and r_P the distance between the port and the receiver position. The total pressure is given by the summation of diaphragm and port pressures

$$p_{Tot} = p_D + p_P = j\omega\rho_0 c_0 \left[q_D \frac{e^{-jkr_D}}{4\pi r_D} + q_P \frac{e^{-jkr_P}}{4\pi r_P} \right]. \tag{2.11}$$

By using the cosine rule [55], the distances of diaphragm and port at an angle θ around the source are given by

$$r_D = \sqrt{r_0^2 + l^2/4 + r_0 l\cos\theta} \qquad r_P = \sqrt{r_0^2 + l^2/4 + r_0 l\cos(\pi - \theta)}, \tag{2.12}$$

where l represents the separation between port and diaphragm of the source. Substituting Eq. 2.8 into Eq. 2.11 and taking the far-field approximation of r_D, the resultant far-field total pressure is given by

$$p_{Tot} = \frac{j\omega\rho_0 c_0}{4\pi r_D} q_D e^{-jkr_D} \left(1 - e^{-jk(l\cos\theta + d)}\right) = p_D \left(1 - e^{-jk(l\cos\theta + d)}\right). \tag{2.13}$$

It is then assumed that the equivalent path length, d, is small compared with the wavelength if

$$k(l\cos\theta + d) = \frac{2\pi l\cos\theta + d}{\lambda} \ll 1 \tag{2.14}$$

Using a small angle approximation [111], the exponential term in Eq. 2.13 can be written as

$$e^{-jk(l\cos\theta + d)} = 1 - jk(l\cos\theta + d). \tag{2.15}$$

The total pressure radiated by the loudspeaker can be approximated by

$$p_{Tot}(\theta) = jkp_D(l\cos\theta + d). \tag{2.16}$$

This means that by choosing the right ratio between the parameters l and d, a directive limaçon can be obtained. The hypercardioid limaçon is the first order pattern that minimises the power input to the reverberant field [104]. This can be obtained if $d = l/3$ [112], which is got by using $R_P = \frac{l\rho_0 c_0}{3V}$.

2.2 Understanding the behaviour of a phase-shift source

2.2.1 First performance measures of a phase-shift source

The first study of a PS source was performed when searching for a suitable source for the array presented in [94, 113]. However, little was known at this time about the behaviour of such sources. A sample speaker was built, in which the rear aperture size was made variable, with the frequency dependent resistance constituted by a fine metallic mesh. A set of measurements was then produced, with different port apertures, to determine the aperture size which was giving a radiation pattern closer to a hypercardioid. Once an optimum aperture size was determined, 8 sources where built, posteriorly used in the array shown in [113] and in a personal audio system for a car cabin [58]. The actual aspect of the prototype source and of the actual sources is shown in Fig. 2.2a, and the horizontal and vertical radiation patterns obtained with one of these sources can be observed in Fig. 2.3, in where the directional response is shown at various discrete frequencies.

It is possible to observe that at some frequencies the radiation pattern is quite close to that of a hypercardioid, as for example at 500 Hz, 1 kHz and 2 kHz. At other frequencies, the response is also quite directional, as the loudspeaker has started to beam due to the fact that the wavelength of radiation is smaller than the dimensions of its diaphragm.

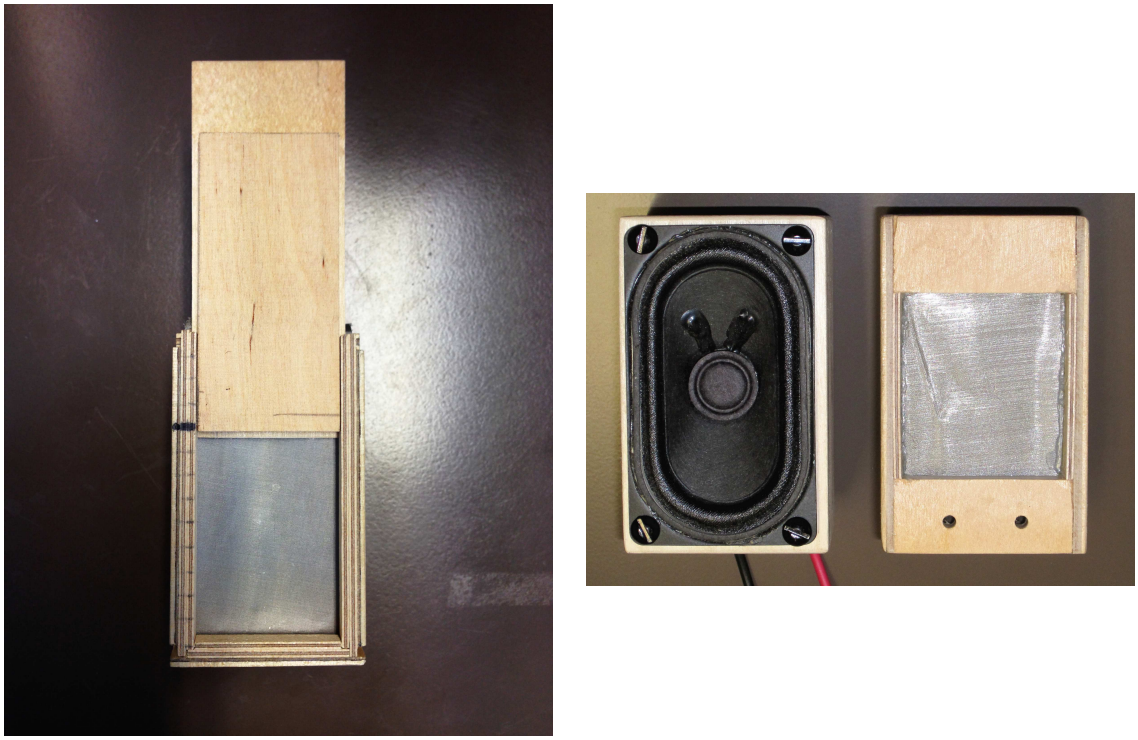


Figure 2.2: Prototype and actual source for the array described in [94, 113].

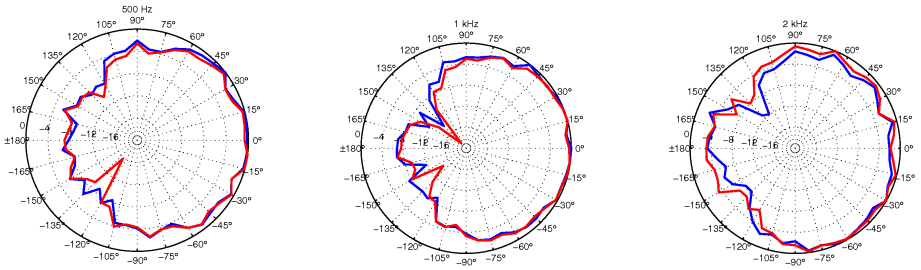
2.2.2 Comparison of transfer functions between a phase-shift source and a closed-back speaker

A large number of the simulations included in this thesis use point hypercardioid sources to simulate a phase-shift source array. Using the "product theorem" [114], the transfer impedance of a point monopole source simulating a driver in an array, Eq. 2.9, can be multiplied by the directional pattern of a hypercardioid source to give the transfer impedance of an idealized first order point hypercardioid source, whose magnitude depends on the orientation angle

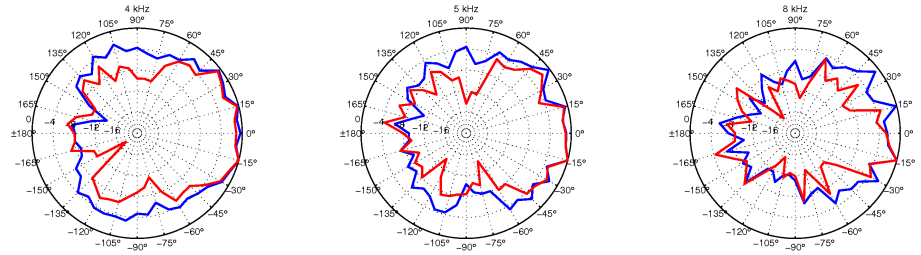
$$Z_D = ZD(\Psi, \theta, \phi) = j\omega\rho\frac{e^{-jkr}}{4\pi r}\left(1 - \Psi + \Psi\cos\theta\cos\phi\right), \tag{2.17}$$

where for the case of a hypercardioid source Ψ is set to 0.75.

The directional performance of various arrays is always compared with these using the same amount of electrical energy, which is quantified by the array effort,



(a) 500 Hz **(b)** 1 kHz **(c)** 2 kHz



(d) 4 kHz **(e)** 5 kHz **(f)** 8 kHz

Figure 2.3: Measured directivities for the PS source shown in Fig. 2.2a at various discrete frequencies. Horizontal directivities shown in blue and vertical directivities shown in red. Results reproduced from [94].

described in Section 3.1.5, which requires the knowledge of the individual sources transfer impedances. Looking at Eq. 2.17, it can be observed how the radiation pattern is invariant with frequency and how the value of the transfer impedance at $\theta = 0, \phi = 0$ is equal to that of a point monopole source. If an array of point monopole sources is to be compared with one of point hypercardioid sources, when both of them use the same array effort, it is needed that the transfer impedances of the *practical* source which they simulate, i.e., a closed back loudspeaker or a phase-shift source, have a similar measured transfer impedance.

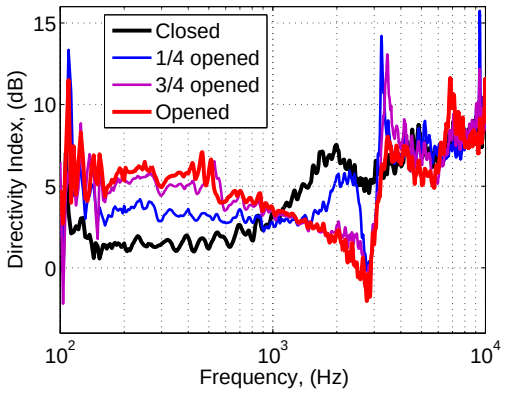


Figure 2.4: Measured DI of the prototype source shown in Fig. 2.2a for different openings of the rear port.

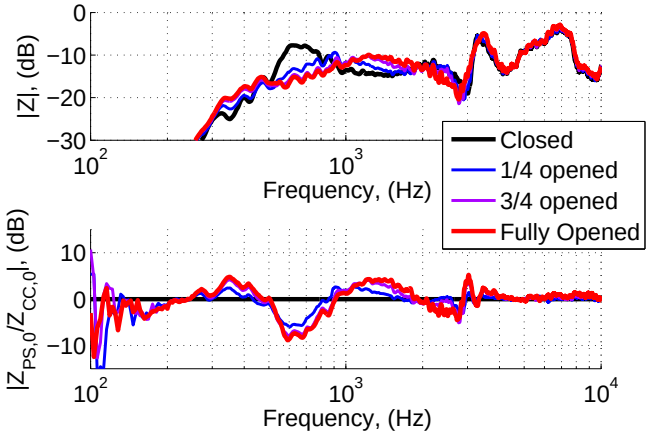


Figure 2.5: On axis frequency response of the prototype source (upper plot) and relative differences in level with respect to the closed back configuration (lower plot) for different opening sizes of the rear port.

In order to make sure that the transfer impedances of a closed back source and a phase-shift source are similar, a comparison can be made of the transfer impedances of the prototype source, when this is working in closed back and phase-shift configuration Fig. 2.5 shows the corresponding frequency response for different openings of the rear port. The absolute value of the transfer impedance obtained for different aperture configurations is relatively similar for the whole of the frequency range. Around 800 Hz, where the cardioid effect is greater, the closed back source is about 8 dB more efficient, as all the radiation is directed towards the front and not through the rear port. Above 1 kHz the PS source is about 4 dB more efficient, and above 3 kHz the transfer impedance is very similar for all configurations. As the variability of the transfer impedances is small, the point hypercardioids can be modelled using Eq. 2.17.

2.2.3 Directional response

Apart from observing the radiation pattern of a PS source at a given frequency, it is also very useful to quantify its directional characteristics along frequency. To this end, the directivity index is used. The DI is widely used to evaluate the directional response of a radiator, and it is defined as the magnitude, in decibels, of the ratio between the on axis pressure that a source produces compared to that produced by an omnidirectional source, when both radiate the same acoustical power. Considering a loudspeaker placed on the x, y plane of Fig. 2.6, the

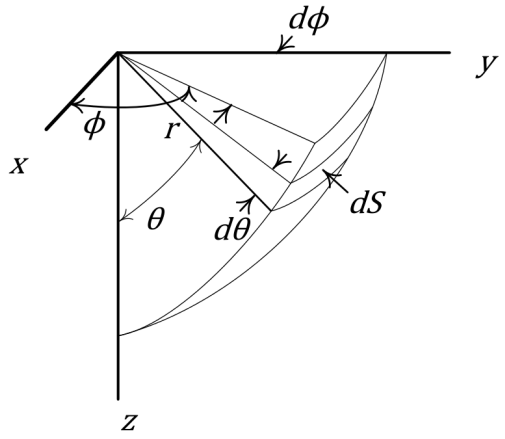


Figure 2.6: Coordinate system defining the angles θ and ϕ and the radius r used to estimate the directivity index (DI). The area of the surface differential is $dS = \sin\theta d\theta d\phi$.

mathematical expression of the DI is given by [82]

$$DI = 10\log_{10}\left(\frac{4\pi|p_{ax}|^2}{\int_0^{2\pi}\int_{-\pi/2}^{\pi/2}|p(\phi,\theta,r)|^2|\cos\theta|d\theta d\phi}\right), \tag{2.18}$$

where p_{ax} refers to the on axis pressure. If the pressure is obtained at a set of discrete points the DI is given by

$$DI = 10\log_{10}\left(\frac{4\pi|p_{ax}|^2}{\sum_{m=1}^{2\pi/\Delta_\theta}\sum_{n=1}^{\pi/\Delta_\phi}|p(\phi_n,\theta_m,r)|^2|\cos\theta_n|\Delta_\theta\Delta_\phi}\right), \tag{2.19}$$

where Δ_θ is the number of horizontal measurement points between 0 and 2π and Δ_ϕ is the number of vertical control points between $-pi/2$ and $pi/2$. If the estimation is only performed in a horizontal measurement slide, the above expression reduces to

$$DI = 10\log_{10}\left(\frac{2\pi|p_{ax}|^2}{\sum_0^{2\pi}|p(\phi)|^2\Delta_\phi}\right). \tag{2.20}$$

The above expression, also known as the "partial directivity index", is a generalisation of the DI formula used when the directivity estimation is performed in horizontal measurement slide which extends between 0 radians and 2π radians. In this case, the DI is the ratio between the on axis pressure that a source produces compared to that produced by an omnidirectional source in a two-dimensional measurement slide. As the majority of the measurements presented in this thesis

are done in a two-dimensional horizontal measurement slide, Eq. 2.20 is used to quantify the directional behaviour against frequency.

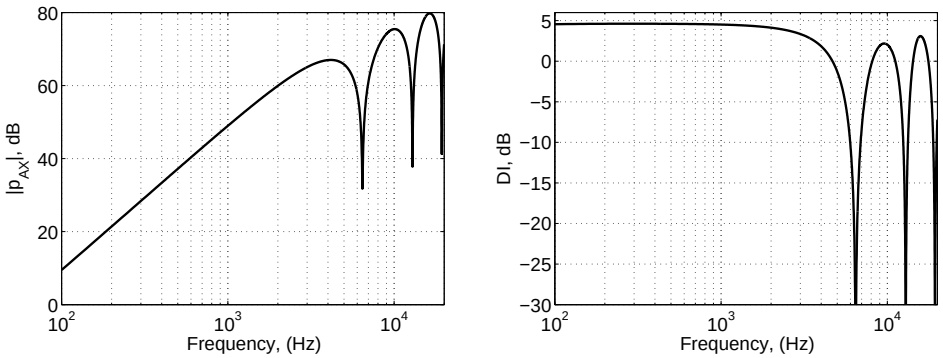


Figure 2.7: On-axis pressure (left plot) and DI along frequency (right plot) of an ideal hypercardioid source with $d = 4$ cm in a horizontal measurement slide.

The frequency response of the on axis pressure, p_{AX}, and the directional response of a PS source depend entirely on the separation between driver and rear aperture of the source. As a first design tool, these responses can be obtained using point monopole sources as shown in Eq. 2.11, with the source strengths of driver and rear port related as

$$q_P = -q_D e^{-jkc_0 C_B R_P}, \tag{2.21}$$

where in order to give a hypercardioid radiation pattern $c_0 C_B R_P = l/3$. Using such formulation, the response for a source with $l = 4$ cm is shown in Fig. 2.7. The frequency response grows proportionally to the frequency of radiation, as expected from Eq. 2.9 until around 6.5 kHz, where a big notch can be observed, which is common with the simulated DI. For a hypercardioid source the null in the responses is found at a frequency

$$f_{Notch} = \frac{3c_0}{4d}, \tag{2.22}$$

which for a source with spacing d of 4 cm is 6.45 kHz. At this frequency the radiated wavelength is 4/3 times the separation between both monopole sources, i.e., driver and rear port in a practical source. Driver and rear port transfer impedances at $\theta = 0$ can be related as

$$Z_P = Z_D e^{jkl}. \tag{2.23}$$

The pressure at $\theta = 0$ can be calculated using Eq. 2.21 for an optimal case and

Eq. 2.23. At f_{Notch} it is obtained that $j2kl/3 = 2\pi$. which leads to

$$p(\theta = 0) = Z_D q_D - Z_D e^{jkl} q_D e^{-jkl/3} = Z_D q_D - Z_D q_D e^{j4kl/3} = Z_D q_D - Z_D q_D = 0. \tag{2.24}$$

which explains the null on the radiated pressure. This effect can be visualised in Fig. 2.8, where it is seen how the radiation is directed towards the back of the source. For the scope of this thesis and for the majority of the applications which require a directional source, this effect is undesirable, as it contributes directly to inject energy into the reverberant field. In further sections of this chapter it is shown how the reduction of directivity at f_{Notch} is diminished by the use of an improved rear port resistance.

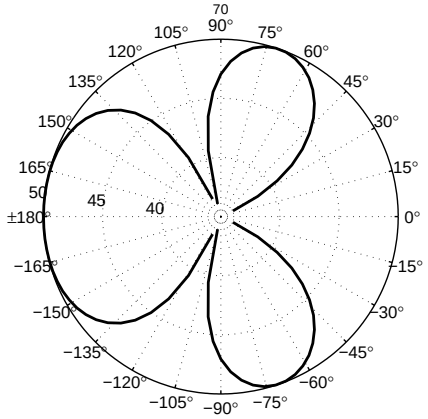


Figure 2.8: Directivity pattern of a theoretical PS hypercardioid source at f_{Notch}, where it can be observed how the radiation at 0° is largely reduced.

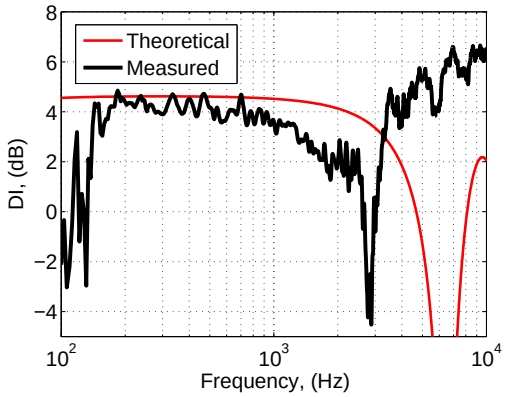


Figure 2.9: Measured and simulated directivity index of a PS source with $d = 4$ cm.

Fig. 2.9 shows a comparison between the theoretical DI of a PS source tuned to give a hypercardioid directivity pattern, and the measured DI of the source shown in Fig.2.2a. At low frequency the DI is close to the predicted theoretical value, however the notch this case is centred around 2.5 kHz, a much lower frequency

than the theoretically predicted one. In a practical source, this result is caused by the finite size of the cabinet. The *effective point* at which the soundwaves start forming is denominated *acoustic centre*. The IEC standard 61094-3 defines the acoustic centre as [115]: "For a sound emitting transducer, for a sinusoidal signal of a given frequency and for a specified direction and distance, the point from which the approximately spherical wavefronts, as observed in a small region around the observation point, appear to diverge". This problem has been studied previously by Elliott and Jones [48] and by Vanderkooy [116, 117, 118]. Elliot and Jones [48] analytically showed that the acoustic centre is placed half a radius away from the driver and Vanderkooy stated that as a general rule of thumb the acoustic centre lies in front of a loudspeaker diaphragm at a distance approximately equal to half-width of its cabinet. The approaches presented by both authors agree with the results observed for the practical source, as if f_{Notch} is calculated in this way

$$f_{Notch}(42 \text{ mm}) = \frac{3 \cdot 343}{4(l + 2w/2)} = 3062Hz, \tag{2.25}$$

where l = 4.6 cm and w=4.2 cm, which gives a value close to the frequency of the null shown in Fig. 2.9. However, if the same approximation for the acoustic centre is used for smaller values of d, the theoretical values do not match with the correspondent f_{Notch}. The acoustic centre was later studied by Cheer [56, 57], who observed the phase difference of the signal arriving from two sources which where either acting as a line array or placed in a baffle. The author showed that placing the sources in a baffle greatly increased the effective separation between the two sources.

Although it is well understood why the directivity is reduced at f_{Notch}, and although it is possible to find a fair approximation of the acoustic centre for the cases when d is relatively large, a more accurate model of a PS source was desired, which suggested the building of a finite element model.

2.2.4 Study of the dimensions effect through a finite element model

The model has been built in the commercial software Comsol v4.3, with a geometry as that shown in Fig. 2.10. The model is built up using similar specifications as those used by Cheer *et al.* [56], in which the authors studied the effect of the baffle introduced by a mobile phone. The domain wherein the model is placed is made of two concentric spheres with dimensions varying depending on the frequency range of study as shown in Table 2.1, which allows the study between 100

Hz and 8 kHz.

Table 2.1: Dimensions of the finite element models.

Model	**Bandwidth**	Sensor array radius, (cm)	PML radius, (cm)
LF	25.94	23.89	26.14
MF	25.94	23.89	26.14
HF	25.94	23.89	26.14

The inner sphere is the air domain, wherein the response between the source and a set of 48 estimation points is obtained, in order to posteriorly calculate the DI. The inner sphere has to be large enough to allow the sensors being in the geometric far-field, which is obtained if the sensors are placed at a radius $r \geq l$ [105]. The second, outer sphere forms a shell around the air domain, and is constituted by a perfectly matched layer (PML) [119]. The PML is a method of fulfilling the Sommerfield radiation condition, which ensures that outgoing waves propagate to infinity in the absence of reflecting boundaries. The PML is a layer of anisotropic damping which increases with distance into the domain perpendicularly to the interference between the physical domain, the air domain, and the PML [119], therefore allowing to simulate an anechoic-like environment. Apart from that, the domain of study can be truncated, which leads to a reduction of the computational demand without significant loss in the accuracy of the solution [119].

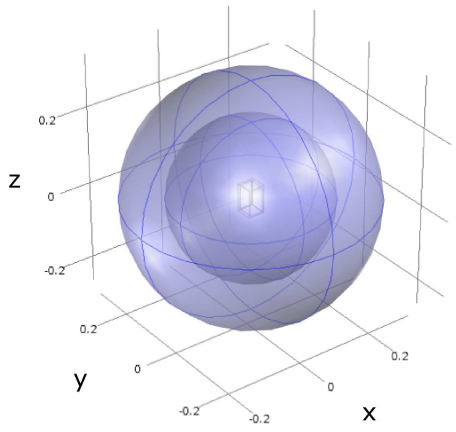


Figure 2.10: Finite element (FE) geometry used for the study of the PS source.

The phase-shift source is modelled as a hollow baffle made out of wood, with 4 mm walls. The driver is constituted by a vibrating square piston and the rear port by an acoustic impedance. In order to assess the accuracy of the finite element model, the baffle and piston simulating the driver are first built with dimensions

as the source shown in Fig. 2.2a, so that the result can be compared with DI measurements of the source in closed back and in hypercardioid configuration. These results are shown in Fig. 2.11. For modelling the speaker working in hypercardioid configuration, a value of port impedance of 350 Pasm^{-1} has been used, taken from measurements of the same material performed by Cheer [55] . Although the impedance of the metallic mesh used in the practical source is assumed to vary with frequency, a constant value has been used as a starting design parameter. Despite some differences, the results offered by the finite element simulations are close to the measured DIs, which validates the finite element model. When the source is operating in PS configuration both FE simulated and measured results are in agreement, with the drop in the DI also occurring at a similar frequency. The reduction of DI at f_{Notch} is, however, underestimated by the finite element model. Up to 5 kHz the measured results give a higher DI, which is believed to be caused by a higher impedance at high frequencies of the practical phase-shift source rear port (a constant value is used in the FE model along frequency), and due to the beaming of the loudspeaker diaphragm. The measured and FE simulated results for the closed back source are also in a reasonably good agreement, matching closely above 2 kHz. Although the model presents some differences with respect to the practical source, it allows a good estimation of the reduction of directive characteristics at f_{Notch}.

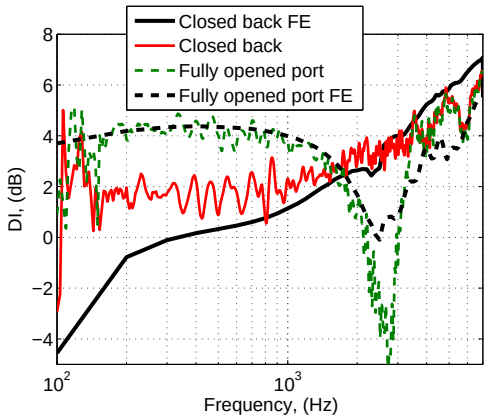


Figure 2.11: Measured and FE simulated DI of the source shown in Fig.2.2a, for hypercardioid and closed back configurations.

In order for the array to give a proper amplification that allows increasing speech intelligibility, it is intended that the array is able to control the bandwidth between 500 Hz and 8 kHz, which implies that f_{Notch} has to be moved towards higher frequencies and/or its effect has to be reduced. To study the effect that the effective separation between driver and rear port has on the DI, a first set of simulations was carried with sources with a smaller length. With views to a

posterior construction of the source, the dimensions used for the piston simulating the driver were taken from a miniature speaker, the Visaton k20.40, shown in Appendix A. The dimensions of the two models built are shown in Table. 2.2. A smaller cabinet means a lower internal volume, which then modifies the ratio $\frac{VR_P}{\rho_0 c_0}$, which has to be equal to $l/3$ to lead to a hypercardioid radiation pattern. In order to keep this ratio constant the acoustic resistance of the port is increased. In a practical source the acoustic resistance can be modified by selecting a different metallic mesh or by just adding one layer of metallic mesh after the other. Assuming that linearity holds, two layers of resistive materials correspond to a resistance of $2 \cdot (350)$ Pasm^{-1}. The DI obtained using such model sources are shown in Fig. 2.12. It can be observed how the DI below 4 kHz is close to that predicted by a hypercardioid, and how at about 5 kHz for the 12 mm and 5.5 kHz for the 6 mm source, a small reduction in the DI is found due to the explanation of Section 2.2.1. The reduction in the DI at f_{Notch} is, however, underestimated.

Table 2.2: Dimensions of the "short" PS sources studied in the FE model.

Source	h, (cm)	w, (cm)	l, (cm)
Array 1 [113]	7.50	4.60	4.20
Source 1	44.00	26.00	6.00
Source 2	44.00	26.00	12.00

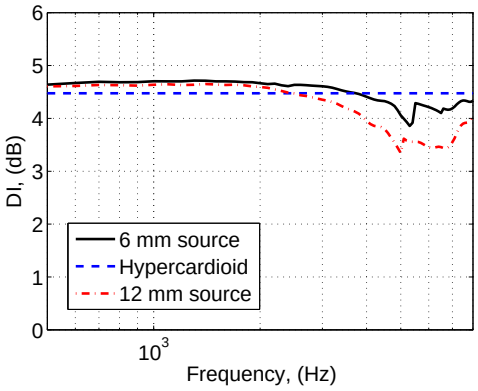


Figure 2.12: DI performance of the FE model for the sources with dimensions shown in Table 2.2, compared with the DI of a theoretical hypercardioid source.

The results obtained with the finite element simulations, shown in Fig. 2.12, lead to the construction of two prototype sources with a shorter cabinet length, denominated "Source 1" for the 6 mm source and "Source 2" for the 12 mm source. A close up of both sources can be observed in Fig. 2.13. As it can be observed in Fig. 2.14, although a flat DI is seen between 500 Hz and 4 kHz, a large null is present at about 5.5 kHz, a very different result from that obtained in

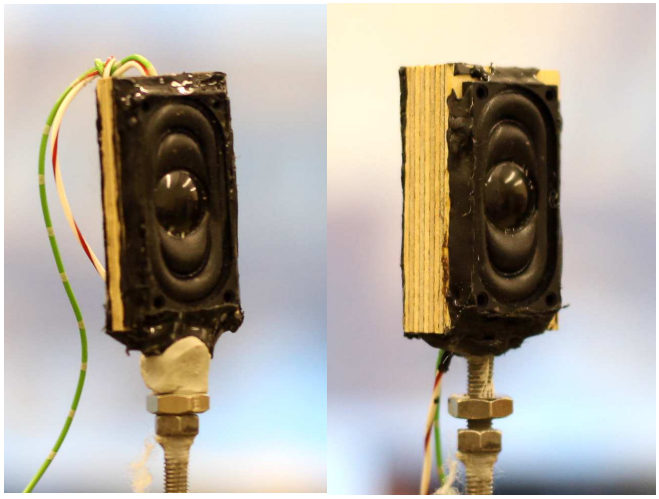


Figure 2.13: Prototype " short" PS sources.

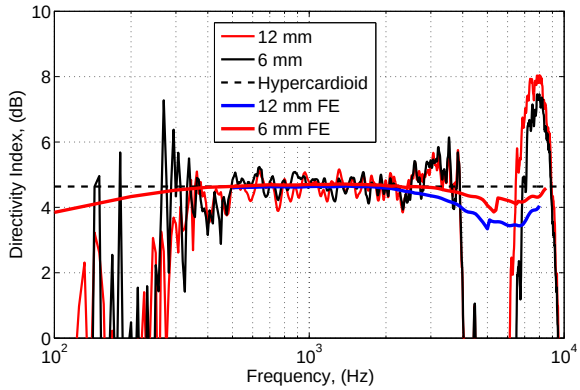
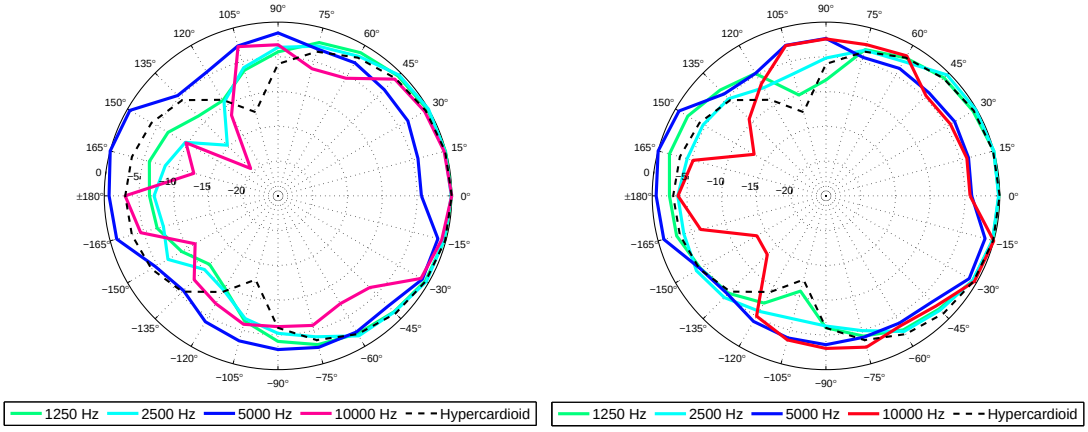


Figure 2.14: Measured DI of the "short" prototype PS sources in comparison with the FE simulations. The theoretical DI given by a hypercardioid is also shown.

the FE simulations. It also can be observed in the directivities, presented in Fig. 2.15, how at 5 kHz the directivity tends to that shown in Fig. 2.8. The results for both prototype loudspeakers are very similar, as only 6 mm difference exist between both sources. It is needed to consider the width of the cabinet in this case, w=26 mm. If f_{Notch} is calculated using the rule of thumb of Vanderkooy, the null frequency is obtained at

$$f_{Notch}(26) = \frac{3 \cdot 343}{4(w + 2w/2)} = 4950Hz, \tag{2.26}$$

where in this case the wide of the speaker is of w=2.6 cm, which gives a figure closer to the result obtained in the measurements result. The FE simulations show a small descent in the DI centred at the frequency of the null, however, not so deep as the measurements result. Being the width of the cabinets bigger than their length, it is reasonable to say that this dimension dominates in the creation of the acoustic centre.



(a) 12 mm source **(b)** 6 mm source

Figure 2.15: Measured directivity patterns of the 12 mm and 6 mm prototype sources at different frequencies.

2.3 Towards an improved source: mixed acoustic mass-resistance source

As both prototype sources show a big null centred at about 5 kHz, they are not suitable for accomplish the desired objectives of a very directional source from 500 Hz to 8 kHz. The next approach taken has been to include a frequency dependent resistance of the rear-port and hence reduce the back radiation at f_{Notch}. The idea is to reduce the radiation of the rear port as the frequency is increased, so that the PS source tends to work in closed cabinet configuration, minimising the reduction in DI at f_{Notch} and obtaining directivity due to the beaming of the driver itself.

An initial analytical model has been developed to illustrate the behaviour with frequency of such a port. A practical loudspeaker which is able of avoiding the lack of directional performance at f_{Notch} is shown at the end of this section.

2.3.1 Analytical model

A modification of the mixed-mass resistance element model defined by Beranek [82] is used here to simulate the effect that such a rear port impedance can have on the directivity of a PS source, denominated here "inertance" model. Although in the practical speaker the frequency dependent resistance has been built with a high frequency foam, the model presented here is based on a perforated grill for the frequency dependent behaviour and a metallic mesh for the rear port resistance. According to Beranek [82], the equations that define the impedance

of a hole of the perforated sheet are

$$Z_A = R_A + j\omega M_A, \tag{2.27}$$

where

$$R_A = \frac{1}{\pi a^2}\rho_0\sqrt{2\omega\mu}\left[\frac{t}{a} + 2\left(1 - \frac{A_h}{A_b}\right)\right], \tag{2.28}$$

and

$$M_A = \frac{\rho_0}{\pi a^2}\left[t + 1.7a\left(1 - \frac{a}{b}\right)\right], \tag{2.29}$$

where μ=1.56·10^{-5} m^{2}s^{-1} is the kinematic coefficient of viscosity, a is the radius of the holes, t is the thickness of the plate, and b is the side of a square that surrounds the holes. A_h is the area of the hole in square meters and A_b is the area of the square that surrounds the holes in square meters. If there are n holes in the perforated sheet, the above equations must be divided by n. A lumped parameter model can be used now, with the circuit represented by an electroacoustic analogy. The perforated sheet is formed by a resistance plus a coil in series, with the thin metallic grid also as a resistance and a coil. The volume velocity of the port depends on that of the driver as

$$q_P = \frac{\frac{-j}{\omega C_B}}{Z_P - \frac{j}{\omega C_B}} q_D = \frac{1}{1 + j\omega Z_P C_B} q_D, \tag{2.30}$$

where

$$Z_P = R_A + R_P + j\omega(M_A + M_P). \tag{2.31}$$

being R_P and M_P the real and complex part of the impedance of the thin metallic grid, normalised by the area corresponding to the holes of the perforated sheet. This is done because in the model it is assumed that the thin metallic grid is placed beside the perforated sheet. The perforated sheet is built up with the parameters of Table 2.3.

Table 2.3: Parameters used in Equations 2.28 and 2.29 for the perforated sheet constituting the impedance of the inertance model.

Parameter	mm
a	1
b	2
t	20

A bode diagram of the ratio q_P/q_D is shown in Fig. 2.16 for different rear port configurations. The "optimal" case refers to the use of $R_P = \frac{l\rho_0 c_0}{3V}$, the "measured" case refers to the use of $R_P = 350 + j75$ Pasm^{-1} and the inertance case

refers to the use of Eq. 2.31. For the optimal case the ratio q_P/q_D is of about 0 dB until about 2 kHz, whilst that for the case when the measured impedance is used the ratio q_P/q_D falls apart above 200 Hz. If the inertance model is applied the magnitude of q_P/q_D becomes slightly positive at about 2 kHz, but sees a fast decrease as the frequency is increased. Fig. 2.17 shows the DI for the three rear port configurations. Whilst the highest one is given by the optimal case, that of the measured case is also very similar. The results for the inertance rear port are almost identical to that of the optimal case until above 500 Hz, decreasing then as frequency increases and getting negative at around 4 kHz. The directivities that are obtained with the inertance rear port configuration are shown in Fig. 2.18.

As the frequency is increased, the response of the point monopole source simulation using the inertance rear port becomes closer to that of an omnidirectional source, as only the driver is radiating, leading to a DI which is 0. In a practical source, however, the directivity of the driver itself increases with frequency, which then leads to a higher DI at high frequency.

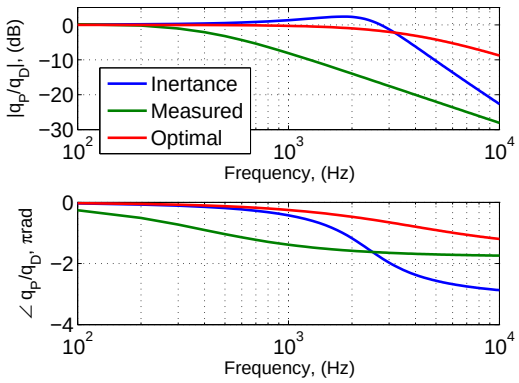


Figure 2.16: Bode diagram of the ratio q_P/q_D for different PS source resistance configurations. "Inertance" uses the model here defined. "Measured" uses the impedances measured by Cheer [55] and "Optimal" uses the optimal impedance, where $R_P = \frac{l\rho_0 c_0}{3V}$.

2.3.2 Practical sources

Based on the formulation presented above, it is possible to build a source which at low frequency presents a hypercardioid directivity but that at high frequency works as a close cabinet. At the moment of the construction of a practical PS source, the most difficult part is, however, to find the appropriate materials for the rear port impedance.

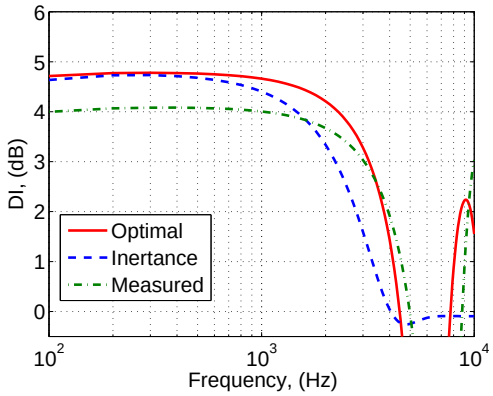


Figure 2.17: DI performance obtained using the monopole model for the different rear port impedance configurations of Fig. 2.16

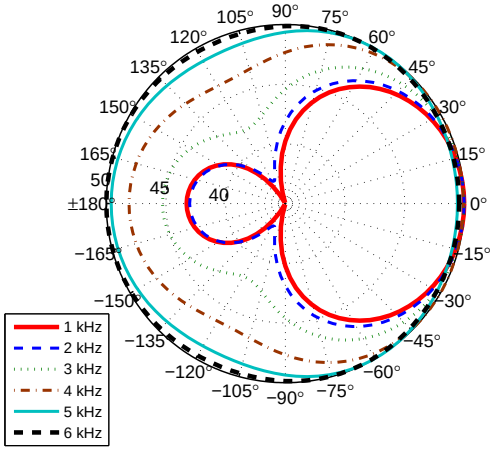


Figure 2.18: Directivities at various frequencies using the inertance model as rear port resistance.

Due to the difficulties using a grid of perforated elements, the inertance-type impedance for the rear port has been constructed by placing layers of high frequency absorbent material at the rear of the loudspeaker. The model of Eq. 2.30 can be now generalised so that driver and rear port transfer impedances are related by

$$q_P = \frac{1}{1 + j\omega R_P C_B - \omega^2 L_P C_B} q_D, \tag{2.32}$$

where in this case L_P is the frequency dependent acoustic resistance.

In the same way as performed when selecting the area of the metallic grid needed to obtain a hypercardioid radiation pattern, see Sect. 2.2.1, an exhaustive set of measurements was carried out until finding an appropriate material for the rear port. The construction adopted has been to superimpose various layer of a high frequency absorbent material. The number of layers needed has been selected by

looking at the DI and also by looking at predictions of acoustic contrast[1]. An aspect of the first prototype can be seen in Fig. 2.19. An improved version of the prototype source is used in the 1×8 array, which can be observed in Fig. 2.21

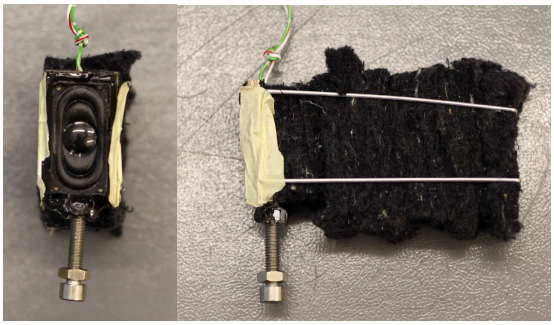


Figure 2.19: Close up of the 6 mm prototype PS source with high frequency absorbent material to form an inertance-type radiation port.

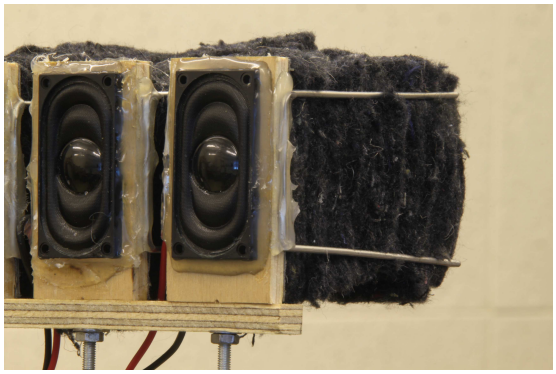


Figure 2.20: Close up of the sources used in the 1×8 array [53].

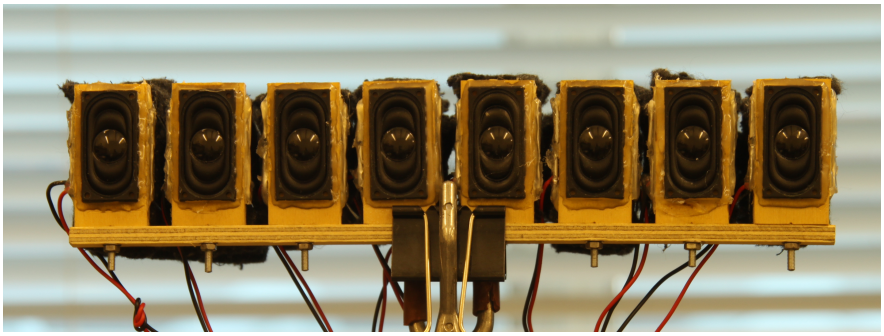


Figure 2.21: The 1×8 array [53].

The resulting PS sources have a *damped* hypercardioid radiation pattern, as it can be observed in Fig. 2.23, where the polar plots tend more towards a cardioid type than to a hypercardioid type, however the measured DI is still very high. The

[1]See Appendix D for an explanation on how the acoustic contrast is predicted.

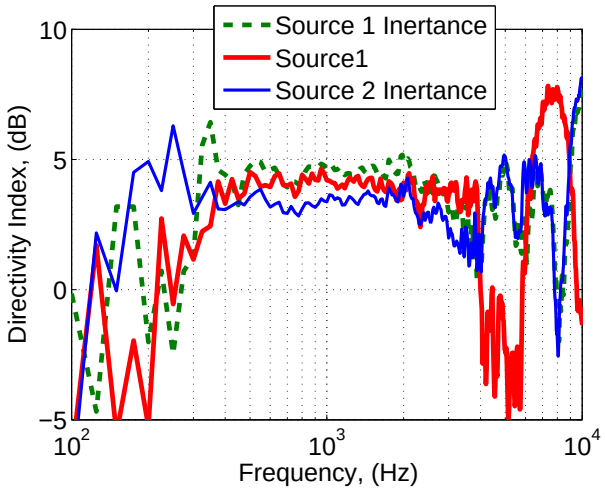
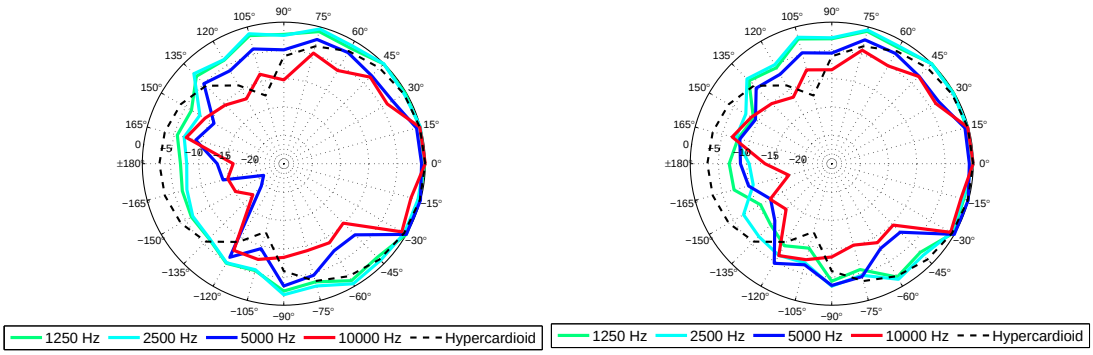


Figure 2.22: Measured DI of the two prototype sources shown in Fig. 2.12 with high frequency absorbent material added to the rear port, in comparison when only a metallic grid is used for the rear port resistance.

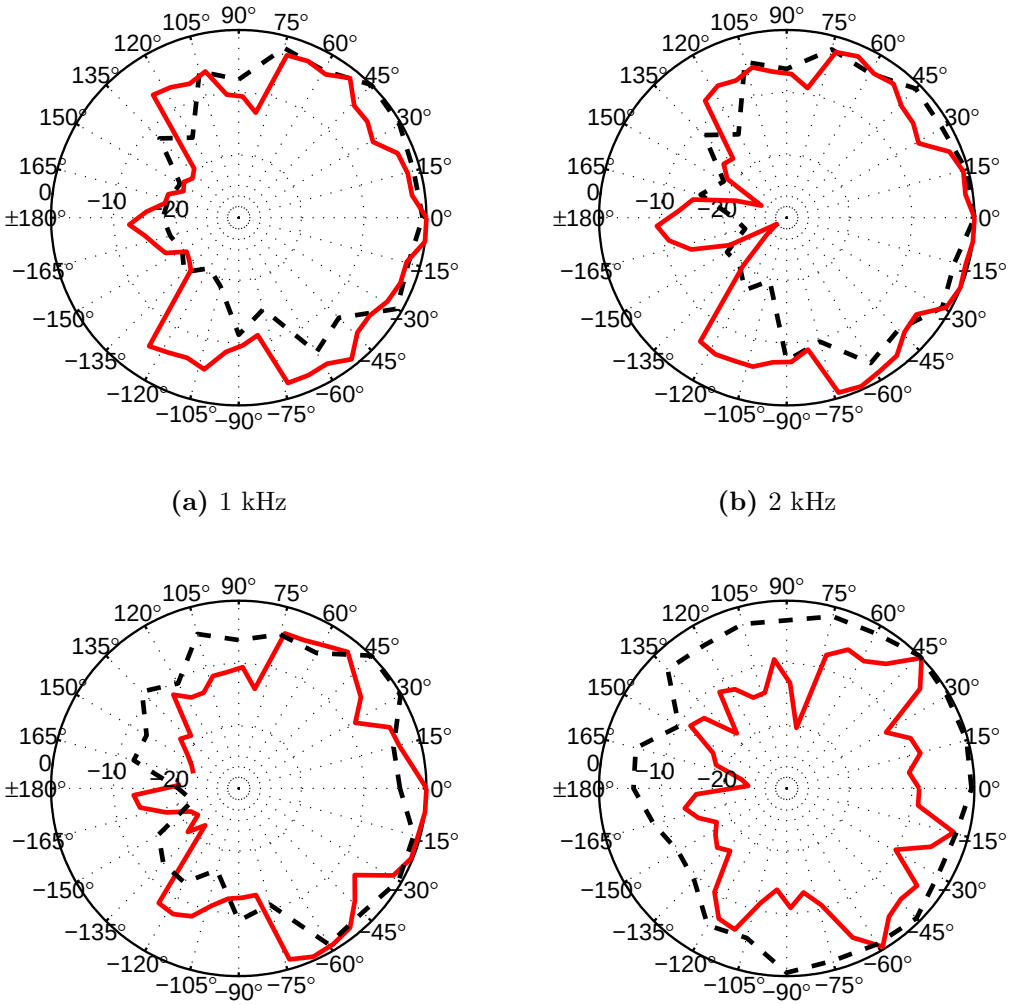


(a) 12 mm source **(b)** 6 mm source

Figure 2.23: Measured directivity patterns of the prototype PS sources with inertance-type radiation port at different frequencies.

effect of the high frequency resistance is quite notable, with the nulls of directivity greatly reduced, obtaining a DI which is more constant along frequency. This can be observed as well comparing the directivities of Fig. 2.23 against those of Fig. 2.15 at 5 kHz, due to the reduction of radiation from the back at high frequencies.

A comparison of the directivities between these obtained with the prototype source and those for the sources of the 1×8, which include the effect of the baffling from other sources, is shown for various frequencies in Fig. 2.24. Both sources show a similar directivity at 1 kHz and at 2 kHz. At 4 kHz it can be observed the effect of the baffle by other sources, as for example at 90° the radiation pattern is lower. At 6 kHz, which approximately coincides with f_{Notch}, the directivity of the single source radiates more omnidirectionally.



(a) 1 kHz **(b)** 2 kHz

(c) 4 kHz **(d)** 6 kHz

Figure 2.24: Measured directivity patterns of the prototype PS source with inertance-type radiation port (black dashed lines) compared to that obtained with the 4th source of the 1×8 array (red solid lines).

2.4 The 4×8 array sources

The 4×8 array, which is described in Section 3.4, uses sources which are directional in a 3D sense. The array, shown in Fig. 2.25, is a planar array with 32 sources arranged in 8 columns of 4 sources. Each column of the array is considered an individual source, which is made up of 4 sources as that shown in Fig. 2.20, so that a vertical beamforming is obtained and the energy input towards the reverberant field is reduced [95].

In order to obtain a *good* vertical beamforming it is important that each source radiates with equal amplitude and phase. Generally, however, every loudspeaker

Figure 2.25: The 4×8 line array, which uses a total of 32 sources, arranged in 8 columns of 4 sources each. The sources of each column are connected to a single wire, hence radiating in phase.

is somehow different in sensitivity and phase response, which will reduce the performance of a delay and sum beamformer. Such differences between the sources can be regularised if they are coupled via the cabinet mechanical impedance, increasing the robustness of the beamformer.

2.4.1 Robustness of the 4×8 source to source mismatches

In the vertical sense, the sources of the 4×8 array behave a like broadside delay and sum line array, and so its directional performance directly depends on each source radiating with same amplitude and phase. This is a relevant issue when standard drivers are used, which present a slightly different response from one another. In the upper plot of Fig. 2.26 it is shown how the directivity index is altered due to a mismatch in the transfer functions of a of 4 source delay and sum, as for the case of one of the cabinets of a 4×8 array, which is simulated by introducing random uncertainties following a Gaussian distribution in the volume velocities of each individual source. The mismatch of the sources is rated on mean square error (MSE) [120]. In the lower plot of the same figure, it is shown the relative magnitude of the volume velocities of each source, respect to a value of 1, considered to be the case when all sources are identical, as in an ideal source. It is possible to observe how the error distribution between the sources is more important than the actual magnitude of the error.

2.4.2 Mechanical coupling model

The robustness of the vertical delay and sum beamforming of the 4×8 source can be increased, however, if the sources are placed inside a cabinet where they

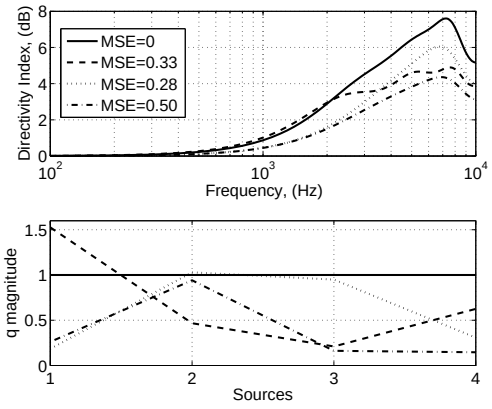


Figure 2.26: Point monopole simulation of the variation on the DI of a 4×8 source with different mean square error distributions on its sources (upper plot). Relative magnitudes of the volume velocities of each source with respect to a magnitude of one, which is considered to be an ideal source (lower plot).

share a common volume. If the mechanical impedance of the cabinet volume is considerably larger than the mechanical impedance of the individual source, the effect of mismatches will be reduced. A simulation to illustrate how this effect is likely to happen can be performed by looking at the impedances of the cabinet compliance and the source of a lumped parameter model.

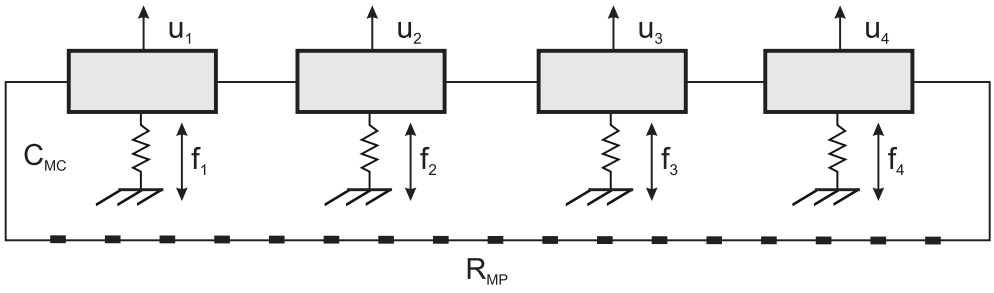


Figure 2.27: Lumped parameter model of the quadruple source, in where each source driven with a velocity u_i produces a force f_i, inside a ported enclosure with a mechanical compliance C_{MC} and a mechanical port resistance R_{MP}.

A lumped parameter representation of a 4×8 cabinet can be observed in Fig. 2.27, in where each single degree of freedom system (SDOF) represents a loudspeaker, which is placed inside a cabinet with a mechanical compliance C_{MC} and a ported resistance R_{MP}. Each source is driven with a velocity u_i, producing a force f_i.

A matrix representation of Fig. 2.27 can be defined as

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} = \begin{bmatrix} Z_{M_1} & 0 & \cdots & 0 \\ 0 & Z_{M_2} & \cdots & \vdots \\ \vdots & \vdots & Z_{M_3} & 0 \\ 0 & \cdots & 0 & Z_{M_4} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} - Z_{MC} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}, \tag{2.33}$$

where Z_{M_i} represents the mechanical impedance of each source and Z_{MC} the mechanical impedance of the cavity at low frequency, where the wavelength is much greater than the cabinet dimension, so that

$$\mathbf{f} = \mathbf{Z}_{M,Tot}\mathbf{u}, \tag{2.34}$$

where the total mechanical impedance matrix, $\mathbf{Z}_{M,Tot}$, is defined as

$$\mathbf{Z}_{M,Tot} = \begin{bmatrix} Z_{M_1} - Z_{MC} & -Z_{MC} & \cdots & -Z_{MC} \\ -Z_{MC} & Z_{M_2} - Z_{MC} & \cdots & \vdots \\ \vdots & \vdots & Z_{M_3} - Z_{MC} & -Z_{MC} \\ -Z_{MC} & \cdots & -Z_{MC} & Z_{M_4} - Z_{MC} \end{bmatrix}. \tag{2.35}$$

Assuming each source has the same Bl product, a vector of forces driving each source, $\mathbf{f}$, is equivalent to

$$\mathbf{f} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} BlI, \tag{2.36}$$

where B represents the magnetic field induction, l the length of the coil and I the intensity that flows through the speaker coil. The matrix of source velocities will be given by

$$\mathbf{u} = \mathbf{Z}_{M,Tot}^{-1}\mathbf{f}, \tag{2.37}$$

which suggest that in a case where Z_{MC} is much greater than that of each individual source, the resultant driving velocities of each source, u_i, will be all identical. In this case coupling via the cavity acts as if adding a regularisation coefficient that minimises the differences between the arrays sources.

2.4.3 Source and cabinet mechanical parameters

In order to compare source and cavity mechanical parameters an analysis is presented, with values estimated from measurements in a practical source and data

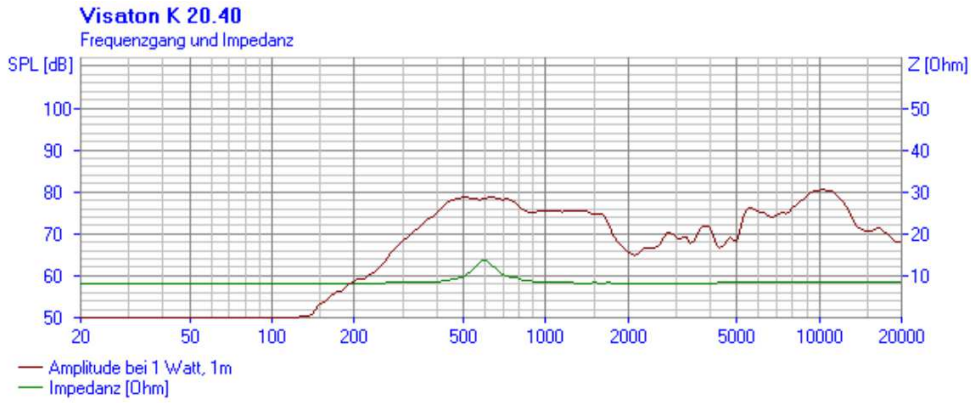


Figure 2.28: Frequency response and electrical impedance of the *Visaton k20.40* sources, which are used on the 4×8 phase-shift cabinets. Data provided by the manufacturer [121]. The complete commercial brochure of the source is shown in Appendix A.

given by the manufacturer. The selected source to build the 4×8 array PS sources is the same as that used in the 1×8 array. The manufacturer of the individual sources, *Visaton*, gives the frequency response and electrical impedance, which is shown in Fig. 2.28. The manufacturer also gives the resonance frequency, f_N, to be 600 Hz, and the total weight of the speaker which is of 0.007 Kg. From Fig. 2.28 the damping ratio of the speaker can be found

$$\zeta = \frac{f_{1,-3dB} - f_{2,-3dB}}{2f_N} \approx 0.07. \tag{2.38}$$

If the mass of the diaphragm, M_D, is estimated to be of 2 grams, the diaphragm stiffness, K_D, can be obtained from the resonance frequency definition

$$f_n = \frac{1}{2\pi}\sqrt{\frac{K_D}{M_D}} \rightarrow K_D = (2\pi f_N)^2 M_D = 2.13 \cdot 10^4 \text{ Nm}^{-1}. \tag{2.39}$$

The speaker damping, R_D, can be obtained by the definition of the damping ratio [122]

$$\zeta = \frac{R_D}{4\pi M_D f_N} \rightarrow R_D = \zeta 4\pi M_D f_N \approx 0.7917 \text{ Kgs}^{-1}. \tag{2.40}$$

The transduction factor of the speaker, $(Bl)^2$, can be obtained from the electrical impedance graph, Z_E. At the resonance peak, the electrical impedance value equals $D + R_E$, where R_E is the speaker nominal resistance equal to 8 Ω, and D is the peak value of the mechanical impedance, which from the graph of Fig. 2.28 has a value of 6 Ω. At the natural frequency, D and $(Bl)^2$ are related by the speaker mechanical resistance, R_M. [123]

$$D = \frac{(Bl)^2}{R_M} \rightarrow (Bl)^2 = DR_M \approx 5.54. \tag{2.41}$$

The inductance of the coil, L_E, can also be obtained from the electrical impedance graph, by knowing that its value at 20 kHz is 8.2 Ω

$$Z_E = \sqrt{R_E^2 + L_E^2} \rightarrow L_E = \approx \sqrt{8.2^2 - 8^2} = 0.014 \text{ mH}. \tag{2.42}$$

The total mechanical impedance of the speaker is then defined as [123]

$$Z_{M,Tot} = R + j\omega M + \frac{1}{j\omega C} + \frac{(Bl)^2}{R_E + j\omega L_E}. \tag{2.43}$$

Two configurations can now be simulated for the model: one considering that the cabinet has a closed back configuration, and another considering that the cabinet has a vented aperture, as in a PS source.

For a closed cabinet speaker, the mechanical impedance that a driver sees is given by a series mass and compliance [82]

$$Z_{MC} = S_D^2 \left(j\omega M_C + \frac{1}{j\omega C_B} \right), \tag{2.44}$$

where S_D^2 is the mechano-acoustical transduction factor, determined by the surface of the source diaphragm, which in this case is S_D =610 mm^2. The quantity M_C represents the acoustic mass in kilograms of the air load in the rear side of the diaphragm, and C_B the acoustic compliance of the cabinet, given by

$$C_B = \frac{V}{\rho c_0^2}. \tag{2.45}$$

For the case of a vented speaker this expression is more complicated, as it is needed to take into account the parallel system formed by the cabinet compliance, C_B, and the rear port resistance, R_P. In this case, Z_{MC} is defined as

$$Z_{MC} = S_D^2 \left(\frac{1}{R_P + j\omega L_P} + j\omega C_B \right)^{-1}. \tag{2.46}$$

A 4×8 cabinet is estimated to have a volume V=1.73·10^{-5} m^3. R_P is estimated to be as 2(350+j75) Pasm^{-1}, as shown in Section 2.3.1. L_P has been chosen to be that of melamine foam, as measured in [124], which is about 600 Pasm^{-1}.

A comparison of the mechanical impedances of the driver and of the two types of enclosures are shown in Fig. 2.29. The estimated impedances of both types of enclosures, closed cabinet and ported cabinet, are higher than the mechanical impedance of the source, which suggests that the individual sources are coupled

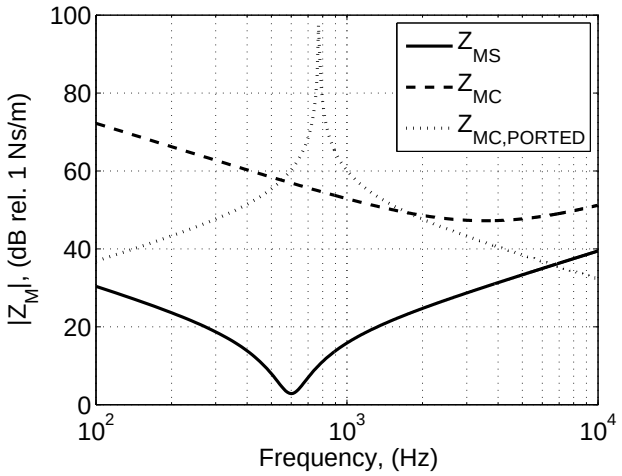


Figure 2.29: Comparison of mechanical impedances: Z_{MS} represents the estimated mechanical impedance of one of the Visaton k20.40 sources, Z_{MC} represents the estimated mechanical impedance of a 4×8 cabinet working in closed back configuration, and $Z_{MC,PORTED}$ represents the estimated mechanical impedance of 4×8 working in PS configuration.

via the cavity, ideally radiating with the same amplitude and phase. The mechanical impedance of the ported cabinet is, however, lower than that of the driver for frequencies above 6.5 kHz, which suggests that the mismatches above this frequency will be less regularised.

2.4.4 Directional performance

The three dimensional DI of a 4×8 source is shown in Fig. 2.30 in comparison with that obtained by a simple source, as that of Fig. 2.23. Due to the vertical beamforming obtained by the 4 sources radiating in phase, a 4×8 source provides a large enhancement in DI with respect to simple source. Below 800 Hz the 1×8 source gives a higher DI. Above 1 kHz the vertical beamforming starts to take place, and the 4×8 source increases in performance, until about 5.5 kHz, where the directivity breaks up. This large reduction in the directional characteristics is attributed to various reasons: the vertical spacing of the sources is 5 cm, which places the spatial aliasing frequency at 6.8 kHz, frequency at which the directivity breaks up, the reduction of directivity at f_N, and also the fact that above 6.5 kHz, the regularisation of the source cabinet is less effective.

Fig. 2.31 shows the horizontal and vertical directivities of a 4×8 source and a simple source. Both sources are very similar in a horizontal sense, however, it can be observed how the vertical directivity pattern given by 4×8 source is much narrower.

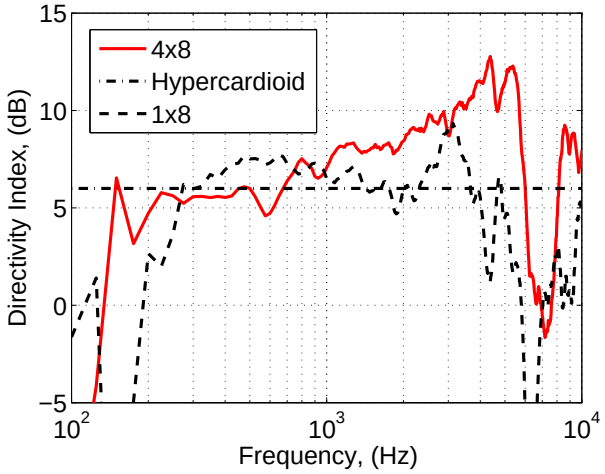
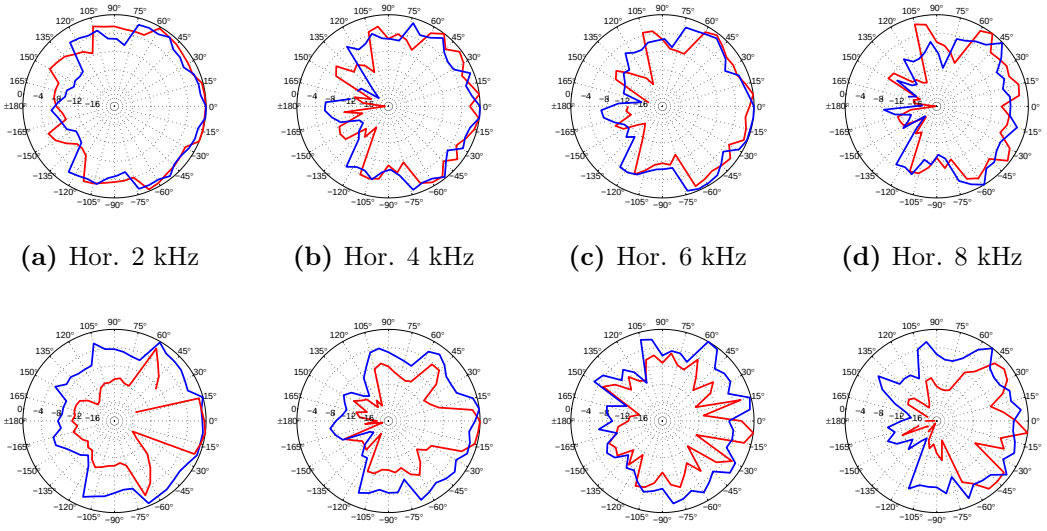


Figure 2.30: The 3D DI, as calculated with Eq. 2.18, of the 4-th source of the 4×8 array compared with that obtained by a single inertance-type source, as that shown in Fig. 2.19.



(a) Hor. 2 kHz **(b)** Hor. 4 kHz **(c)** Hor. 6 kHz **(d)** Hor. 8 kHz

(e) Ver. 2 kHz **(f)** Ver. 4 kHz **(g)** Ver. 6 kHz **(h)** Ver. 8 kHz

Figure 2.31: Measured horizontal, "Hor.", and vertical, "Ver.", directivity patterns of the 4-th source of the 1×8 (blue) and 4×8 (red) arrays at various frequencies.

2.5 Conclusions

This chapter has studied the phase-shift source as a passive directional radiator. A literature review has shown that although such concept was invented many decades ago, there is little evidence of available systems using such arrangement.

A lumped parameter model of the phase-shift source is first considered, which gives a relation between the volume velocity of the moving-coil driver and the rear port. By knowing such relation, the cabinet volume and rear port resistance

needed to obtain a certain directivity pattern can be determined. Such theory has been used to perform initial simulations using point monopole Green functions to model the driver and the rear port of a phase-shift source.

Measurements on a prototype phase-shift source have been presented. Such measurements have shown to obtain a very good directivity until a certain frequency, at which due to the finite size of the cabinet, the radiation is directed towards the back of the source.

The response predictions given by the point monopole model give a result which does not match with that obtained in measurements on the prototype source. This has motivated the development of a finite element model, which has lead to more accurate predictions of the phase-shift prototype source response. This model helped in the design of two sources with cabinets 6 mm and 12 mm long. After these sources were built, they did show a hypercardioid directivity up to 4 kHz, however, between 4 kHz and 6.5 kHz the radiation was directed to the back of the source.

In order to overcome such lack of directivity at the frequency band where the phase-shift effect breaks, a more developed design was introduced, using a mixed acoustic mass-resistance for the rear port. This arrangement makes the source work in closed cabinet configuration at the frequencies where the phase-shift effect fails. In order to develop such loudspeaker, the feasibility was first studied with an analytical model. Measurements with different materials were then made in the laboratory, to find a suitable absorbent for the rear port resistance. Such sources are used in the 1×8 array, which is shown in Chapter 3, and provide a directivity index close to that of a hypercardioid between 500 Hz and 8 kHz.

Although the sources developed for the 1×8 array showed a very good performance, their 3D directivity had to be larger in order to further minimise the power input into the reverberant field. This lead to the construction of the 4×8 sources, which are made up of four phase-shift individual cabinets in a column arrangement. Such sources share a common interior volume, which helps to regularise the mismatches in response by each of the drivers. Due to such arrangement, the radiators are able of obtaining a 3D directivity index about 7 dB higher than a normal phase-shift source. This sources are employed in the 4×8 array, which will be shown in Chapter 3.

Chapter 3

Personal audio array performance

The overall aim of this work is to improve the speech intelligibility of the hard of hearing by the use of a superdirective line array. The array must provide a suitable audio signal in a restricted spatial zone aimed for the hearing impaired. At the same time, it must produce a reduced acoustic response to the regions where normal hearing viewers are sat, so that these are not affected by the amplified audio. It is anticipated that this system will work in addition to the normal TV audio system, which will radiate the low-medium frequencies content.

This chapter describes the design of such an array. Classical array theory is first reviewed, together with practical methods of creating superdirective beamforming in the frequency domain, such the acoustic contrast maximisation and the least squares inverse filters. The behaviour of a superdirective beamformer is analysed by means of simulations, using point source Green functions. After that, the measured performance of the two arrays used in this thesis, the 1×8 array and the 4×8 array, is described. The steering capability of a phase-shift source array is also investigated, in comparison with that of a dual-layer array.

An analysis of the effect of regularisation on the audio quality, and the process used to obtain the time series of the superdirective filters is contained in Appendices B and C. The description of methods for creating superdirective filters in the time domain is described in Section 4. The methodology used in the acoustical measurements contained in this section is presented in Appendix J.

3.1 Source array theory and signal processing

According to the Huygen-Fresnel principle, directional acoustic radiation is obtained when the size of the source is comparable to that of the wavelength it is generating. In this way, a large loudspeaker is naturally more directional than a small source due to its size. However, a source with an equivalent directivity can be made by utilising an array of small sources, all driven together in phase. Systems of smaller sources as directional radiators were introduced in a paper about directional radiation of sound by Wolf and Malter [108], where the authors showed that by combining systems of point sources, highly directional patterns could be obtained. The concept of a line-array was also discussed by Olson [104], where the author described a straight-line source as a large number of point sources of equal strength and phase on a line, separated by equal and very small distances. Olson also stated that the directivity of the line-array increases as its length gets larger, with the condition that the distance between the sources should be smaller than the wavelength of the reproduced signal. Large linear arrays of sources are typically used for PA applications in open air arenas, thus obtaining an efficient radiation that avoids disturbing close neighbourhoods.

Nowadays, the general idea described by Olson [104] can be extended, and although the description for line arrays apply for devices of linear geometry, many other array geometries which are used for creating directional sources can be found in the literature. For example, the speakers of the car audio system have been used as an array to perform personal audio reproduction between front and rear car spaces [58]. The use of spherical source arrays has also been exploited for soundfield reproduction [125, 126] and for sound zoning control [47, 70]. Cylindrical arrays have also been considered for spatial sound reproduction [88].

The use of smaller miniature arrays for domestic purposes inside normal rooms has not been so widely studied, due to the difficulties that reverberation introduces towards creating a directional sound field. However, there are some commercial products aimed for HI-FI and 3D home cinema systems. These are, for example, the OPSODIS sound-bar [127], the Yamaha YSP-1100 [128] or Bose Cinemate 1SR [129].

3.1.1 Superposition of sources

By using superposition of a number of acoustic waves, a directional radiation pattern in a desired direction can be created. Considering a number M acoustic waves, these can be produced by M acoustic radiators. The number of sources of the array and the distance that separates these, determines the frequency and directional response of an array. A schematic representation of a line array is shown in Fig. 3.1.

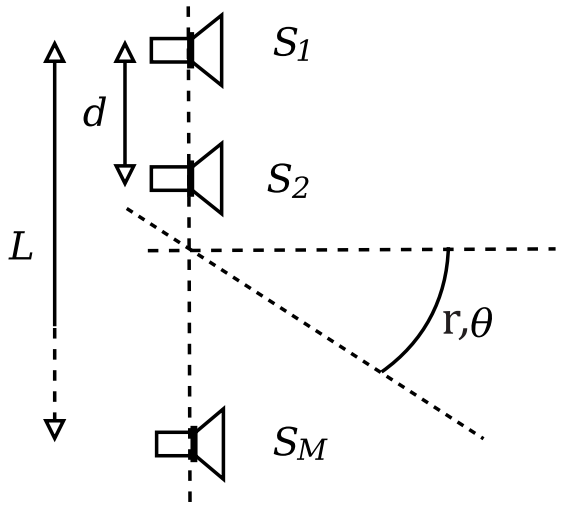


Figure 3.1: Schematic drawing for a M elements line array. Each element is separated a distance d with the array having a total aperture L.

Considering the array of Fig. 3.1, with M monopole sources, each of them spaced a distance d apart. If each source i_{th} has the same amplitude and radiates in phase with each other, the resultant pressure at a field point (r, θ) is given by the summation [130]

$$p(r, \theta, t) = \sum_{i=1}^{M} \frac{A}{r_i} e^{j(\omega t - k r_i)}, \tag{3.1}$$

where A represents the excitation of each of the elements, and r_i is the distance from the source i to a field point at (r, θ). The total pressure at a field point depends on the phase of each source of the array. In the field points where the phase of all the propagating waves is identical, constructive interference will be created and hence the acoustic pressure will be maximised. In the field points where the propagating waves arrive with different phase, destructive interference will be generally produced, with a corresponding decrease in radiated pressure. If all the sources of an array radiate with identical amplitude and phase, a pressure map as this shown in Fig. 3.2 will be created. As can be observed, the directivity is symmetrical with respect to the main axis of the array.

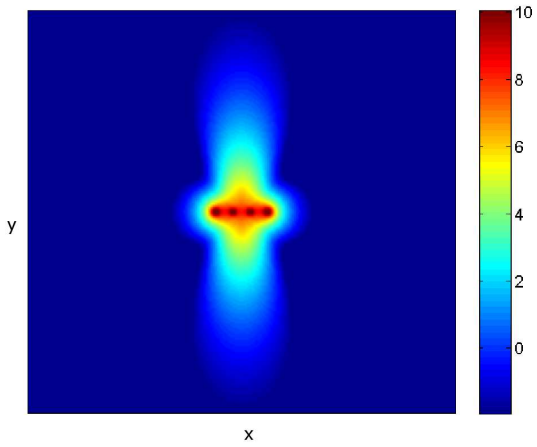


Figure 3.2: Free-field absolute pressure map created by an array of 4 point monopole sources radiating in phase, at $kd = \pi$. The colour map shows the magnitude in dB.

It is often desirable to transmit or receive in various directions without physically rotating the array, which can be accomplished by electronic steering, applying a delay, $i\tau$, into the electric driving signal for the i_{th} element of the array [130]. By mean of this technique, it is possible to select the angle from the array where the M pressures will arrive in phase, so that a maximum of acoustic pressure is propagated at the selected angle. This technique is known as *delay and sum beamforming*. For this case Eq. 3.1 becomes

$$p(r, \theta, t) = \sum_{i=1}^{M} \frac{A}{r_i} e^{j[\omega(t+i\tau)-kr_i]}. \tag{3.2}$$

By using these equations the acoustic response of an idealised delay and sum beamformer can be obtained, which allows to simulate its performance. An example of the soundfield created by an array of 4 point monopole sources in which a delay is applied to each of its sources can be observed in Fig. 3.3.

In Figures 3.2 and 3.3 the response of the array is calculated at a single frequency. The response of an array does, however, change with the frequency of radiation, and it is important, to quantify this variation.

3.1.2 Soundfield control metrics

A measure of the soundfield can be obtained by sampling the amplitude and phase of the acoustic pressure at a set of points. In this thesis, it is important

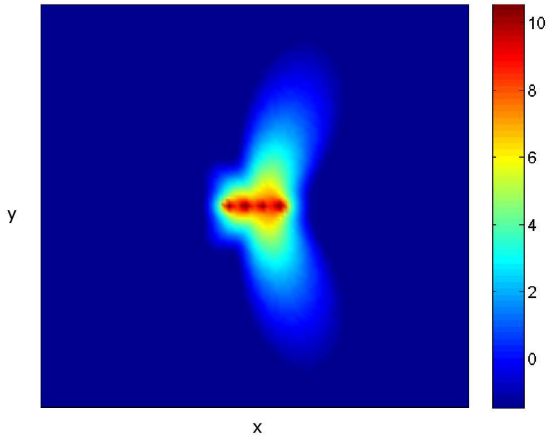


Figure 3.3: Free-field absolute pressure map created by an array of 4 point monopole sources radiating with a phase delay introduced to each of them, at $kd = \pi$. The colour map shows the magnitude in dB.

to know how the array performs in creating a high acoustic pressure zone whilst it reduces the pressure radiation to another zone. To this end, it is convenient to use the notation presented by Choi and Kim [46], where the performance of the radiator is rated in terms of the contrast of mean square pressure that exists between two spatial zones; a zone where the acoustic pressure is to be maximised, known as *bright zone* and a zone where the acoustic pressure is to be minimised, the *dark zone*. A practical control geometry used along this thesis is depicted in Fig. 3.4. A line array of 8 sources is surrounded by a circular microphone array at 2 m from the radiator. The span between each microphone is of 7.5°, which gives a total of 48 microphone positions. The bright zone is formed by three microphones which extend between 0°and 15°to the right hand side of the array. The dark zone is constituted by the rest of the microphones.

The acoustic contrast defines the ratio of mean square pressure between two subcontrol regions inside a control region. In the case of the control geometry shown in Fig. 3.4, the acoustic contrast is the ratio of the mean square pressures of the three microphones of the bright zone divided by the ratio of the mean square pressures of the 45 microphones of the dark zone. At a single radiation frequency, the pressures of both control zones can be rearranged into two vectors

$$\mathbf{p}_B = \left[p_{B_1} \cdots p_{N_B}\right]^T, \ \mathbf{p}_D = \left[p_{D_1} \cdots p_{N_D}\right]^T, \tag{3.3}$$

where N_B is the total number of microphones in the bright zone and N_D is the total number of microphones in the dark zone. The acoustic contrast can then

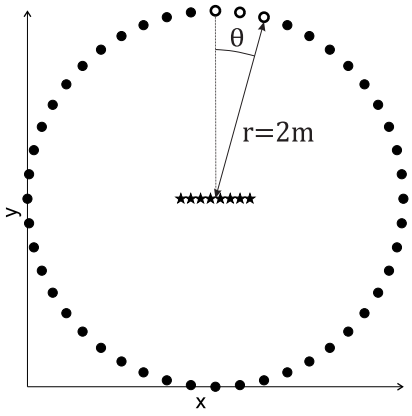


Figure 3.4: Control zone used for simulating and measuring the performance of the arrays presented in this thesis. The open circles represent the microphones defining the bright zone, the black dots represent the dark zone microphones, and the black stars represent the eight sources of the array (scaled).

be written as

$$C = \frac{N_D}{N_B} \frac{\mathbf{p}_B^H \mathbf{p}_B}{\mathbf{p}_D^H \mathbf{p}_D}, \tag{3.4}$$

where the ratio N_D/N_B is included to make the contrast independent of the number of control points and H represents the Hermitian transpose. The contrast can be calculated using this equation if the pressure is known at each of the control points of both zones. The acoustic contrast is normally plotted in dB relative to the acoustic contrast that a single monopole can create in each control geometry.

In order to calculate the acoustic pressure created in each control zone by an array of M sources, we need the individual transfer impedances from each source to each control point. Considering M control sources, two control matrices are created. The matrix correspondent to the bright zone is defined as $\mathbf{Z}_B$, with a size $(N_B \times M)$

$$\mathbf{Z}_B = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1M} \\ Z_{21} & Z_{22} & \cdots & Z_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{N_B 1} & Z_{N_B 2} & \cdots & Z_{N_B M}. \end{bmatrix} \tag{3.5}$$

Similarly, a matrix with a size $N_D \times M$ matrix is created for the dark zone, $\mathbf{Z}_D$

$$\mathbf{Z}_D = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1M} \\ Z_{21} & Z_{22} & \cdots & Z_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{N_D 1} & Z_{N_D 2} & \cdots & Z_{N_D M} \end{bmatrix}. \tag{3.6}$$

The pressure vectors for bright and dark zone can now be defined by multiplying the transfer impedances matrices by the contribution of all the sources of the array, which is contained in a vector of volume velocities $\mathbf{q}$. These pressure vectors are defined as $\mathbf{p}_B$ and $\mathbf{p}_D$, given by

$$\mathbf{p}_B = \mathbf{Z}_B\mathbf{q}, \qquad \mathbf{p}_D = \mathbf{Z}_D\mathbf{q}, \tag{3.7}$$

with the vector of volume velocities defined as

$$\mathbf{q} = \Big[q_1 \cdots q_M\Big]^T. \tag{3.8}$$

The acoustic contrast now can be related to the transfer impedances of the source as

$$C = \frac{N_D}{N_B}\frac{\mathbf{p}_B^H\mathbf{p}_B}{\mathbf{p}_D^H\mathbf{p}_D} = \frac{N_D}{N_B}\frac{\mathbf{q}^H\mathbf{Z}_B^H\mathbf{Z}_B\mathbf{q}}{\mathbf{q}^H\mathbf{Z}_D^H\mathbf{Z}_D\mathbf{q}}. \tag{3.9}$$

The acoustic contrast is used as the main metric of directional performance in this thesis, as it is desired to control an extended listening zone, rather than a single bright point. Apart from a monitoring metric, the acoustic contrast defined in Eq. 3.9 is also used as cost function by the acoustic contrast maximisation (ACM) method.

Apart from the acoustic contrast, the directivity index (DI) is normally used to characterise the directional response of loudspeakers or arrays. In this work, the DI has been used to assess the performance of the phase-shift sources and eventually to test the on axis response of the different arrays which are presented here, as defined in Eqs. 2.18 and 2.20.

3.1.3 Arrays response limitations

The directivity of an array is related to its size, compared to the wavelength of the sound it is generating. The larger the sources compared to the wavelength of the sound waves, the more directional beam results. The specific transduction method has no impact on the directivity of the produced sound field; the analysis relies only on the dimensions, or *aperture* function of the source, according to the HuygensFresnel principle.

Due to the Huygens-Fresnel principle, the directional response of a line array is frequency dependent, with the performance only being high at a narrow frequency range. At frequencies higher than those at which the array perform at its max-

imum, the directional performance is again decreased, and secondary radiation lobes appear in the radiation pattern. It can be said at this point, that the line arrays have a maximum operating frequency. This limitation of performance at high frequency occurs as an analogous process to digital sampling, in where the maximum frequency able to be converted is given by half of the frequency of the converter. In the case of a discrete line array, this phenomena is due to the finite spacing between the sources of the array. This effect is termed as *spatial aliasing*, and occurs when sound waves are emitted at uniform intervals in space [131], which causes the array to have multiple zones of maximum response. For a plane wave generated in the far field by the array in the direction θ, the anti-aliasing condition is given by

$$f_{AL} \leq \frac{c_0}{d|1+\sin\theta|}, \tag{3.10}$$

where c_0 =343 ms^{-1} is the wave speed of the sound particles in the air and d the inter element spacing. The above relation determines, hence, the high frequency limit of operation of a conventional line array for sound field reproduction as a function of θ [132, 133], i.e., above this frequency the array will have multiple main beams. An example is shown in Fig. 3.5, where the array is radiating at $f = f_{AL}$. It can be observed how the array has now 2 distinguished main beams.

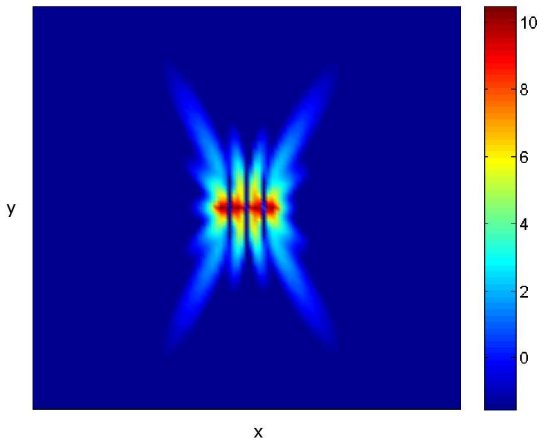


Figure 3.5: Absolute pressure map created by an array of 4 point monopole sources radiating at a frequency above its aliasing frequency, at $kd = 2\pi$. The colour map shows the magnitude in dB.

The low frequency performance of an array is given by its length or *aperture*, since the source is inefficient for wavelengths bigger than its largest dimension [131]. Figures 3.6 and 3.7 show the DI obtained with *broadside* and *endfire* line arrays

[0]Broadside and endfire refer to radiation patterns which are orientated respectively 0° and 90°, using the reference system of Fig. 3.1.

with different number of sources, plotted against the non-dimensional parameter kd.

The aliasing frequencies for broadside and endfire configurations are given by

$$\begin{aligned} Endfire: \quad & f_{AL} = \frac{c}{2d} \rightarrow k = \frac{\omega}{c} = \frac{2\pi f}{c} \rightarrow kd = \pi, \\ Broadside: & f_{AL} = \frac{c}{d} \rightarrow kd = 2\pi. \end{aligned}$$

It is possible to observe in Figures 3.6 and 3.7 how the maximum DI is then obtained at a kd just below the aliasing frequency, as after this frequency, the number of directions of maximum radiation is increased. It is also possible to observe how at low kd numbers the DI is 0, as the wavelength is much greater than the array aperture, and how both endfire and broadside obtain the same peak DI value. When kd tends to the infinite the DI of a line array of M sources approaches $10 \log_{10} M$.

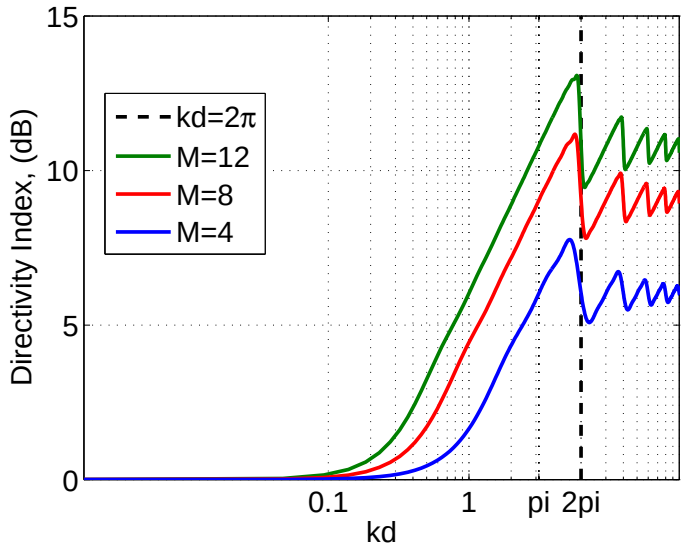


Figure 3.6: Directivity index (DI) of a discrete element delay and sum line array, having equal source strengths, i.e., in a broadside configuration. Simulation performed for arrays with different source numbers, M.

However, the limits presented in these figures only hold for a conventional delay and sum beamformer. The validity of these limits for other types of array control approaches is discussed in Section 3.1.4. Due to the low capacity of a delay and sum beamformer of presenting a constant lobe width over an extended frequency range, signal processing techniques as the ones shown in Section 3.1.4, are used to increase the performance of the array.

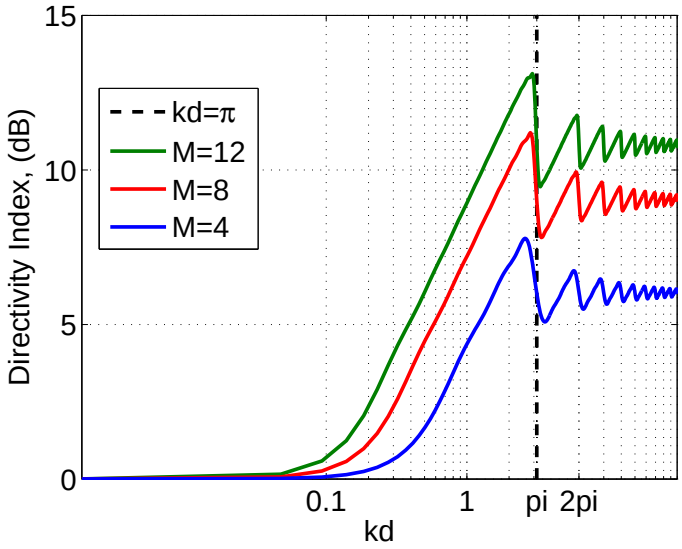


Figure 3.7: Directivity index (DI) of a discrete element delay and sum line array having delayed source strengths, i.e., in an endfire configuration. Simulation performed for arrays with different source numbers, M.

3.1.4 Superdirectivity

The terms *superdirective beamformer, optimal beamformer* and *frequency invariant beamformer* refer to techniques that extend the performance of a conventional delay and sum array. These techniques are based in previous knowledge of the transfer impedances from the individual sources of the array to each field sampling point. Once the transfer impedances are known, a quadratic optimisation, constrained by a desired soundfield is performed, where the optimal solution is given by the radiation pattern which is closer to the desired soundfield.

The use of superdirectivity techniques is popular in the field of sensor arrays [45], and, using acoustic reciprocity, this knowledge can be exported to the field of loudspeakers [130]; The term "super" comes after a paper wherein the design of a high directional microphone array robust to errors was presented by Cox *et al.* [44]. In this case the optimization was performed using Lagrange multipliers, as in the acoustic contrast maximisation [46] method. Optimal beamforming is also applied in radiofrequency arrays [45]. In this case, the signal processing is very similar as that found in acoustical arrays, as there are clear analogies between electromagnetic and acoustic waves [134].

In the following formulation, three superdirective approaches are presented; the least squares inverse filters (LSIF) [135, 50], the acoustic contrast maximisation (ACM) [46] and a third technique that combines this previous two, developed

by Chang and Jacobsen [67], which is known as mixed control or combined solution approach. The main differences between both, least squares inverse filters and acoustic contrast maximisation techniques, is that the least squares method allows to control the phase of the acoustic pressure waves that arrive at each microphone, whilst that acoustic contrast maximisation just maximises the ratio of mean acoustic pressure between the two zones, without taking into account the phase information.

Least Squares Inverse Solution

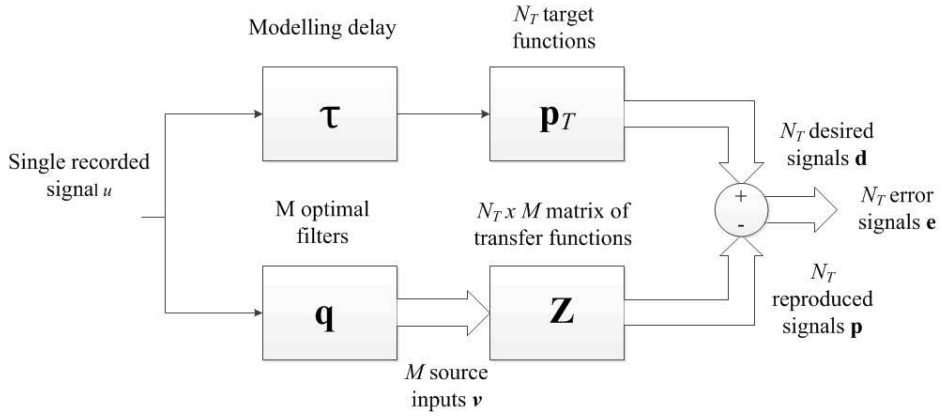


Figure 3.8: Block diagram of the reproduction system used in the least squares inverse filters (LSIF). Reproduced from [50].

A block diagram of the problem used to obtain LSIF is shown in Fig. 3.8, where a system formed by M sources and N_T control points is presented. The system is fed by an input signal, u, which represents the audio input to the system at a single frequency, a factor that scales the amplitude of the vector of volume velocities, $\mathbf{q}$. The least squares formulation requires the selection of a target pressure, which provides a desired value of the amplitude and phase of the sound field at each control point. In the least squares formulation both zones control matrices are rearranged to form a new one, $\mathbf{Z}$, with dimensions $N_T \times M$, of the form

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_B \\ \mathbf{Z}_D \end{bmatrix}, \tag{3.11}$$

where $N_T = N_B + N_D$ is the total number of control points. The vector of pressures at the N_T control points due to the array is then defined as

$$\mathbf{p} = \mathbf{Z}\mathbf{q}. \tag{3.12}$$

An error vector, $\mathbf{e}$, is also introduced, which determines the difference between

the physical acoustic field and the desired, target, acoustical field, $\mathbf{p}_T$, written as

$$\mathbf{e} = \mathbf{p}_T - \mathbf{p}. \tag{3.13}$$

The least squares solution is the distribution of array source strengths that achieves an acoustic field closest to $\mathbf{p}_T$. In order to minimize the mean error, a cost function, J, is defined as

$$J = \mathbf{e}^H\mathbf{e} = (\mathbf{p}_T - \mathbf{p})^H(\mathbf{p}_T - \mathbf{p}), \tag{3.14}$$

which if expanded gives the Hermitian quadratic form of the cost function

$$J = \mathbf{p}_T^H\mathbf{p}_T - \mathbf{Z}^H\mathbf{q}^H\mathbf{p_T} - \mathbf{p}_T^H\mathbf{Z}\mathbf{q} + \mathbf{q}^H\mathbf{Z}^H\mathbf{Z}\mathbf{q}. \tag{3.15}$$

Since the Hermitian quadratic form of the cost function is real and grater than zero for any $\mathbf{q}$, the matrix $\mathbf{Z}^H\mathbf{Z}$ must be positive definite. In this case, the cost function describes a bowl shaped surface in terms of the real and imaginary components of $\mathbf{q}$, with a unique global minimum. The minimal solution is got by equating the partial derivative of the cost function respect to $\mathbf{q}$, to zero [105].

$$\frac{\partial J}{\partial \mathbf{q}} = 0. \tag{3.16}$$

Depending on the dimensions of the plant matrix, the optimal solution is given by a different equation. In the case of a fully determined case, where $N_T = M$ and hence the matrix $\mathbf{Z}^H\mathbf{Z}$ is square, the solution is given by

$$\mathbf{q} = \mathbf{Z}^{-1}\mathbf{p}_T. \tag{3.17}$$

If the number of sources is smaller than the number of control points, an overdetermined system is formed ($N_T > M$)

$$\mathbf{q} = (\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H\mathbf{p}_T. \tag{3.18}$$

Once the vector of optimal volume velocities $\mathbf{q}$ is obtained, the acoustic contrast defined in Eq. 3.9 is used for rating the array performance along frequency. For a given control geometry, the personal audio performance of an array optimised using LSIF depends strongly on the definition of the vector of target pressures, $\mathbf{p}_T$.

Contrast Maximisation Solution

The acoustic contrast maximization (ACM) was presented by Choi and Kim [46]. The authors introduced an analytical method for obtaining the optimal solution of a cost function, which maximised the mean square pressure contrast between two control regions. The acoustical potential energy determines the perceived sound. As the acoustical potential energy is proportional to the acoustic pressure squared, the ACM algorithm maximizes the difference in potential energy between two defined control zones. These are, a bright zone with maximum potential energy, and a dark zone with minimal potential energy, as described in Section 3.1.2.

Being the control points placed along a desired control volume, these are divided into bright and dark control points, as shown in Section 3.1.2, leading to two vectors of pressures respectively, $\mathbf{p}_B$ and $\mathbf{p}_D$. The maximum value for the cost function C, as defined in Eq. 3.9, is obtained by solving a constrained optimization where $\mathbf{p}_D^H\mathbf{p}_D$ is minimized with the condition that $\mathbf{p}_B^H\mathbf{p}_B$ is held constant to a value c. By the use of Lagrange multipliers the function to be minimized with respect to $\mathbf{q}$ and the Lagrange multiplier λ is given by

$$J = \mathbf{q}^H\mathbf{Z}_D^H\mathbf{Z}_D\mathbf{q} - \lambda(\mathbf{q}^H\mathbf{Z}_B^H\mathbf{Z}_B\mathbf{q} - c). \tag{3.19}$$

By setting the differential of this function respect to the complex and real parts of $\mathbf{q}$ to 0

$$\frac{\partial J}{\partial \mathbf{q}} = 0, \tag{3.20}$$

leading to

$$\mathbf{Z}_D^H\mathbf{Z}_D\mathbf{q} - \lambda\mathbf{Z}_B^H\mathbf{Z}_B\mathbf{q} = 0 \tag{3.21}$$

which rearranged gives,

$$\mathbf{q} = \lambda[\mathbf{Z}_D^H\mathbf{Z}_D]^{-1}\mathbf{Z}_B^H\mathbf{Z}_B\mathbf{q}. \tag{3.22}$$

The optimal solution which maximises the directional response is given when $\mathbf{q}$ is proportional to an eigenvector correspondent to the largest eigenvalue of the matrix $[\mathbf{Z}_D^H\mathbf{Z}_D]^{-1}\mathbf{Z}_B^H\mathbf{Z}_B$, as stated by Choi and Kim [46].

In order for the bright zone pressure to have a value c the differential of J with respect to λ is set to 0 so that

$$\frac{\partial J}{\partial \lambda} = 0, \tag{3.23}$$

leading to

$$\mathbf{Z}_B^H\mathbf{Z}_B\mathbf{q} = c, \tag{3.24}$$

which can be performed after obtaining the optimal vector of source strengths $\mathbf{q}$.

If the bright zone is shrunk down to a single point, the pressure in the bright zone becomes a scalar, equal to

$$p_B = \mathbf{z}_B \mathbf{q}, \tag{3.25}$$

where $\mathbf{z}_B$ is now formed by a $1 \times M$ vector. If $\mathbf{p}_D^H \mathbf{p}_D$ is minimized with p_B held constant with a value c, in order to find the optimal solution, the Lagrange multiplier is presented

$$J = \mathbf{q}^H \mathbf{Z}_D^H \mathbf{Z}_D \mathbf{q} - \lambda(\mathbf{z}_B \mathbf{q} - c). \tag{3.26}$$

By partially differentiating J respect to $\mathbf{q}$ it is obtained

$$2\mathbf{q}^H \mathbf{Z}_D^H \mathbf{Z}_D \mathbf{q} = \lambda \mathbf{z}_B^H, \tag{3.27}$$

leading to a direct analytic solution where the vector of optimal volume velocities is given by

$$\mathbf{q} = \frac{\lambda}{2} [\mathbf{Z}_D^H \mathbf{Z}_D]^{-1} \mathbf{z}_B^H, \tag{3.28}$$

having the Lagrange multiplier dimensions of pressure. Similarly as performed in Equations 3.23 and 3.24, by setting the differential of J with respect to λ in Eq. 3.26 it is obtained that $\mathbf{z}_B \mathbf{q} = c$. However, it is only the relative values of the terms in the vector $\mathbf{q}$ which are important for the directional response [52].

Mixed control solution

A solution that combines the ACM and LSIF control techniques has been recently published by Chang and Jacobsen [67]. This solution allows to switch between a pure contrast maximisation technique, in where the maximum contrast of mean square pressure is obtained, and a solution in where the vector of target pressures in the bright zone, $\mathbf{p}_T$, is matched. In this thesis, this solution is referred as mixed control (MC) or combined solution, and it is obtained by introducing the cost function

$$J = \xi \mathbf{P}_D^H \mathbf{P}_D + (1 - \xi)(\mathbf{P}_B - \hat{\mathbf{P}}_B)^H (\mathbf{P}_B - \hat{\mathbf{P}}_B), \tag{3.29}$$

where $\xi (0 \leq \xi < 1)$ is a weighting factor that determines the balance between the potential energy in the quiet zone and the mean square error in the listening zone and $\hat{\mathbf{P}}_B$ refers to the desired vector of pressures in the bright zone. Substitution

of the vectors of pressures defined in Eq. 3.7 into Eq. 4.10 leads to

$$J = \mathbf{q}^H \left[\xi \mathbf{Z}_D^H \mathbf{Z}_D + (1-\xi)\mathbf{Z}_B^H \mathbf{Z}_B\right] \mathbf{q} + (1-\xi)(\mathbf{P}_B^H \mathbf{P}_B - \mathbf{P}_B^H \mathbf{Z}_B \mathbf{q} - \mathbf{q}^H \mathbf{Z}_B^H \hat{\mathbf{P}}_H). \tag{3.30}$$

The matrix in the square brackets is positive definite, therefore J has an optimal solution at a point given by [105]. In this case $\mathbf{Z}_B^H \hat{\mathbf{P}}_H = \mathbf{Z}_T^H \mathbf{p}_T$, so that

$$\frac{\partial J}{\partial \mathbf{q}} = 2\left[\xi \mathbf{Z}_D^H \mathbf{Z}_D + (1-\xi)\mathbf{Z}_B^H \mathbf{Z}_B\right] \mathbf{q} + 2(1-\xi)(-\mathbf{Z}_T^H \mathbf{p}_T). \tag{3.31}$$

The optimal solution is hence obtained as

$$\mathbf{q} = \left[\xi \mathbf{Z}_D^H \mathbf{Z}_D + (1-\xi)\mathbf{Z}_B^H \mathbf{Z}_B\right]^{-1} (1-\xi)\mathbf{Z}_T^H \mathbf{p}_T. \tag{3.32}$$

The directional characteristics of the solution obtained vary depending on the value of ξ. As it approaches to 1, the solution tends to minimise the acoustic potential energy on the dark zone, so that the acoustic contrast is maximised. As it approaches to 0, the solution tends to minimise the error in the listening zone. For ξ=0.5 the solution is equal to that of the LSIF given by Eq. 3.18.

3.1.5 Control effort and ill-conditioning issues

Introduction

Common to all the superdirective techniques presented here, there are two problems that increase proportional to the number of sources used [94]; these are the great amount of energy at low frequency needed to obtain superdirectivity, and the corresponding ill-conditioning problem. Both problems are caused by the low efficiency of the array radiation at low frequency [136].

As the condition number with respect to the inversion of the transfer impedances matrix grows with the number of sources, the power needed for obtaining superdirectivity at low frequencies also grows proportional with the number of sources of the array [52, 94, 53]. Mathematically, when the matrix $\mathbf{Z}$ is full rank, the inverse $(\mathbf{Z}^H\mathbf{Z})$ exists. However, this result might not be stable since the matrix $\mathbf{Z}$ can present high order singular values close to zero. In the transfer impedances matrix of an array, this is caused due to the similarity of each source transfer impedances at low frequency. Therefore, although a vector of optimal digital filters that minimizes the cost functions J or C exists, the norm of this solution is very large, giving even so small singular values that make the solution not possible using numerical methods [65]. Furthermore, ill-conditioning makes the

solution very sensitive to small changes in the input variables [137, 91] (in this case the input variables are the transfer impedances). As the norm of the solution is proportional to the required power, the array needs filters with a very large amplitude at low frequency.

To avoid these effects, a constraint can be added into the cost function. This action is performed by adding a positive number to the matrix ($\mathbf{Z}^H\mathbf{Z}$) which increases the amplitude of its singular values, making the inversion of the matrix easier. This technique was first introduced by Tychonoff [138], having taken its name and being known as "Tychonoff's regularization".

The regularisation factor is denoted as β. This parameter is known in the context of active control as coefficient-weighting factor [105]. In the context of the line arrays, Boone *et al.* [64] refer to this term as stability factor; however, in this work, it is going to be defined as the control effort weighting factor, as termed by Elliott *et al.* [52]. As shown in the following sections, the use of regularization is based on a modification of the cost functions defined in Sections. 3.1.4 and 3.1.4. By selecting a different value for β at each frequency, the amount of regularisation can be selected so that some frequencies are more penalised than others.

When $\beta\mathbf{I}$ is added to the matrix $\mathbf{Z}^H\mathbf{Z}$, the low amplitude singular values of the matrix get larger, making the inversion easier. Furthermore, the difference between the largest and the smallest singular values is reduced, so that the norm of the solution is reduced and consequently the amount of electrical energy used in the filters. In the case that large dips exist in the magnitude of the frequency response of the original plant response, the optimal filters try to correct these by creating large boosts at those frequencies where the dips occur. In this case, regularization can be used to limit the existence of those large boosts of the optimal filters and hence limit the power output from them.

Regularisation for the least squares inverse filters (LSIF)

In order to introduce regularisation, the cost function defined in Eq. 3.14 is modified to give

$$J = \mathbf{e}^H\mathbf{e} + \beta\mathbf{q}^H\mathbf{q}, \tag{3.33}$$

where $\beta\mathbf{q}^H\mathbf{q}$ is the previously defined effort weighting factor, which is a measure proportional to the total input power to all sources. The parameter β is a positive real number that determines the weighting of the effort factor [50]. The cost

function to minimise is given in this case by

$$J = \mathbf{e}^H\mathbf{e} + \beta\mathbf{q}^H\mathbf{q} = (\mathbf{p}_T - \mathbf{p})^H(\mathbf{p}_T - \mathbf{p}) + \beta\mathbf{q}^H\mathbf{q}. \tag{3.34}$$

Similar to Eq. 3.14, the optimal vector $\mathbf{q}$ that minimizes the cost function in this case is written as

$$\mathbf{q} = (\mathbf{Z}^H\mathbf{Z} + \beta\mathbf{I})^{-1}\mathbf{Z}^H\mathbf{p}_T, \tag{3.35}$$

where depending on the value of β, the optimal solution will tend to minimise the squared error or the amplitude of the control filters.

Regularisation for the acoustic contrast maximization

For the case of the acoustic contrast maximisation, the procedure is similar to that used in the least squares optimal solution, wherein the control effort weighting factor is added to the original cost function. The maximum value for the cost function C defined in Eq. 4.20 is optimized by minimizing $\mathbf{q}^H\mathbf{Z}_D^H\mathbf{Z}_D\mathbf{q} + \beta\mathbf{q}^H\mathbf{q}$ whilst maintaining $\mathbf{q}^H\mathbf{Z}_B^H\mathbf{Z}_B\mathbf{q}$ equal to a value c. In order to obtain the optimal solution, the differential of the cost function J respect to $\mathbf{q}$ is set to zero

$$\frac{\partial J}{\partial \mathbf{q}} = 0, \tag{3.36}$$

giving finally

$$\left[\mathbf{Z}_D^H\mathbf{Z}_D\mathbf{q} + \beta\mathbf{I} - \lambda\mathbf{Z}_B^H\mathbf{Z}_B\right]\mathbf{q} = 0. \tag{3.37}$$

Rearranging the previous equation it leads to a final form

$$\mathbf{q} = \lambda[\mathbf{Z}_D^H\mathbf{Z}_D + \beta I]^{-1}\mathbf{Z}_B^H\mathbf{Z}_B\mathbf{q}, \tag{3.38}$$

where the optimal value of volume velocities is given by the eigenvector corresponding to the maximum eigenvalue of the matrix $[\mathbf{Z}_D^H\mathbf{Z}_D + \beta I]^{-1}\mathbf{Z}_B^H\mathbf{Z}_B$ [55]. Note that for the acoustic contrast maximisation formulation the role of β is different that for the least squares inverse filters. As the filters for the acoustic contrast maximisation are given by an eigenvector, they have a unitary norm. Hence, by increasing β in Eq. 3.38, the obtained filters have still unitary norm, but the reproduced pressure in the bright zone is lower. This requires of further equalisation and regularisation until both the effort constrains and the desired pressure in the listening zone are met.

Regularisation for the combined solution

Chang and Jacobsen [67] presented an alternative formulation to perform matrix regularisations, however, Tychonoff's regularisation is used in this thesis, as performed for the ACM and LSIF approaches. In this case the cost function to optimise is

$$J = \xi \mathbf{P}_D^H \mathbf{P}_D + \beta \mathbf{q}^H \mathbf{q} + (1-\xi)(\mathbf{P}_B - \hat{\mathbf{P}}_B)^H (\mathbf{P}_B - \hat{\mathbf{P}}_B). \tag{3.39}$$

If taking the partial differential of J respect to $\mathbf{q}$ and rearranging, it leads to

$$\mathbf{q} = \left[\xi \mathbf{Z}_D^H \mathbf{Z}_D + (1-\xi)\mathbf{Z}_B^H \mathbf{Z}_B + \beta \mathbf{I}\right]^{-1} (1-\xi)\mathbf{Z}_T^H \mathbf{p}_T. \tag{3.40}$$

Controlling the amount of regularisation

In order to have a good knowledge of the amount of superdirectivity being used and the requirements it needs, a metric of the electric power used by the array is introduced, which is termed *normalised array effort* [53, 91]. The normalised array effort is defined as the norm of the optimal set of volume velocities divided by the norm of the volume velocity that a single monopole requires to obtain the same acoustic pressure as that produced by the array in the centre of the bright zone, q_{MON}. The normalized array effort is thus written as

$$AE = \frac{\mathbf{q}^H \mathbf{q}}{|q_{MON}|^2}. \tag{3.41}$$

This quantity is proportional to the electrical power required to drive the array, assuming there are no significant electroacoustic interactions between the transducers. The array effort is generally plotted in dB.

The parameter β can be adjusted iteratively at each frequency so that the array effort is constant or lower than a desired power figure. Other researches control the amount of regularisation proportionally to the value of the biggest singular value of the transfer impedances matrix [65]. The regularisation parameter has also been selected for sound field reconstruction applications so that an optimal balance between performance and energy is obtained, using the so called "L-Curve" method [69]. The array effort is used along this thesis, however, as it allows a more precise control of the magnitude of the electrical power used in the solution.

To illustrate the effects of the use of regularisation to control the array effort, an example is introduced with an endfire array. The array has been optimised to

maximise the acoustic contrast in a control geometry as that shown in Fig. 3.9. This control geometry is formed by a single bright point in endfire position and 258 dark points. The control points are positioned in order to cover an almost equal distribution of area around each control point [139] and hence obtain a good estimate of the radiated sound field. Since the bright zone consists of a single point, the optimal solution for the ACM produces in this case an identical directional responses to that of the LSIF [52], as there are no phase constraints for the LSIF method.

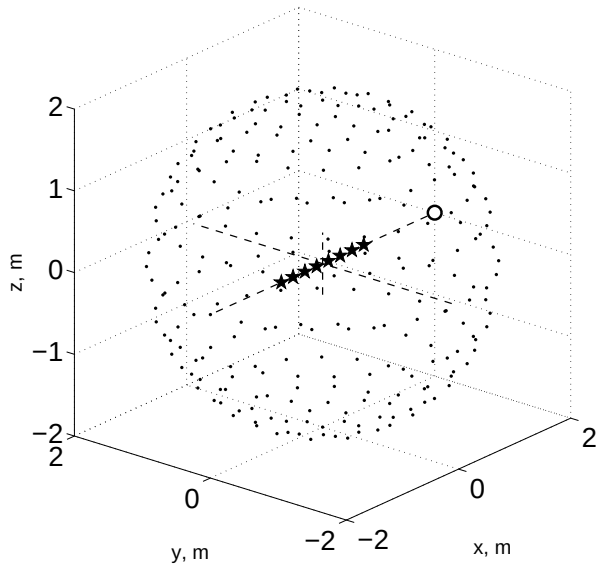


Figure 3.9: Geometric arrangement for an endfire array. Black points represent the microphones in the dark region. The open circle represents the microphone defining the bright zone and the black stars represent the 8 sources of the array. Reproduced from [53].

The acoustic contrast performance obtained with ACM filters is shown in Fig. 3.10, for the case when the array effort is unconstrained (dashed line) and for the case when the regularization parameter β is adjusted at each frequency to limit the normalized effort to 6 dB. A maximum value of 6 dB is used in most of the simulations and measured results presented throughout this thesis, as it is considered to be a good upper limit to prevent damage to the array sources by overdriving. The value of the parameter β needed for limiting the power of the endfire array below 6 dB is shown in Fig. 3.11. The amount of β required grows with frequency until 3 kHz, where it decreases greatly. Above 3 kHz, the array effort is not longer large, as the array starts to be comparable with the radiation wavelength. It can be observed how the lower frequencies are more sensitive to the amount of regularisation used, and how the value of β increases

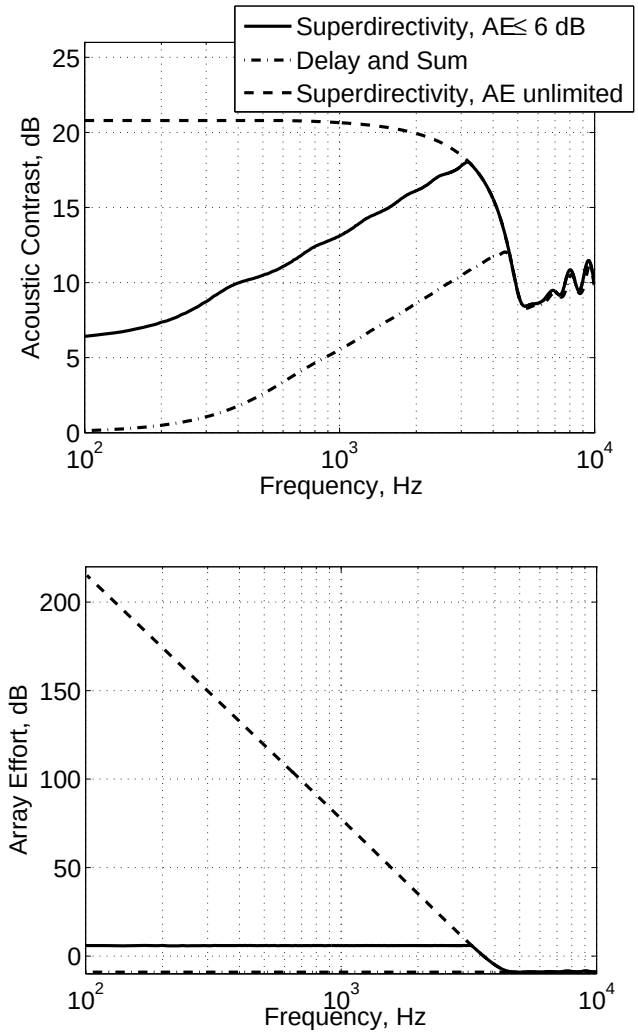


Figure 3.10: Acoustic contrast and array effort produced by an 8 source endfire array controlled via ACM filters obtained without any array effort (AE) limitations and when the AE is limited to 6 dB at each frequency. The response of the array when driven in delay and sum configuration is also shown for comparison. Reproduced from [53].

with frequency until 3 kHz.

It is interesting to note that when the normalized array effort is limited to 6 dB at each frequency, the trend set by the delay and sum beamformer is continued to lower frequencies. There is thus a clear trade-off between low frequency limit and array effort in the regularized design of such superdirective arrays. The effort required for the delay and sum beamformer is only 1/8 of that of a monopole, as each of the 8 sources requires a mean square volume velocity of $(\frac{|q_{MON}|}{8})^2$, so that the level of the normalized effort is about -9 dB.

The amount of superdirectivity used in the array, and hence the electrical energy required, has a direct effect on the audio quality obtained with the array. An

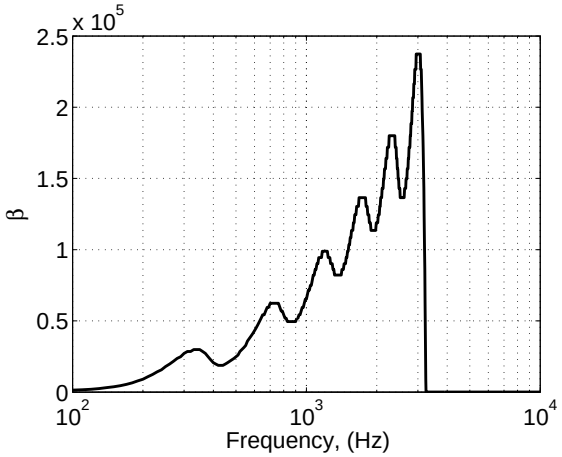


Figure 3.11: Magnitude of the regularisation parameter, β, needed to limit array effort of the endfire array shown in Fig. 3.10 to 6 dB.

analysis of this effect is presented in Appendix B. The appendix shows how the array effort has to be constrained not to be below a certain figure to obtain an acceptable audio quality.

The design of the sources for the two arrays mainly used in this thesis is described in Chapter 2. In the following sections, the practical performance of these arrays will be presented, when driving using the superdirective techniques described above.

3.2 The 1×8 array

Fig. 4.6 shows a close up of the array. This array comprises 8 phase-shift sources (so its name of 1×8) which can be driven to give a highly directive response, while minimising the reverberant excitation of the room. The array is aimed to be used in a broadside configuration or close to broadside configuration, as it is desired to control the soundfield until a high frequency. For the case of this array, the separation between each source is of $d = 3.5$ cm, so that the aliasing frequency is present at 7.8 kHz for a broadside configuration.

Although line arrays have been studied for many years, the approach previously used to reduce the radiation at the rear of the array, and hence reduce reverberant levels for other listeners, has been to position a second set of loudspeakers at the back of the array to cancel the sound in that direction, as for example in [67, 69, 109]. Instead of using an extra array to form a dual layer, the 1×8 array uses phase-shift sources with hypercardioid directivity. The phase-shift sources

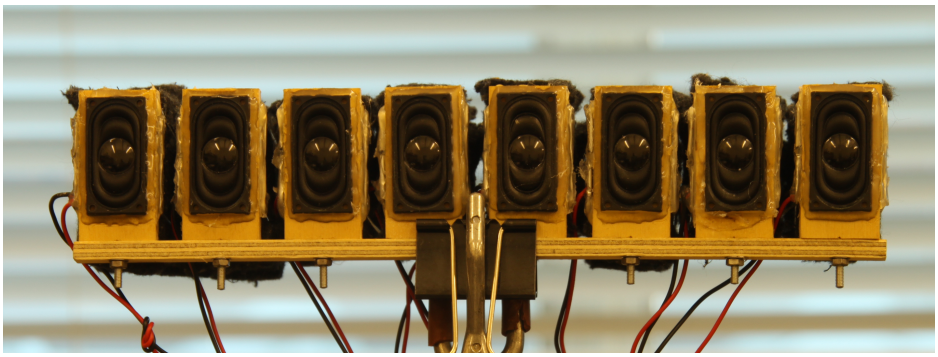


Figure 3.12: Close up of the 1×8 array.

used by the array are explained in detail in Section 2.3.2. The use of phase-shift sources with respect to a dual layer array is in great advantage, as it allows to obtain the same directional characteristics whilst using half the number of individual sources. The lower the number of sources used by an array, the more robust with respect to ill-conditioning and mismatches in the transfer impedances the array is. Furthermore, each source needs a digital filter to form the superdirective beamforming, what requires to use a DSP with many independent channels. From this point of view, an array which uses M sources instead of $2M$ sources to obtain the same acoustic performance is in clear advantage.

A complete analysis of the performance of a phase-shift sources array with respect to a dual layer array is presented in Appendix E, using point source transfer functions, and in Section 3.3, using the measured transfer functions of the 1×8 array and of a dual layer array.

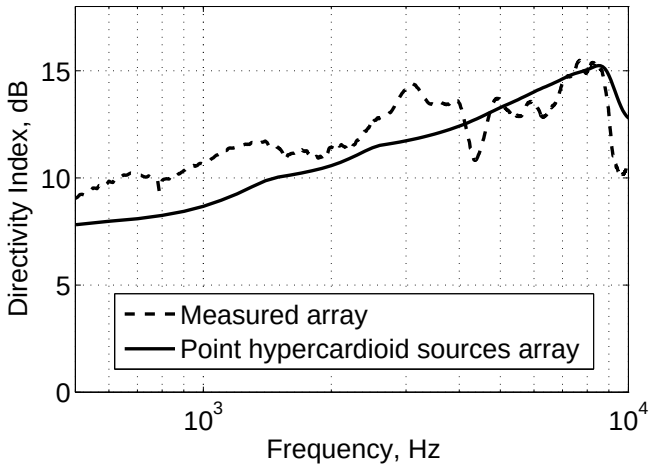


Figure 3.13: DI presented by the 8 source array using phase-shift sources, measured in the free-field and calculated from simulations using point hypercardioid sources. The array effort is limited to be below 6 dB.

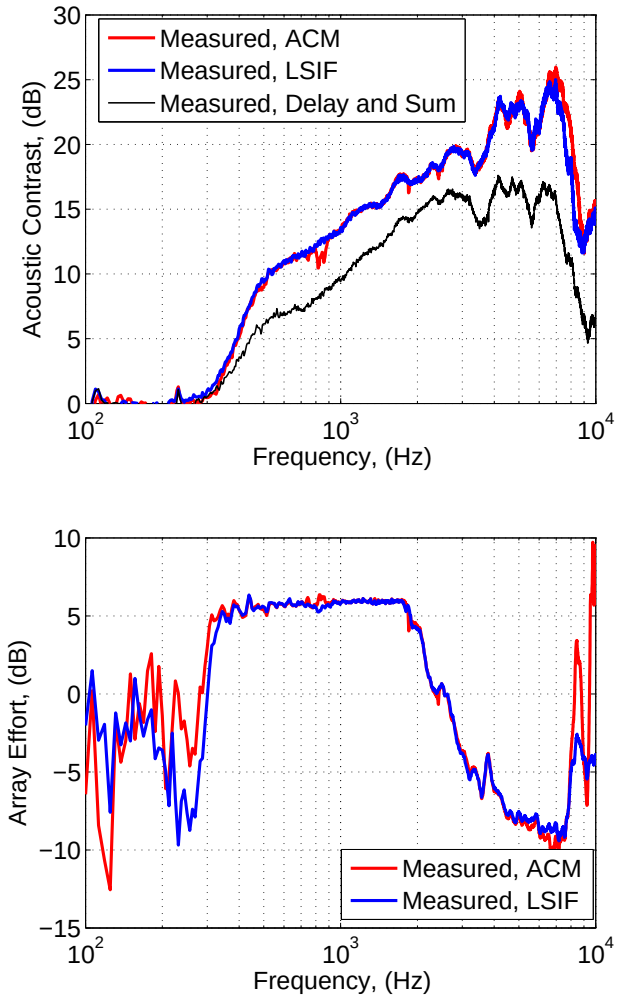


Figure 3.14: Measured free-field acoustic contrast performance of the 8 source array using phase-shift sources. Also shown for comparison is the measured acoustic contrast performance of the array in delay and sum configuration. Reproduced from [53].

In order to characterise the array and being able to apply the formulation presented in Section 3.1.4, the transfer impedances from each source of the array to each microphone of the control zone illustrated in Fig. 3.4 have been measured. A listening distance of 2 m has been chosen to create the filters for the array, as this is considered a typical distance for watching TV and, furthermore, it is far away compared with the size of the array. The free-field directional performance obtained with the array is shown, in terms of DI performance, in Fig. 3.13, compared to that obtained with an 8 point hypercardioid sources array, with the AE limited to 6 dB for both arrays. In this case, the array filters are created so that a main beam is directed in a broadside configuration, i.e., the bright zone is a single microphone placed some distance away in a perpendicular line that passes through the centre of the array. The filters have been obtained initially with $I = 2048$ and have been then truncated to give impulse responses of $I = 1024$,

as shown in Appendix C.

Sometimes the bright zone is required to be broader than a single point, as in the case of the original application of the array as an aid for the hard of hearing. To this end, a control geometry as shown in Fig. 3.4, with a bright zone subtending an angle of 15° from broadside has been used to create the filters of the array and to assess the directional performance, so that it provides an area wide enough for a hearing impaired listener. The measured real-time performance that the array provides in such a control zone in a free-field environment is shown in Fig. 3.14, using the ACM and LSIF control algorithms. The response of the array working as a delay and sum beamformer is also shown for comparison. As observed, the acoustic contrast offered by LSIF and ACM is very similar along the whole frequency band. The ACM is able, however, of obtaining a slightly higher performance between 6.5 kH and 9 kHz, at the expense of using a larger array effort.

The impulse of the filter corresponding to the 4th source of the array used to obtain these results are shown in Fig. 3.15. The impulse response has been obtained by inverse Fourier transformation of the frequency response also shown in the same image, for frequency responses calculated with N=4096 frequency samples. The impulse responses for each filter have been then truncated to a shorter length of 1024 samples. It can be observed how the impulse response has a small non-minimum phase part, and how the magnitude of the frequency response is similar along frequency to the array effort. More information about the audio performance of the filters with respect to the use of regularisation is presented in Appendix B and an study of the directional performance for filters of different lengths is shown in Appendix C.

The measured directivities of the array at various frequencies are also shown in Fig. 3.16, where it can be observed the great performance of the individual phase-shift sources in reducing the back radiation of the array.

By measuring the individual transfer functions from each source of the array to each control point, it has been shown that a large directionality has been obtained. The measuring of these quantities is, however, a time consuming and delicate process, which needs of an anechoic environment. In order to make the filter creation process easier, it is possible to model the transfer function of each source based on a point source Green function or in the frequency response of a single source. Nevertheless, this process does not obtain the same free-field di-

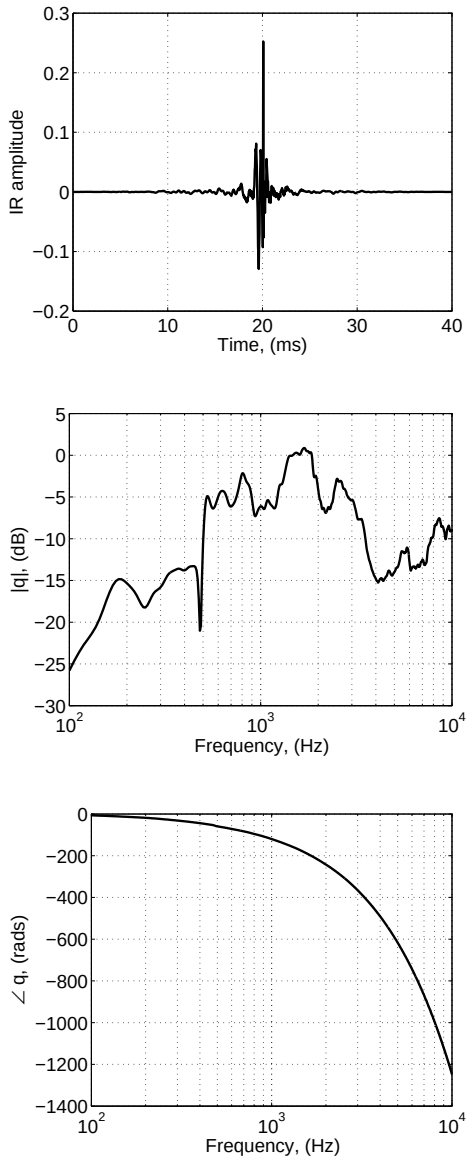
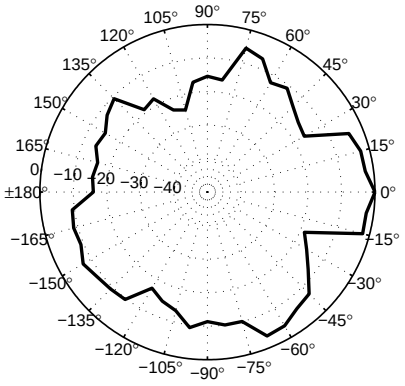


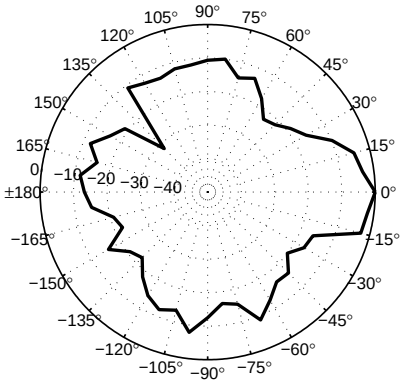
Figure 3.15: Impulse and frequency responses of the filter for the 4th source of the 1×8 array, with a length of 1024 samples.

rectionality as by measuring the individual transfer functions of each source, but as will be shown in Section 5.4, it leads to the same directional performance in a normal room. An analysis of how the array filters can be created using analytical transfer functions is presented in Appendix D.

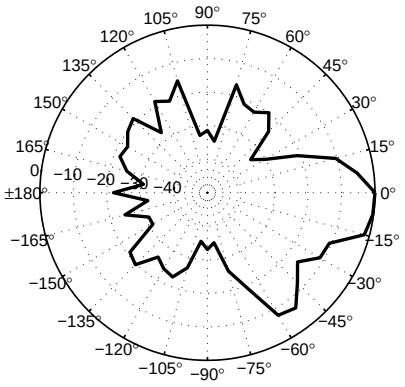
The performance of this array inside an anechoic chamber shows a great capacity to be used as a hearing aid, as an amplification equal to the acoustic contrast can be applied at each frequency in the bright zone without annoying the normal hearing listeners of the dark zone. However, when the array is placed inside a normal, reverberant, environment, its performance is significantly reduced, as will be discussed in Section 5.3. This has motivated the design of an array which is



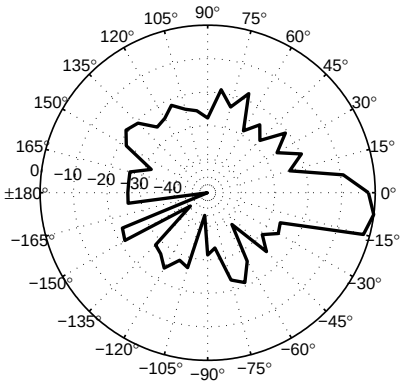
(a) 500 Hz



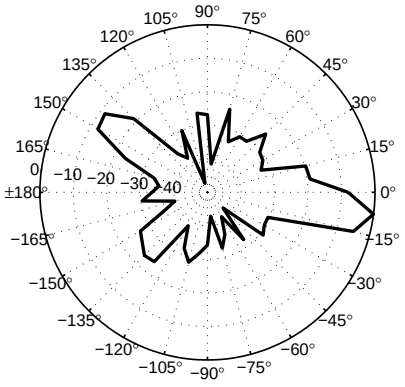
(b) 1 kHz



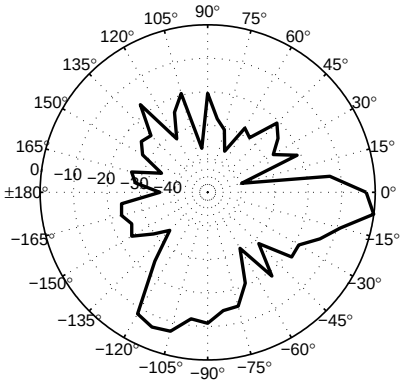
(c) 2 kHz



(d) 4 kHz



(e) 6 kHz



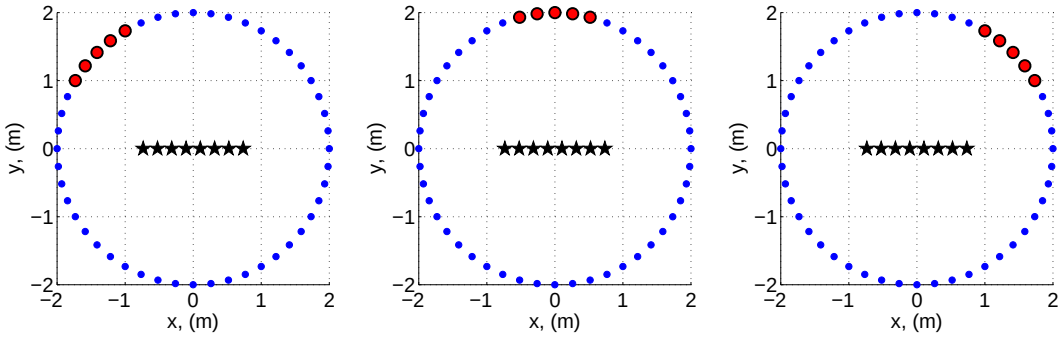
(f) 8 kHz

Figure 3.16: Measured directivity patterns of the 1×8 array at different frequencies. Results calculated with acoustic contrast maximisation filters with the AE limited to 6 dB.

more directional in a 3D sense; the 4×8 array presented in Section 3.4.

3.3 Steering performance of the array and comparison with a dual layer array

The ability to steer the array to different directions is now illustrated, using the geometry shown in Fig. 3.17. The opportunity is also taken to compare the results from a phase-shift array, the 1×8, with a dual layer array of 8 frontal and 8 rear loudspeakers, as shown in Fig. 3.18. An initial simulation using point source Green functions for hypercardioid and double monopole arrays is presented in Appendix E. This suggested that the performance of the two types of arrays was similar for angles that were surprisingly large compared with the broadside direction in which the hypercardioids were pointing.



(a) Left bright zone **(b)** Central bright zone **(c)** Right bright zone

Figure 3.17: Control geometries for creating three different sound programs in a 2D horizontal plane. Each bright zone is extended over 30°. The left bright zone covers the space comprehended between -60° and -30°, the central dark zone extends between -15° and 15 ° and the right dark zone between 30° and 60°.

(a) Dual layer array

Figure 3.18: Close up of the dual layer array of 16 sources used by Shin *et al.* [66]. Reproduced from [66].

It can be desirable to radiate different sound programmes inside a room. One example could be given by three TV viewers that are watching a film, but each one is listening to the audio in a different language. This can be obtained with a line array which creates three different bright zones, as shown in Fig. 3.17. The

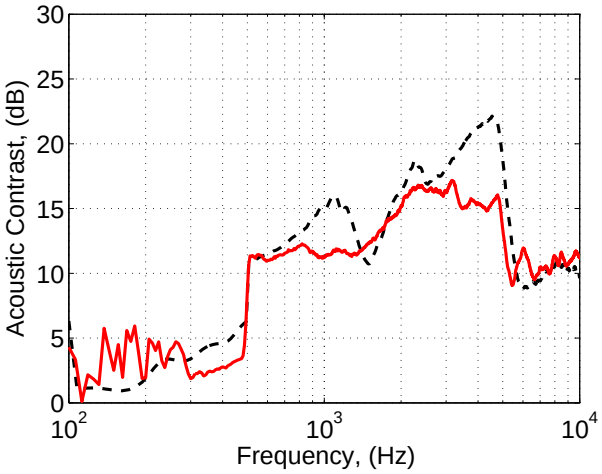
array is being driven by three different sets of filters, so that each set of filters is created to obtain a control in the soundfield which maximises the acoustic pressure and obtains a good audio quality in the bright zone. The mean square pressure is minimised in the rest of the control geometry, so as not to annoy the rest of the listeners which are listening to another language. To create these three control zones, two practical arrays are considered, and the performance is evaluated using the manipulation of measured frequency responses.

The first of the arrays is the 1×8 array, shown in the previous section. The second array, shown in the lower plot of Fig. 3.18, is a dual layer array formed by 16 independently controlled sources, as this presented by Shin *et al.* [66]. The transfer functions of both arrays have been measured in the control geometry of Fig. 3.4 in an anechoic chamber, which allows us to perform off-line performance simulations. The sources of the phase-shift array are separated 35 mm horizontally and the sources of the dual layer array 38 mm, which gives a broadside aliasing limit of about 9.8 kHz and 9 kHz respectively. The two rows of speakers of the dual layer array are separated by 12 cm, giving an spatial aliasing limit of 1.5 kHz. For the creation of the array control filters, the least squares method has been used. The bright zone transfer functions have been used as target pressures, with the dark points set to zero.

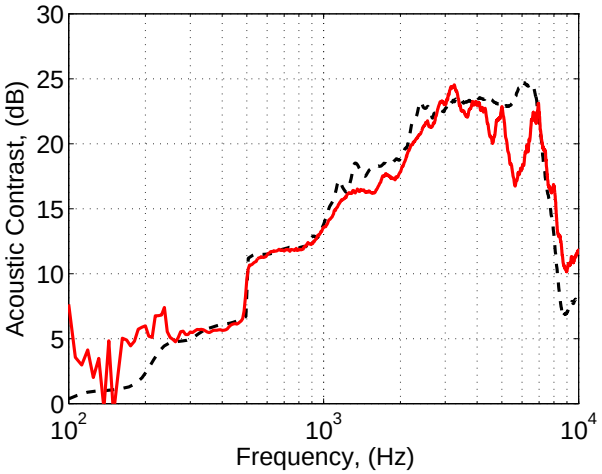
The acoustic contrast results calculated for the two arrays in each of the control geometries of Fig. 3.17 is presented in Fig. 3.19, wherein the array effort has been limited to be lower than 6 dB for both arrays. Both arrays give a very similar result in the central control zone, as the phase-shift sources behave in a similar way to hypercardioids. The results for the left control geometry show a better performance of the dual layer array. Around 1 kHz and 4.5 kHz the dual layer array is able to give about 3 and 6 dB more contrast than the phase-shift sources array. However, looking at the acoustic contrast performance in the right bright zone control geometry, it can be seen that the performance that both array give is similar. Theoretically, the acoustic contrast performance should be the same when the phase-shift array is steered for both sides, however, this is not the case. The sources of the phase-shift array have been hand built, and do not present a perfect, symmetrical, hypercardioid, which causes the performance of the phase-shift array when it is steered to the right to be larger than that when it is steered to the left, which presents more equally radiating sources. At around 1.5 kHz there is a dip in the response of the dual layer array when this is steered to both sides, as at this frequency the array is working right on the spatial aliasing frequency.

The directivities for both arrays when these are steered according to the different control geometries are shown for various frequencies in Fig. 3.20. It can be observed how these are very similar for all the steering configurations at 1 kHz, 2 kHz and 4 kHz. The results at 6 kHz and 7 kHz show how large level secondary lobes start to appear for the left and right control configurations. The spatial aliasing limit to generate a plane wave with an array having a separation d in a direction θ is given by Eq. 3.10. When the array is steered to the left or to the bright it is constrained "to extend a control" until 60° from broadside configuration, which corresponds to a spatial aliasing frequency of about 4.8 kHz for the 1×8 array and of about 5.2 kHz for the dual layer array. This fact can be noticed from the level of the secondary lobes presented by the dual layer array, which are greater than those presented by the dual layer array. At 7.5 kHz the secondary lobes also appear for the central control zone, as the frequency is close to the spatial aliasing limit for that control configuration.

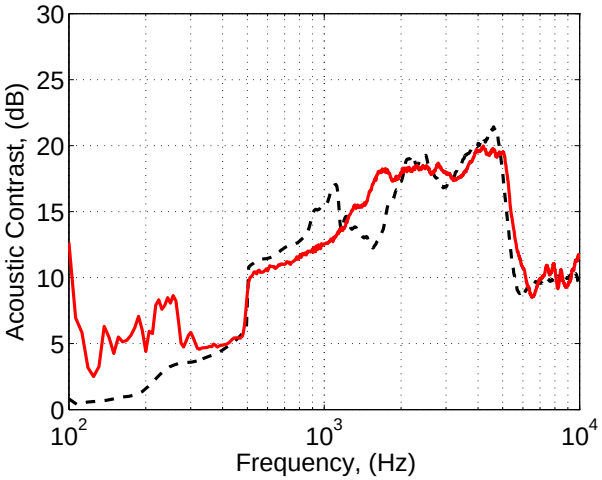
A robustness analysis has also been performed by calculating the condition number with respect to the inversion of the matrix $\mathbf{Z}^H\mathbf{Z}$, which is shown in Fig. 3.21. It is shown in this figure how the condition number of the phase-shift sources array is much lower than that of the dual layer array at low frequency. The condition number of the phase-shift sources array increases to be very similar to that of the dual layer array at about 1 kHz and then decreases again, to be around 10 dB lower at around 2 kHz. After 3kHz the condition number presented by both arrays is very similar. As the condition of the phase-shift array is lower, the array is more robust respect to mismatches in sensitivity and position of the sources.



(a) Left bright zone

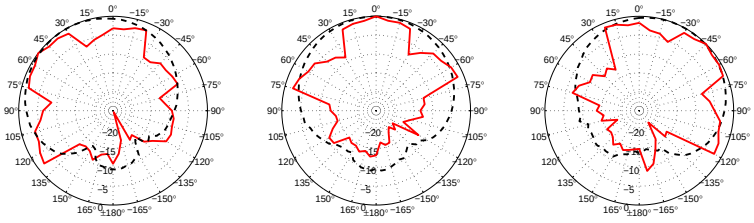


(b) Central bright zone

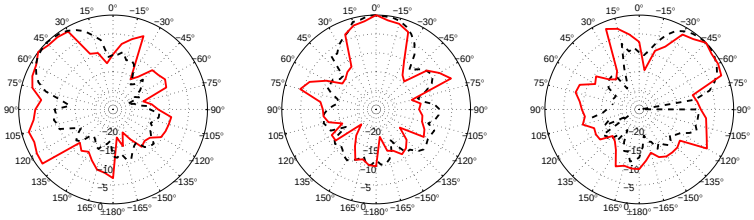


(c) Right bright zone

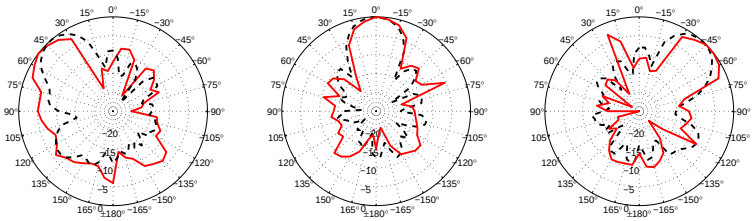
Figure 3.19: Acoustic contrast against frequency for an 8 phase-shift sources array (red solid line) and for an 8 double monopole sources array (black dashed line), calculated from measured responses for each of the control geometries disposed in Fig. 3.17. The array effort has been limited for both arrays to be lower than 6 dB.



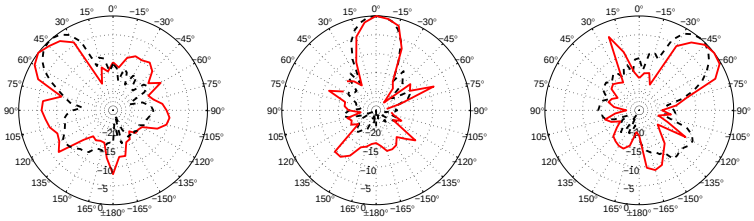
(a) Left, 500 Hz **(b)** Central, 500 Hz **(c)** Right, 500 Hz



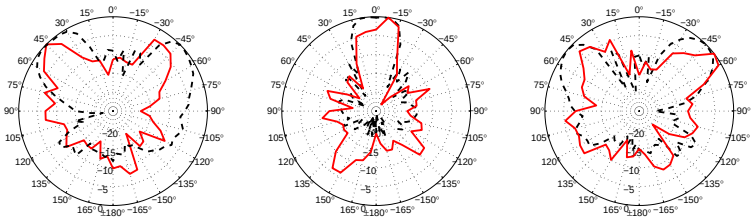
(d) Left, 1 kHz **(e)** Central, 1 kHz **(f)** Right, 1 kHz



(g) Left, 2 kHz **(h)** Central, 2 kHz **(i)** Right, 2 kHz



(j) Left, 4 kHz **(k)** Central, 4 kHz **(l)** Right, 4 kHz



(m) Left, 6 kHz **(n)** Central, 6 kHz **(o)** Right, 6 kHz

Figure 3.20: Directivities at various frequencies obtained with the 8 phase-shift sources array (red solid line) and with the 8 double monopole sources array (black dashed line), for the different control zones shown in Fig. 3.17. The array effort has been limited for both arrays to be lower than 6 dB.

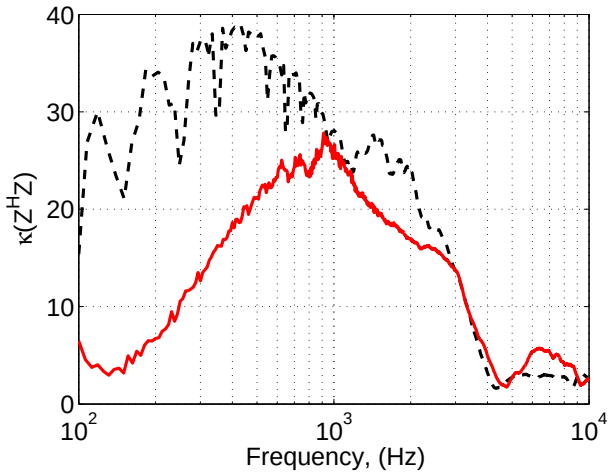


Figure 3.21: Condition number respect to the inversion of $\mathbf{Z}^H\mathbf{Z}$. The results for the phase-shift array are shown in red solid line and the results for the dual layer array are shown in black dashed line.

3.4 The 4×8 array

Although an acoustic array can show good directive characteristics in the free-field, this directivity can be significantly reduced when the source is placed in a closed room. In order to overcome this limitation, we need to minimise the reverberant input of the array, which can only be obtained by increasing the average absorption coefficient of the room walls or by minimising the radiation of the source to other directions rather than on axis. From the point of view of the application discussed here, the second option is the only valid one.

For a given on axis pressure, the radiated power of the source is inversely proportional to its directivity index, see Eq. 2.18. To reduce the radiated power, the acoustic source thus has to be more directive in a 3D sense. Considering an array of point monopole sources placed in a Cartesian coordinate system where the array is placed along the x axis, a beam can be created along the x axis, however, the radiation pattern will be symmetrical all around it. If an additional beamforming is wanted to be created around y and z axis, the array sources have to be extended also along these axis. In the 1×8 array this was obtained passively using phase-shift sources, however, the radiation along z axis was only controlled by the oval shape of the individual drivers.

The performance of the 1×8 array could be upgraded by adding more sources vertically, however, this needs of the use of more digital filters, which increases the complexity and reduces the robustness of the array [94]. Another possible solution is to increase the directivity along the z axis by increasing the effective size of the source in this direction, e.g., by driving several loudspeakers placed vertically with the same signal. This is the solution which is presented in this section; an array of 32 sources driven by 8 digital filters, in where each *individual* source is formed by a 4 source delay and sum.

3.4.1 Design

In order to obtain a line array which gives good directional characteristics and has a reasonable size, the formulation for the estimation of the reverberant performance of a personal audio array presented in Section 5.3 has been used.

Each individual source is composed of 4 sources similar to those used by the 1×8 array. This leads to a directivity improvement in a vertical sense, as can be observed in Fig. 3.22, where the predicted directivity balloons of a 1×8 and

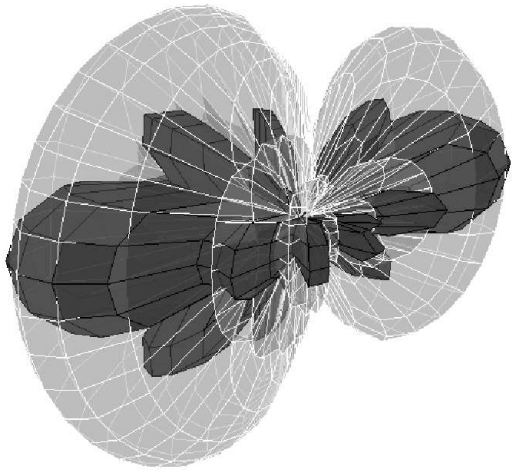


Figure 3.22: Predicted directivity balloons of the 1×8 (light) and the 4×8 (dark) arrays under anechoic conditions.

a 4×8 arrays are plotted. The horizontal directivity is barely modified, but a great improvement in the vertical directivity is obtained. Although more sources could be used, 4 sources can be wired together as shown in Fig. 3.23 to present the same electrical impedance as a single source, which is desirable to reduce the power used to drive the array. More information about the design and the acoustical performance of the sources used by the 4×8 array can be found in Section 2.4.

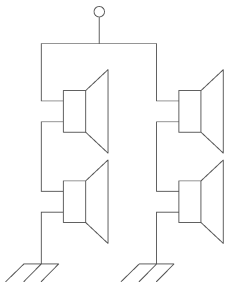


Figure 3.23: Wiring used for each of the 4×8 array's sources.

The separation between each source has been selected to be of 3.5 cm horizontally, as used in the 1×8 array. This distance places the aliasing frequency in the horizontal direction to be of 7.8 kHz in a control zone as that of Fig. 3.4. The vertical separation between the individual drivers which form a 4×8 source has been selected by running a point hypercardioid source simulation using the formulation for the estimation of reverberant directivity introduced in Chapter 5. Fig. 3.24 shows the acoustic contrast performance of both the 1×8 and the 4×8 arrays, calculated for a control zone as the one of Fig. 3.4, for the free-field and

inside a listening room[1]. Both the 1×8 and 4×8 arrays, show a similar free-field performance, as they have the same aperture. However, the performance of the 4×8 in the reverberant room is much higher above 1 kHz. At low frequency, the performance is similar to that of the 1×8, since the array works following an omnidirectional radiation pattern, but as the frequency increases, the array gains directionality, lowering the reverberant input and obtaining a high acoustic contrast.

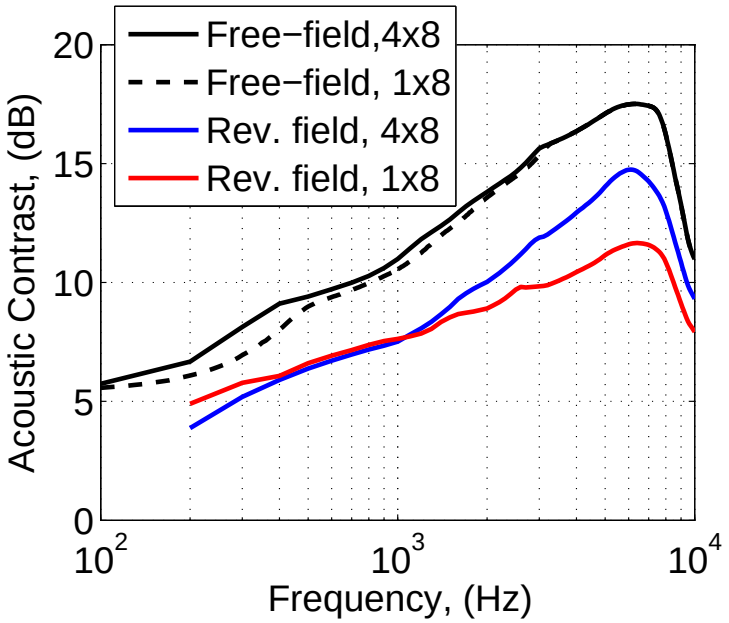


Figure 3.24: Prediction of free-field and reverberant acoustic contrast performance for the 1×8 and the 4×8 array.

The quality of the vertical beamforming obtained by each 4×8 individual source is constrained by the differences in sensitivity between each source. An analysis of the source sensitivity and its estimated robustness to errors is presented in Section 2.4.3.

3.4.2 Measured performance

Free-field performance

After measuring the transfer functions of the 4×8 array in an anechoic chamber, the least squares method has been used to design filters for the array, using the transfer functions of the bright zone of Fig. 3.4 as target pressure. Using this set of filters, the performance has been measured in the control zone of Fig. 3.4.

[1]The reverberation time over frequency of the listening room is shown in Fig. 3.30.

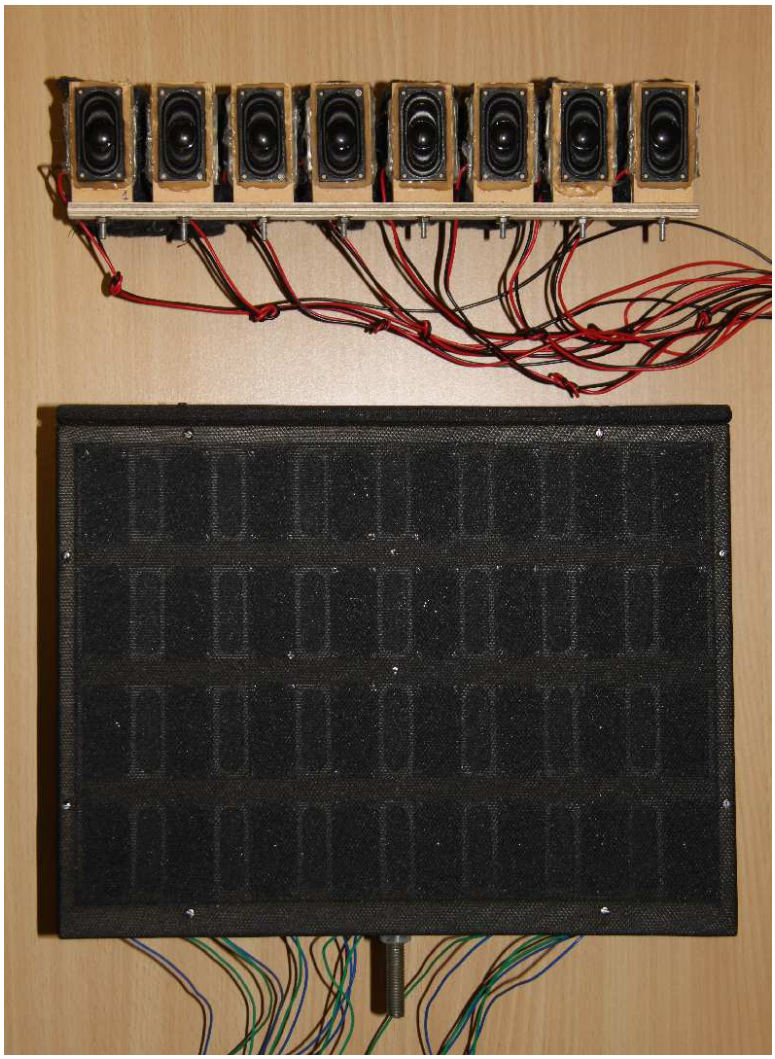


Figure 3.25: View of the 1×8 array (top), and the 4×8 array (bottom).

These results are shown in Fig. 3.26, where the corresponding acoustic contrast of the 1×8 array is also shown for comparison. The array effort has been limited to be below 6 dB in both cases, as this figure allows a good directive pattern to be obtained, while at the same time limiting the maximum magnitude of the array filters sufficiently to prevent over-driving of the sources. Both arrays offer a very similar acoustic contrast figure, as theoretically, they should have practically the same performance in this horizontal control zone. The performance of the 4×8 array, however, is slightly lower than that of the 1×8 array at frequencies below 3 kHz, which can be attributed to the fact that the sources of the 1×8 array are more directional in the horizontal sense, since the bigger baffle of the 4×8 array makes it more difficult to create the phase-shift effect.

A three-dimensional directivity estimation provides a more realistic measure of performance of the 4×8 array. By measuring the transfer responses both in a

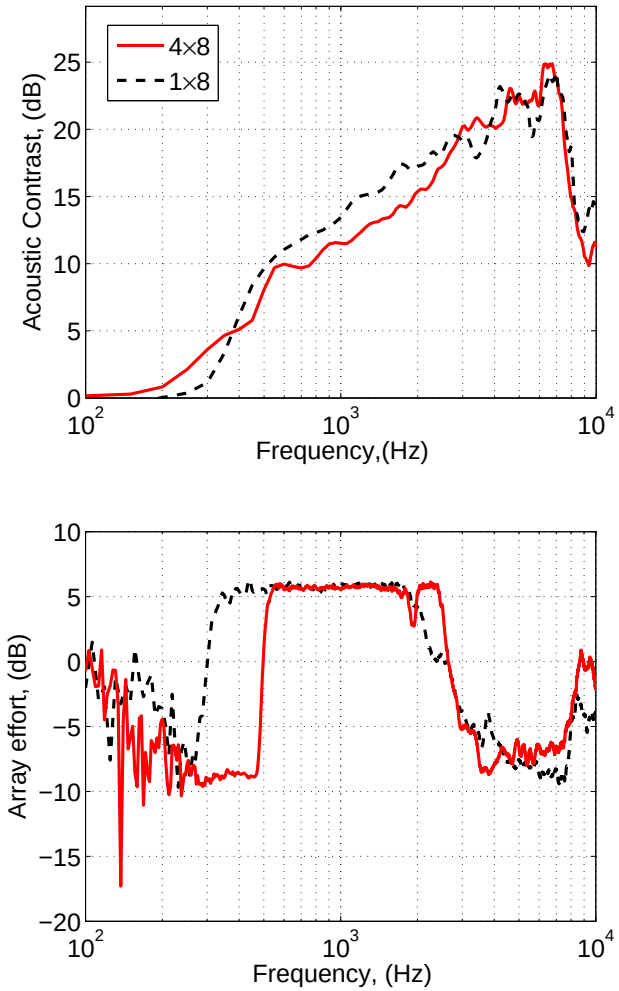


Figure 3.26: Free-field acoustic contrast measured on the 2D control geometry shown in Fig. 3.4 and array effort needed by the LSIF. The solid line refers to the results of the 4×8 array and the dashed line refers to those of the 1×8 array.

horizontal and a vertical plane in an anechoic chamber, values can be linearly interpolated to give an estimation of the transfer responses to a sphere of points surrounding the array. Using these transfer responses, a 3D control zone has been created as shown in Fig. 3.27, consisting of 1106 microphones. The bright zone is the same as that presented in Fig. 3.4 but is extended vertically between -7.5° and 7.5°, so that now it is formed by 9 microphones. The rest of the microphones define the dark zone.

Using the same set of filters as for the results of Fig. 3.26, the performance that both arrays produce in the control zone of Fig. 3.27 has been calculated via off-line simulations. These results are shown in fig. 3.28, where it can be observed how the high frequency performance of the 4×8 array is much better than that of the 1×8 array, due to the benefits of the vertical beamforming obtained by

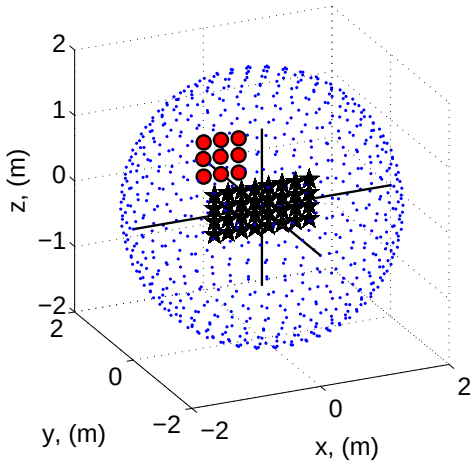


Figure 3.27: 3D control geometry used for measuring the directional performance of both arrays. Dots represent the dark zone's control microphones and closed circles represent the bright zone's control microphones. The 32 stars represent the sources of the 4×8 array.

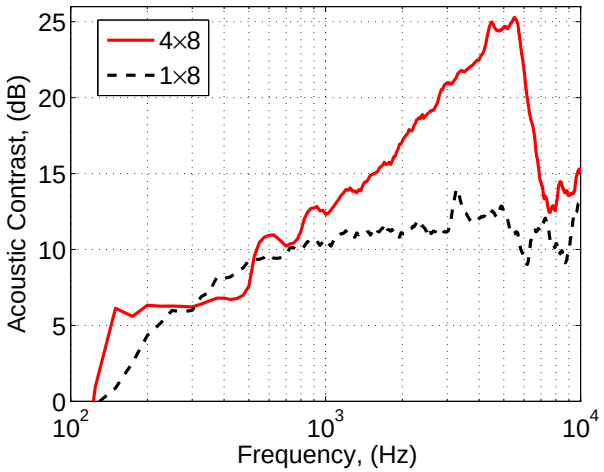


Figure 3.28: Comparison of the free-field acoustic contrast predicted from the measured frequency responses for the 1×8 and 4×8 arrays, in the 3D control geometry of Fig. 3.27. The solid line represents the results of the 4×8 array whilst the dashed line shows those of the 1×8 array.

driving each column source in phase, leading to an improvement of about 12 dB of acoustic contrast at around 5 kHz. This is a very encouraging result, as both arrays use the same number of control filters.

Reverberant performance

The acoustic contrast has also been measured for the two arrays in a listening room, as shown in Fig. 3.29. The room has a surface of about 112 m^2 and a volume of about 73 m^3, with a distribution of reverberation time over frequency as

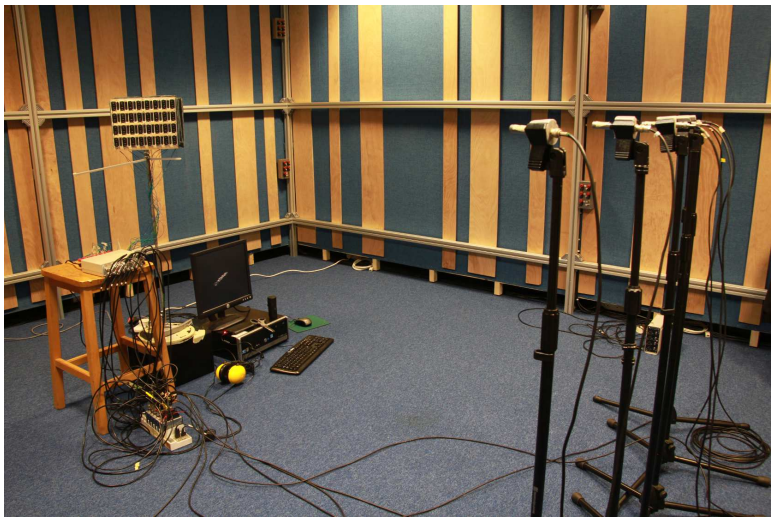


Figure 3.29: Set up of the measurements performed in a listening room.

shown in Fig. 3.30. The measurements have been performed using the horizontal control zone defined in Fig. 3.4, with the same set of filters as for the free-field results. The reverberant contrast results are show in Fig. 3.31, compared to those obtained in the free-field. The reverberant results for the 4×8 array show that a reduction between 3 and 5 dB is obtained with respect to the anechoic case. The 4×8 array is able to provide a much greater performance than the 1×8 array in the reverberant environment. This increases almost constantly from 500 Hz, to obtain around 5 dB of gain in contrast around 5 kHz. These results are in the fashion of the 3D free-field contrast results, shown in Fig. 3.28, where it can be seen how due to the vertical beamforming, the 4×8 array obtains a much higher acoustic contrast, which in the case of the reverberant results, leads to a lower excitation of the reverberant field.

The directivity of the 4×8 array has been measured in the free-field and in the listening room using the procedures of Appendix J, and is shown for some frequencies in Fig. 3.32. In the free-field results the secondary lobes are very small compared with the main lobe, and the back radiation is greatly attenuated due to the phase-shift effect of the sources. The main lobe is centred at about 7.5°, according to the definition of the bright zone. In the reverberant results, the level at other points apart form the bright zone is increased. This is due to the extra pressure component that the reverberant field introduces.

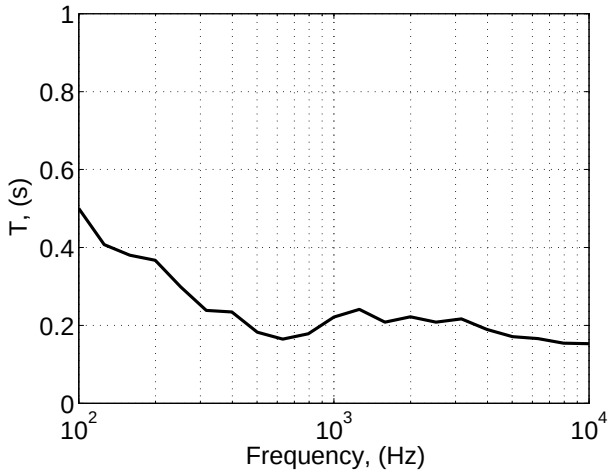


Figure 3.30: Reverberation time of the audio listening room where the performance of the 1×8 and the 4×8 array has been assessed. The room has a surface of about 112 m^2 and a volume of about 73 m^3.

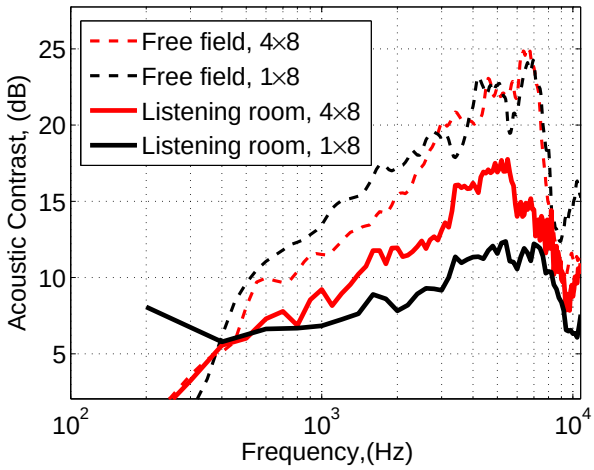
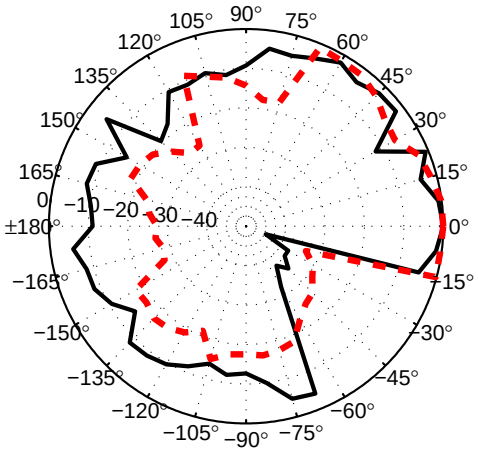
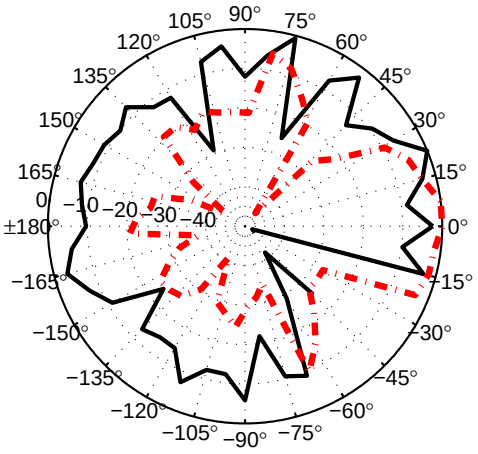


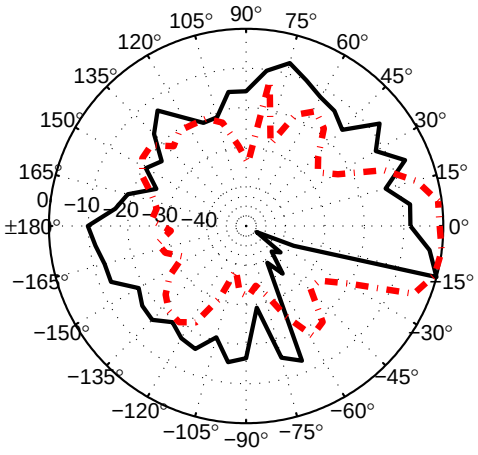
Figure 3.31: The acoustic contrast measured in a listening room in comparison with that obtained in anechoic conditions. The free-field results for the 4×8 array are represented by the red dashed line, the free-field results for the 1×8 array are represented by the dark dashed line, the reverberant results for the 4×8 array are shown by the red solid line and the reverberant results for the 1×8 array are shown by the dark solid line.



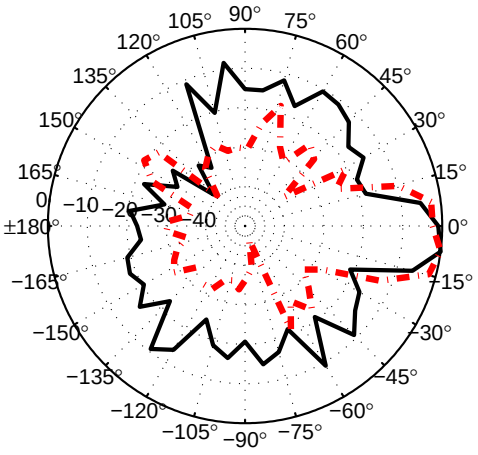
(a) 500 Hz



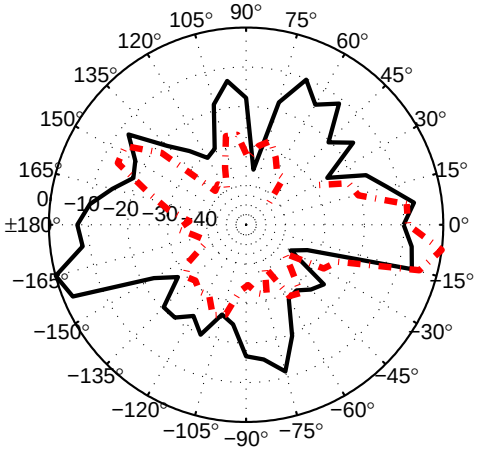
(b) 1 kHz



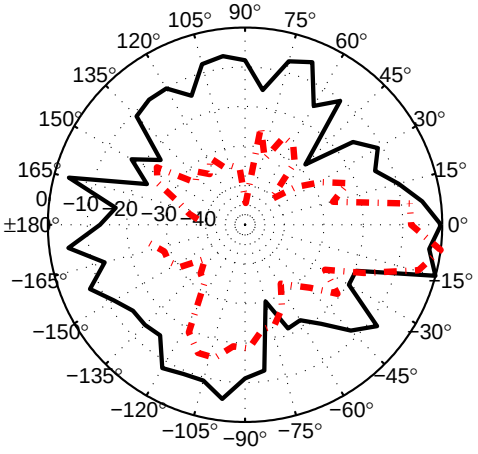
(c) 2 kHz



(d) 4 kHz



(e) 6 kHz



(f) 8 kHz

Figure 3.32: Magnitude of the pressures at the control geometry of Fig. 3.4 at various frequencies for the 4×8 array. The solid line represents the results in the free-field, and the dashed line represents the results in the reverberant room.

3.5 Conclusions

This chapter has first introduced the theory of directional acoustic sources by superposition of a certain number of radiators a certain distance apart. A radiator starts to become directional when its largest dimension is comparable to the wavelength it is radiating. Hence, a radiator which is not directional at a certain frequency can be made more directional by placing various identical radiators and drive all of them in phase, which is commonly denominated as *loudspeaker array*. At certain point in the space the waves propagating from the ensemble of radiators arrive in phase, creating a constructive interference. The directionality of such an array increases with frequency until a certain value, in where more than one point of constructive interference exist, which is denominated as *spatial aliasing*. By applying electronic delays to each of the sources it is also possible to steer the main beam of the array and select the direction of maximum propagation.

At frequencies at which the radiation wavelength is large compared with the maximum dimension of the array, the directional response of such devices is, however, low. In order to overcome such limitation, superdirective beamforming techniques are introduced. Such techniques are based on the a priori knowledge of the transfer functions between the array and the control microphones, which involves the measurement of such quantities in an anechoic chamber or in the environment where the array is being used. As this is a process that can result time and cost expensive, approaches based in analytical transfer functions have been used, which have shown to provide an effective performance. A deeper study of these approaches is presented in Appendix D. Three different formulations have been considered for creating a superdirective beamforming, these are: the acoustic contrast maximisation (ACM), the least square inverse filters (LSIF) and the Chang and Jacobsen mixed control approach. The ACM maximises the contrast of acoustic potential energy between two control zones, without control of the soundfield in the bright zone, whilst that the LSIF allows to control the phase and magnitude in the bright zone by selecting a target pressure. Chang and Jacobsen approach allows an interplay between a pure contrast control approach and the exact matching of the target pressure in the reproduced zone. For the array application, the LSIF approach is preferred to the ACM, as it provides a more uniform distribution of acoustic pressure throughout the listening zone whilst obtaining practically the same directivity. The superdirective approaches do, however, need a huge amount of electrical power to operate, and are very sensitive to errors and mismatches in the characterisation of the transfer functions.

The amount of electrical power and the robustness of the array with respect to such source mismatches can be controlled by introducing a regularisation parameter, which controls the inversion of the matrix of transfer impedances. A deeper analysis of the effect of regularisation with respect to source mismatches is presented in Chapter 5.

The design of time domain filters here obtained by inverse Fourier transformation of frequency responses is described in Appendix C. The design of these filters entirely in the time domain is also discussed in Appendix 4.

Based on the knowledge gained in the design of phase-shift radiators, which is summarised in Chapter 2, two arrays have been designed and built. The first of these two radiators consists of 8 phase-shift sources, and is denominated as the 1×8 array [53]. This array is able of obtaining a contrast that grows from 10 dB at 500 Hz to 25 dB at 7 kHz, using an array effort of about 6 dB, which constitutes an improvement of about 6 dB with respect to a delay and sum line array at such array effort. The free-field acoustic contrast has shown to be greatly reduced when the array is introduced into a reverberant environment. In order to keep the power input to the reverberant field low, the directional characteristics of the individual sources of a radiator are, hence, very important. An example can be brought by the 1×8 array, which uses hypercardioid sources. Such sources theoretically minimise the power input to the reverberant field, however, when the 1×8 is introduced into the listening room, a poor directivity is obtained. A more in deep study of the effect of reverberation in the performance of personal audio systems will be presented in Chapter 4.

Simulations based on the 1×8 array and in a dual layer array have shown that the performance of an array of 8 phase-shift hypercardioid sources is very similar to that obtained with a dual layer array, which uses 16 independently controlled sources. A phase-shift source array presents a clear advantage respect to a dual layer array, as it needs the half amount of sources to function. This makes a phase-shift source array more economic and simple to control. As the number of sources is lower, the array is also more robust to errors in the transfer responses.

In order to further reduce the pressure input into the reverberant field and at the same time keep the same processing complexity, the second array that has been introduced in this thesis uses phase-shift sources which are very directional in a vertical sense. This array is denominated 4×8 array [95] and uses 32 individual phase-shift radiators as that of the 1×8 array arranged in columns of four, in

which each column is driven in phase, hence obtaining a vertical delay and sum source. This array obtains the same free-field directional performance as the 1×8 array, but obtains a large improvement in performance in a reverberant room, of about 5 dB at 5 kHz.

Chapter 4

Time domain formulation of superdirective filters for personal audio applications

Chapter 3 has considered the creation of optimal superdirective filters whose responses are optimised in the frequency domain. Such optimisation is performed frequency by frequency, with the time series obtained afterwards via inverse Fourier transformation. An alternative is that of performing the optimisation in the time domain, where the whole set of frequencies are optimised at a time, as considered in this appendix.

The filters are created from the point of view of a personal audio problem where the acoustic energy contrast between two control regions is to be maximised, paying particular attention to the quality of the reproduced audio. The performance of the time domain formulation is assessed by means of simulation using free-field point source Green functions and measured transfer functions of the 4×8 array. The results of these simulations is compared with the performance of arrays using filters formulated in the frequency domain. The outcome shows that using the least squares algorithm or the mixed control algorithm of Chang and Jacobsen, the filters created in both domains give the same acoustic performance. The filters optimised in the time domain give, however, a better directional performance for short filter lengths. Finally a comparison of how to select the target pressure to maximise speech intelligibility is presented.

4.1 Filtering theory

The control of acoustic radiation in the time domain has been first considered for the deconvolution of room responses by Neely and Allen [140], in order to obtain a flat response and undo the effect of a room or a speaker in the reproduced sound. Kirkeby and Nelson [71] extended this formulation to comprise a larger number of loudspeaker and microphones or target points. The optimisation of acoustic fields in the time domain has also been applied to active control of sound by Elliott [141], using matrices of filtered references. Using the active control formulation, the theory for maximising the acoustic contrast between two control zones has been presented recently by Elliott and Cheer [75]. Cai *et al.* [74, 76] recently studied the control of an 8 sources array using a modified formulation of the time domain acoustic contrast.

The work presented here studies the performance of the filters created in the time domain from the point of view of a personal audio device aimed to increase the speech intelligibility of the hard of hearing. An analysis is presented which compares the performance of filters obtained with the time domain formulation of the acoustic contrast maximisation and the least squares filters with that obtained by the frequency domain formulation by the same algorithms. A first set of simulations is introduced with a 2 point source endfire array, in where the length of the filters is brought to analysis. Later simulations are performed with an 8 point hypercardioid source array, in where the performance of the time domain acoustic contrast maximisation is studied. Finally the performance of the 4×8 array is predicted, and frequency dependent regularisation is applied to prevent the damaging of the array sources. The audio quality of the reproduced sound is assessed in terms of the STI [42].

4.2 Time domain filter optimal filters

4.2.1 Formulation using matrices of impulse responses

The theory presented in [71] is rearranged here to formulate a personal audio problem, in which a pressure maximum is generated in the bright zone and a pressure minimum is created in the dark zone. A block diagram of the problem is presented in Fig. 4.1, where an array of M sources is driven by a set of filters $\mathbf{w}$. The sound field is sampled in the two control zones by N_B bright microphones and N_D dark microphones, distributed throughout the control space. The transfer responses between the loudspeakers and the control microphones are rearranged

into two impulse response matrices for bright and dark control zones, $\mathbf{G}_B$ and $\mathbf{G}_D$.

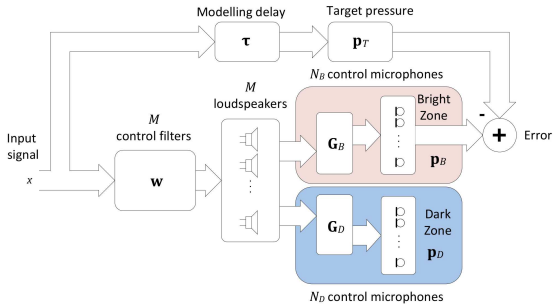


Figure 4.1: Block diagram of the acoustic control problem.

The sampled output at the n_B-th microphone in the bright zone of Fig. 4.1 due to all the control sources is given by

$$p_{Bn_B}(n) = \sum_{m=1}^{M} \sum_{j=0}^{J-1} \sum_{i=0}^{I-1} g_{Bn_Bm}(j) w_m(i) x(n-i-j), \tag{4.1}$$

where $x(n)$ is the sampled input signal to each of the filters with impulse response $w_m(i)$, and $g_{Bn_Bm}(j)$ is the j-th coefficient of the impulse response between the m-th driver and the n_B-th microphone in the bright zone, which is modelled as an FIR filter of length J.

Although, in practice, the optimisation is required over an expectation of a set of input signals, a deterministic problem is obtained if the input signal, $x(n)$, is a transient with an equivalent spectral content. In the original formulation [71], the input signal was assumed to have a flat spectrum so $x(n)$ was as assumed to be a delta function and this is used below. It is, however, possible to generalise this by defining a modified impulse response function in Eq. 4.1 which is the convolution of the physical impulse response function and the response of a filter that produces the spectral content of the input signal.

Following the assumption that $x(n)$ is equal to $\delta(n)$, Eq. 4.1 can be simplified to

$$p_{Bn_B}(n) = \sum_{m=1}^{M} \sum_{j=0}^{J-1} g_{Bn_Bm}(j) w_m(n-j). \tag{4.2}$$

The contribution of the m-th source to the n_B-th microphone of the bright zone can also be obtained if g_{n_Bm} is rearranged into an impulse response matrix, $\mathbf{G}_{n_Bm}$,

and is multiplied by the filter corresponding to that source, $\mathbf{w}_m$, which gives

$$\mathbf{p}_{Bn_Bm} = \mathbf{G}_{n_Bm}\mathbf{w}_m, \tag{4.3}$$

where $\mathbf{p}_{Bn_Bm}$ is the vector of the $J+I-1$ impulse response points of $p_{Bn_Bm}(n)$, and $\mathbf{G}_{n_Bm}$ is defined as

$$\mathbf{G}_{n_Bm} = \begin{bmatrix} g_{n_Bm}(0) & 0 & 0 \\ \vdots & \ddots & 0 \\ g_{n_Bm}(I-1) & \ddots & g_{n_Bm}(0) \\ 0 & \ddots & \vdots \\ 0 & 0 & g_{n_Bm}(I-1) \end{bmatrix}, \tag{4.4}$$

with a size $(J+I-1)\times I$, and $\mathbf{w}_m$ is a vector of the I impulse response points of $w_m(i)$.

A vector of sampled outputs due to all the sources for all of the microphones in the bright zone can then be formed, which is

$$\mathbf{p}_B = \begin{bmatrix} \mathbf{p}_{B1}^T & \mathbf{p}_{B2}^T & \cdots & \mathbf{p}_{BN_B}^T \end{bmatrix}^T, \tag{4.5}$$

which may now be written as

$$\mathbf{p}_B = \mathbf{G}_B\mathbf{w}, \tag{4.6}$$

where $\mathbf{G}_B$ is a matrix composed by the matrices of impulse responses from each source to each control point $\mathbf{G}_{n_Bm}$

$$\mathbf{G}_B = \begin{bmatrix} \mathbf{G}_{11} & \cdots & \mathbf{G}_{1M} \\ \vdots & \mathbf{G}_{n_Bm} & \vdots \\ \mathbf{G}_{N_B1} & \cdots & \mathbf{G}_{N_BM} \end{bmatrix}, \tag{4.7}$$

with a size $N_B(J+I-1)\times MI$, and the vector of all the filter coefficients, $\mathbf{w}$, is defined as

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1^T & \cdots & \mathbf{w}_M^T \end{bmatrix}^T, \tag{4.8}$$

with a size $MI\times 1$.

Similarly, the vector of sampled outputs to all the N_D control microphones of the dark zone can be written as

$$\mathbf{p}_D = \mathbf{G}_D\mathbf{w}, \tag{4.9}$$

where $\mathbf{G}_D$ is the matrix of impulse responses from the M sources to the control points in the dark zone, with size $N_D(J+I-1) \times MI$, created in the same way as $\mathbf{G}_B$.

4.2.2 Optimal time combined solution filters

The vector of superdirective optimal filters is obtained here using the "combined solution" of Chang and Jacobsen [67]. This approach gives a weighted least squares solution, allowing a trade-off between the minimisation of the mean squared error with respect to a desired pressure field, p_T, in the listening zone, i.e., the bright zone, and the maximisation of the contrast of mean squared pressure between both control zones by minimising the potential energy in the dark zone. To this end, a cost function is defined as

$$J = \xi \mathbf{p}_D^T \mathbf{p}_D + (1-\xi)\left[(\mathbf{p}_B - \mathbf{p}_T)^T(\mathbf{p}_B - \mathbf{p}_T)\right], \tag{4.10}$$

where $\xi (0 \leq \xi < 1)$ determines the balance between minimising the potential energy in the dark zone and reducing the mean square error in the bright zone, with $\xi = 0.5$ corresponding to the conventional least squares solution [67]. Substituting Eqs. 4.6 and 4.9 into Eq. 4.10 leads to

$$\begin{aligned} J = \mathbf{w}^T \left[\xi \mathbf{G}_D^T \mathbf{G}_D + (1-\xi)\mathbf{G}_B^T \mathbf{G}_B\right] \mathbf{w} + \\ (1-\xi)\left[\mathbf{w}^T \mathbf{G}_B^T \mathbf{p}_T + \mathbf{p}_T^T \mathbf{G}_B \mathbf{w}\right]. \end{aligned} \tag{4.11}$$

Provided that $\left[\xi \mathbf{G}_D^T \mathbf{G}_D + (1-\xi)\mathbf{G}_B^T \mathbf{G}_B\right]$ is positive definite, the above quadratic cost function has a unique global minimum [141] at a point given by

$$\frac{\partial J}{\partial \mathbf{w}} = 2\left[\xi \mathbf{G}_D^T \mathbf{G}_D + (1-\xi)\mathbf{G}_B^T \mathbf{G}_B\right] \mathbf{w} + 2(1-\xi)(-\mathbf{p}_T^T \mathbf{G}_B). \tag{4.12}$$

The optimal set of coefficients is then given by

$$\mathbf{w} = \left[\xi \mathbf{G}_D^T \mathbf{G}_D + (1-\xi)\mathbf{G}_B^T \mathbf{G}_B\right]^{-1} (1-\xi)\mathbf{G}_B^T \mathbf{p}_T. \tag{4.13}$$

In order to control the magnitude of the optimal filter coefficients, an extra term can be introduced to weight the amount of electrical power used in the above solution. This term is proportional to the sum of the squared filter weights and leads to the optimal solution whilst limiting the maximum amplitude of the filter weights. The cost function is defined in this case as

$$J = \xi \mathbf{p}_D^T \mathbf{p}_D + (1-\xi)\left[(\mathbf{p}_B - \mathbf{p}_T)^T(\mathbf{p}_B - \mathbf{p}_T)\right] + \beta \mathbf{w}^T \mathbf{R} \mathbf{w}. \tag{4.14}$$

where β is the regularisation parameter and $\mathbf{R}$ is the array effort weighting matrix. By carefully selecting the matrix $\mathbf{R}$ it is possible to perform frequency dependent regularisation. To this end, an FIR filter b is used which boosts the response at the frequencies to be regularised. The filter b is arranged to form a matrix $\mathbf{B}$, so that

$$\mathbf{B} = \begin{bmatrix} b(0) & 0 & 0 \\ \vdots & \ddots & 0 \\ b(I-1) & \ddots & b(0) \\ 0 & \ddots & \vdots \\ 0 & 0 & b(I-1) \end{bmatrix}, \tag{4.15}$$

which is employed to form the matrix $\mathbf{R}$ in the cost function of Eq. 4.14, as this is formed by M copies of $\mathbf{B}^T\mathbf{B}$

$$\mathbf{R} = \begin{bmatrix} \mathbf{B}^T\mathbf{B} & 0 & \cdots & 0 \\ 0 & \mathbf{B}^T\mathbf{B} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{B}^T\mathbf{B} \end{bmatrix}. \tag{4.16}$$

The set of optimal filters is given in this case by

$$\mathbf{w} = \left[\xi\mathbf{G}_D^T\mathbf{G}_D + (1-\xi)\mathbf{G}_B^T\mathbf{G}_B + \beta\mathbf{R}\right]^{-1}(1-\xi)\mathbf{G}_B^T\mathbf{p}_T. \tag{4.17}$$

Apart from limiting the power used by the array in certain frequency ranges, it is important to be able to monitor how much energy the array filters use at different frequencies by defining an array effort in the frequency domain. By Fourier transforming the time series of the filters, i.e, $w(n) \xrightarrow{F} q(j\omega)$, the array effort can be defined for a single frequency as

$$AE(j\omega) = 10\log_{10}\left(\frac{\sum_{m=1}^{M} q_m(j\omega)^* q_m(j\omega)}{|q_{MON}^2(j\omega)|}\right), \tag{4.18}$$

where q_m is the source strength in the frequency domain of the source m and q_{MON} is the source strength that a monopole source or one of the central sources of an array needs in order to obtain the same pressure as that produced by the array in the bright zone [53].

It is also necessary to characterise the directional performance obtained with a given set of filters. To this end, the time-averaged contrast between the mean

squared pressures in the two control zones, or acoustic contrast, may be defined as

$$C = \frac{N_D}{N_B} \frac{\mathbf{p}_B^T \mathbf{p}_B}{\mathbf{p}_D^T \mathbf{p}_D}, \tag{4.19}$$

where the ratio $\frac{N_D}{N_B}$ makes the estimation independent of the number of bright and dark control points. Using Equations 4.6 and 4.9 Eq. 4.19 can also be written as

$$C = \frac{N_D}{N_B} \frac{\mathbf{w}^T \mathbf{G}_B^T \mathbf{G}_B \mathbf{w}}{\mathbf{w}^T \mathbf{G}_D^T \mathbf{G}_D \mathbf{w}}. \tag{4.20}$$

However, Eq. 4.20 gives a scalar which averages the contrast over time. In a similar way to the array effort, it is also possible to observe the acoustic contrast as a function of frequency. In this case, the acoustic contrast is calculated at a single frequency as

$$C(j\omega) = \frac{N_D}{N_B} \frac{\sum_{n_B=1}^{N_B} \sum_{m=1}^{M} q_m(j\omega)^* Z_{n_B m}(j\omega)^* Z_{n_B m}(j\omega) q_m(j\omega)}{\sum_{n_D=1}^{N_D} \sum_{m=1}^{M} q_m(j\omega)^* Z_{n_D m}(j\omega)^* Z_{n_D m}(j\omega) q_m(j\omega)}, \tag{4.21}$$

where the subscript * stands for the complex conjugate and $Z_{n_B m}(j\omega)$ and $Z_{n_D m}(j\omega)$ are are obtained by Fourier transformation of $g_{n_B m}(n)$ and $g_{n_D m}(n)$.

4.3 Initial simulations of performance

To gain an understanding of the behaviour of filters optimised in the time domain, initial simulations based on a 2-source endfire array have been carried out. Point source free-field Green functions have been used to calculate the frequency response between each source of the array and the control points. These frequency responses have then been inverse Fourier transformed to obtain the impulse responses from each source to each control point, using $J = 1024$ points at a sampling frequency $f_S = 20$ kHz. To obtain the set of superdirective filters, Eq. 4.13 has been used with a value of $\xi = 0.5$. The result of these simulations is compared with that obtained with filters optimised using the frequency domain method [67]. The control geometry used in the simulations is shown in Fig. 4.2, and has 1 bright control point and 47 dark control points, so that the two source array ideally creates a hypercardioid-like radiation pattern when the array is superdirective [52]. The bright zone pressure target is formed by the impulse response of the source which is closest to the bright control point.

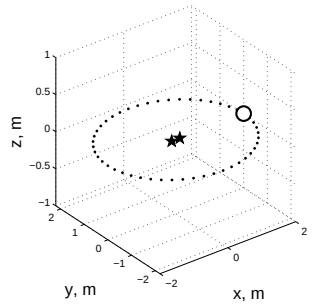


Figure 4.2: Control geometry used to create the filters for a 2 source endfire array. Closed dots symbolise dark zone control points and the open circle symbolises the bright control point. The stars symbolise the 2 sources of the array.

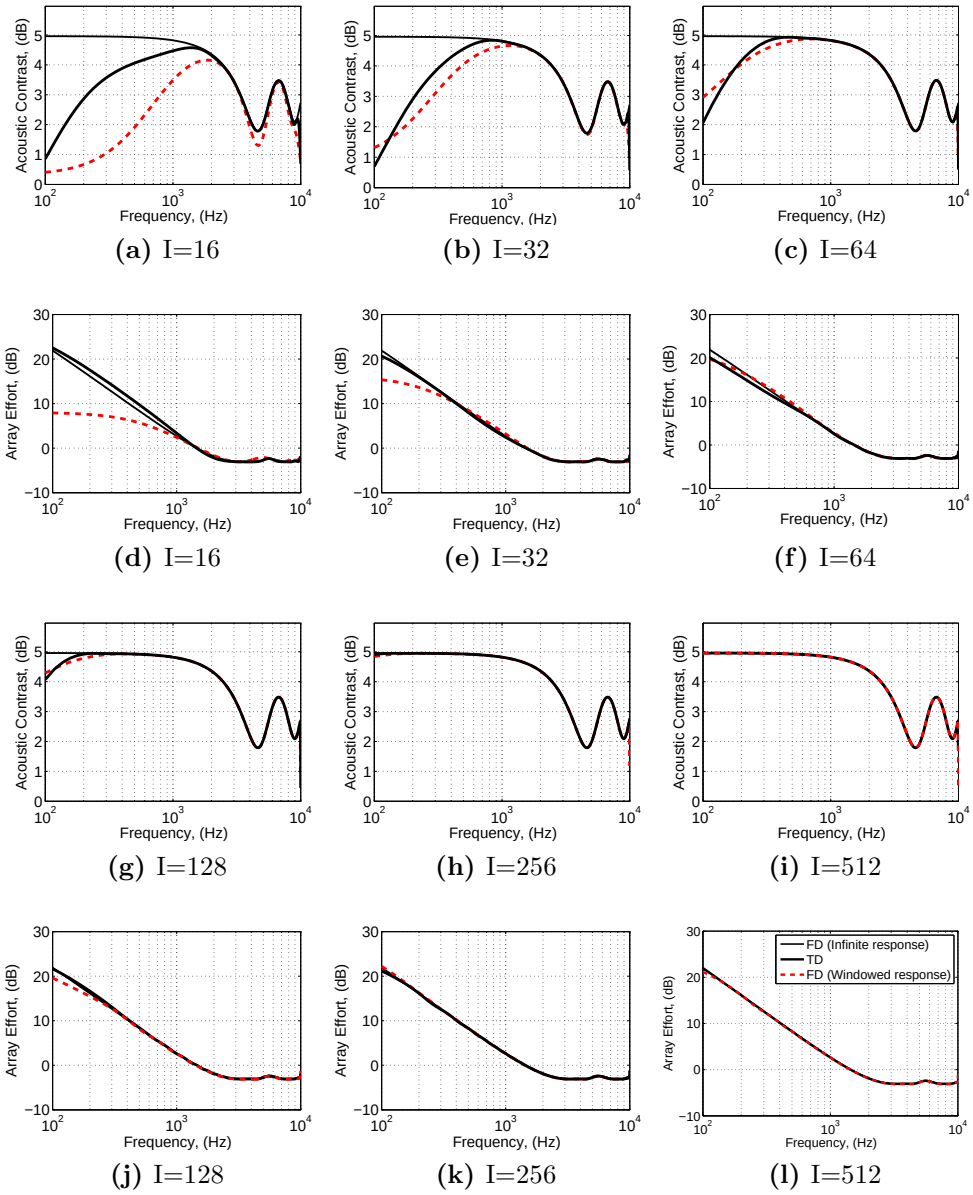
4.3.1 Effect of filter length

Fig. 4.3 shows the effect of the filter length on the acoustic contrast and the array effort for the filters optimised in the time domain. The results are compared with those obtained for a set of filters optimised in the frequency domain with $I = 1024$ frequency bins, and for the set of time-domain filters calculated by inverse Fourier transformation of the frequency domain optimised filter responses and subsequent windowing to the length of the corresponding time-domain filters.

For filters with $I = 16$ coefficients, the performance of the filters formulated in the time domain is much greater than that when they are formulated in the frequency domain and then inverse Fourier transformed and windowed to a length of I=16. As the control filters increase in length, the filters created in both domains are more able to control the low frequencies. The performance of the time domain filters is, however, better than the frequency domain filters for lengths shorter than 128 coefficients. For filters longer than 128, both the filters created in the time domain and in the frequency domain offer a similar performance to that of the optimal filter.

The reason why the short filters created in the time domain outperform those created in the frequency domain is that the time domain optimisation gives the direct least squares solution for a given number of coefficients I, which is not provided by a windowed version of the optimum infinite impulse response obtained using the frequency domain design method. This can also be observed in the array effort, which is reduced in the filters created in the frequency domain and then windowed, as the non-optimal truncation of the filter impulse responses re-

duces the low frequency control. This same effect has previously been observed in practical realisations of cross talk cancellation systems [142], for the equalisation of loudspeaker responses [73] and in microphone beamforming applications [143].



(a) I=16 (b) I=32 (c) I=64

(d) I=16 (e) I=32 (f) I=64

(g) I=128 (h) I=256 (i) I=512

(j) I=128 (k) I=256 (l) I=512

Figure 4.3: Acoustic contrast and array effort for a two source endfire array, with filters created using a time domain optimisation, "TD", or a frequency domain optimisation with windowing to the required length, "FD (windowed response)", all of them when ξ=0.5 in Eq. 4.10 and in the equivalent frequency domain formulation. The thin solid line shows the performance of the unconstrained filter calculated in the frequency domain, "FD (Infinite response)".

The condition number of a matrix, κ, is the difference between the smallest and the largest eigenvalues of such matrix, and indicates the sensitivity of the inverse to random perturbations in the elements of the matrix. The condition number

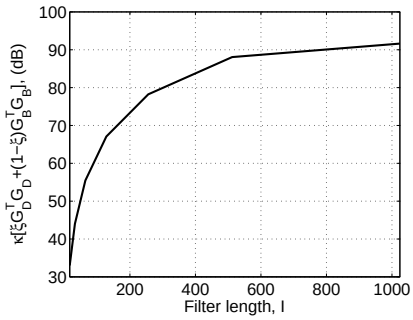


Figure 4.4: Condition number with respect to the inversion of the matrix $\xi\mathbf{G}_D^T\mathbf{G}_D + (1-\xi)\mathbf{G}_B^T\mathbf{G}_B$ plotted against the number of filters coefficients, when $\xi = 0.5$. The plant of the impulse responses is characterised with J=1024 coefficients.

of the matrix $\xi\mathbf{G}_D^T\mathbf{G}_D + (1-\xi)\mathbf{G}_B^T\mathbf{G}$ thus serves as indicator of how robust the filters are with respect to errors and mismatches in the impulse responses of the sources. The effect of filter length on this condition number is shown Fig. 4.4. As expected, the condition number increases with filter length as the dimensions of the matrix increase. The condition number grows faster when increasing the filter length for short filters, but it stabilises for filters greater than I =512 coefficients. This means that although employing longer filters is predicted to produce a higher directional performance, in the presence of errors between the loudspeakers and control microphones this may not be the case.

4.3.2 Delay quantisation error

The sampling frequency, f_S, used by the system plays an important role in the directional response at low frequency. Figure 4.5 shows the acoustic contrast performance of the array described above, optimised using the time domain method, for different sample rates. The length in time of the response used to model $\mathbf{G}$ has been kept constant at 0.06 s for these simulations, which is equivalent to a propagation distance of 20 m, whilst the control filters have been created with responses 0.02 s long. The drop in the acoustic contrast at low frequency does not just depend on how low the sample rate is, but on the actual values of the sample rate. The condition number, with respect to the inversion of the matrix $\xi\mathbf{G}_D^T\mathbf{G}_D + (1-\xi)\mathbf{G}_B^T\mathbf{G}_B$, has also been calculated for these filters. The condition number peaks at certain values of the sampling frequency, which correspond to the values at which the contrast at low frequency is maximum. The reason for this increase in the acoustic contrast is due to the time sampling process used in obtaining the impulse responses, which due to the discrete time resolution, can cause the obtained impulse responses to be sync functions rather than delta func-

tions. This introduces small mismatches in $\mathbf{G}_B$ and $\mathbf{G}_D$, which are also known as *delay quantisation errors* [144], and depend on the sampling frequency which is used. This phenomena has previously been studied in microphone beamforming [145], where it has also been shown that delay quantisation error repeats periodically with sampling frequency.

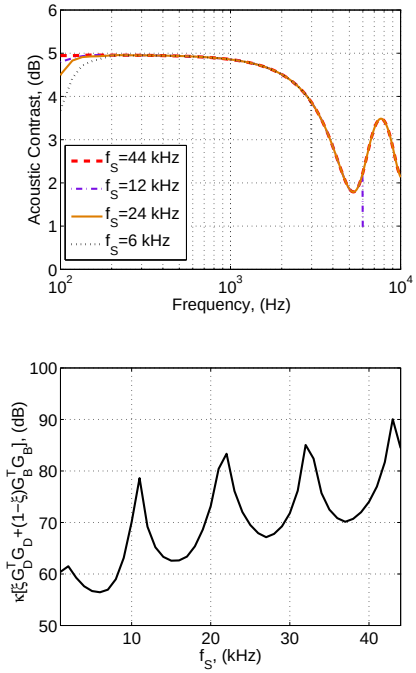


Figure 4.5: Acoustic contrast (upper plot) and condition number with respect to the inversion of the matrix $\xi\mathbf{G}_D^T\mathbf{G}_D + (1-\xi)\mathbf{G}_B^T\mathbf{G}_B$ (lower plot) with $\xi = 0.5$ as a function of the sampling frequency, f_S.

4.4 Simulated performance of a practical array

In this section, the performance of the time domain optimisation method is analysed using the transfer responses measured from a practical array. This array is designed to be placed close to the TV, hence enhancing the audio of the TV in the bright zone, particularly for hard of hearing listeners [53, 95, 146]. The array is shown in Fig. 4.6, and consists of a 4×8 sources planar array, in which each column is driven in phase as a single radiator, reducing the diffuse field energy input by diminishing reflections from ceiling and floor [95]. Each individual source of the array is placed in a phase-shift cabinet, which leads to a hypercardioid-like directivity pattern, further reducing the diffuse field energy injection[104].

In order to quantify the radiation of the array, the control geometry shown in

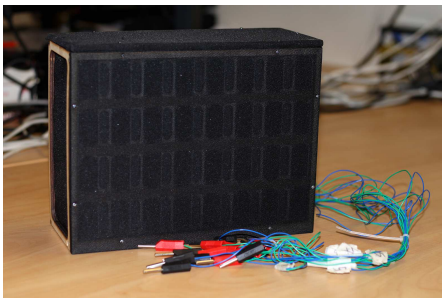


Figure 4.6: The array designed to increase speech intelligibility [95].

Fig. 4.7 has been used, and the transfer responses from each radiator to each microphone have been measured. In order to create superdirective filters and control the soundfield, the geometry has been divided into a bright zone which extends from 0° to 15°, and a dark zone extending between 22.5° and 352.5°, so that the control geometry has a total of 48 microphones. This definition of the dark zone prevents radiation towards the positions where normal hearing listeners may be seated and also to other directions, which otherwise would increase the reverberant pressure and limit the performance of the array. The impulse responses have been obtained via inverse Fourier transformation of the measured transfer responses, leading to responses with J =512 coefficients. These have been used to predict the performance of the array using different sets of control filters by calculating the acoustic contrast in the frequency domain.

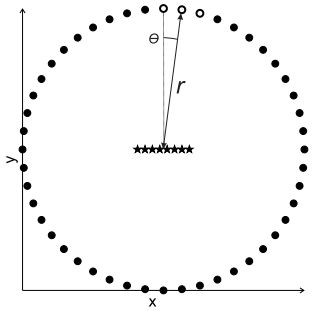


Figure 4.7: 2D control geometry used for creating the filters and for measuring the directive performance of the array. Black dots represent the dark zone control microphones and open white circles represent the bright zone control microphones. The stars represent the 8 sources of the array, and r the observation distance, which is of 2 m.

4.4.1 Frequency dependent regularisation

The formulation described in Section 4.2.2 has been used to predict the performance of the array in the free field. Weighted least squares filters have been

created using I=512 samples to minimise the squared error according to Eq. 4.17 with $\xi = 0.5$ in the control geometry of Fig. 4.7. The target pressure, p_T, has been set to be the impulse responses from the 4-th source of the array to each of the bright zone microphones. In order to assess the performance of the filters created in the time domain, their performance is compared with that obtained using filters optimised in the frequency domain and then windowed to have the same number of coefficients. The array effort in Eq. 4.18 is limited to be below 6 dB to avoid overdriving the sources of the array by selecting the filter $\mathbf{b}$ to be a low pass filters of 128 coefficients with a cut off frequency of 3 kHz. The frequency response of $\mathbf{b}$ is shown in the top graph of Fig. 4.8, where it is compared with the regularisation needed in the frequency domain method to ensure a maximum array effort of 6 dB.

Predictions of acoustic contrast and array effort are also shown in Fig. 4.8, in which the performance of the time domain and frequency domain optimised filters with limited array effort is compared. The filters created in the frequency domain have been inverse Fourier transformed and then windowed to have the same length as those created in the time domain. The acoustic contrast performance for the filters created in both domains using a limited array effort is similar. The acoustic contrast grows from about 10 dB at 1 kHz to about 25 dB at 5.5 kHz as the array aperture becomes more comparable with the wavelength of radiation. The array effort is smaller for the time domain optimisation between 1 kHz and 3 kHz. If there is no constraint on the amount of energy used, a slightly higher acoustic contrast, $\beta = 0$, about 1 dB at 1 kHz, is obtained. This is, however, impractical to use because of the danger of over-driving the array sources since the array effort is large and also leads to filters with a lower sound quality due to a large amount of pre-ringing.

4.4.2 Combined solution response

The effect of the parameter ξ in Eq. 4.17 is now investigated, by selecting ξ to be 0.9 or 0.5. In this case, the pressure target, $\mathbf{p}_T$, has also been set to be the impulse responses from the 4-th source of the array to each of the bright zone microphones. Control with $\xi = 0.9$ represents a case in which the contrast between the two control zones is maximised and less importance is given to the minimisation of the error in the bright zone. Conversely, the control with $\xi = 0.5$ represents a comprise solution, in which both the error minimisation in the bright zone and the acoustic contrast maximisation are taken into account when creat-

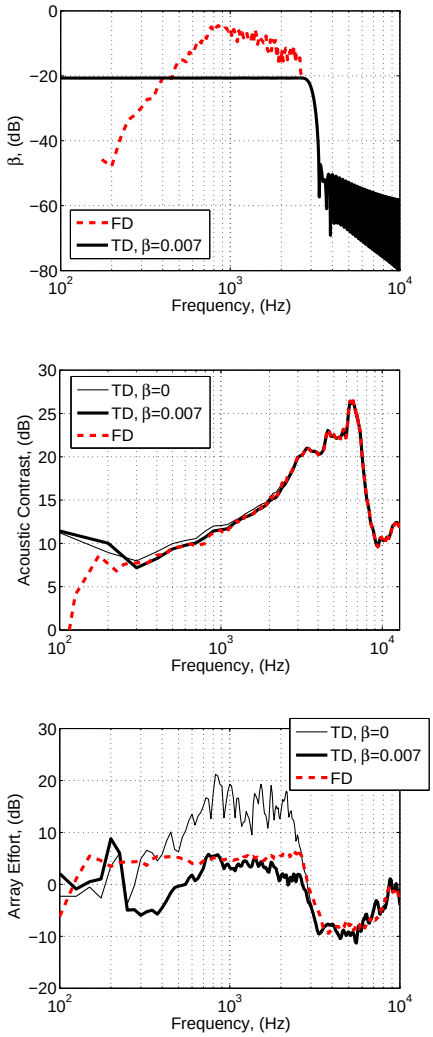


Figure 4.8: The upper left graph shows the regularisation against frequency that has been applied to limit the array effort to be below 6 dB in the time domain, "TD", using the filter $\mathbf{b}$ and $\beta = 0.007$ and in the frequency domain, "FD", one frequency at a time. The medium and lower left graphs show the acoustic contrast and array effort for the array, when the array has 512 coefficient filters, calculated on the time domain or in the frequency domain with windowing.

ing the optimal filters.

The results for the combined solution in the time domain are shown, for the two values of ξ, in Fig. 4.9, for the acoustic contrast performance, array effort and frequency responses of the three bright zone microphones. In the case where $\xi = 0.9$, the acoustic contrast is about 2-3 dB larger between 6 kHz and 7 kHz than for the case where $\xi = 0.5$. The greater directional performance with $\xi = 0.9$ is obtained at the expense of producing a less uniform spatial response in the bright zone, as can be seen in the right hand side plot of Fig. 4.9, which has been previously observed with the combined solution algorithm [67, 69]. From the lower right hand side plot it can be seen that with $\xi = 0.9$ differences of more than 10 dB between the three microphones in the bright zone are present

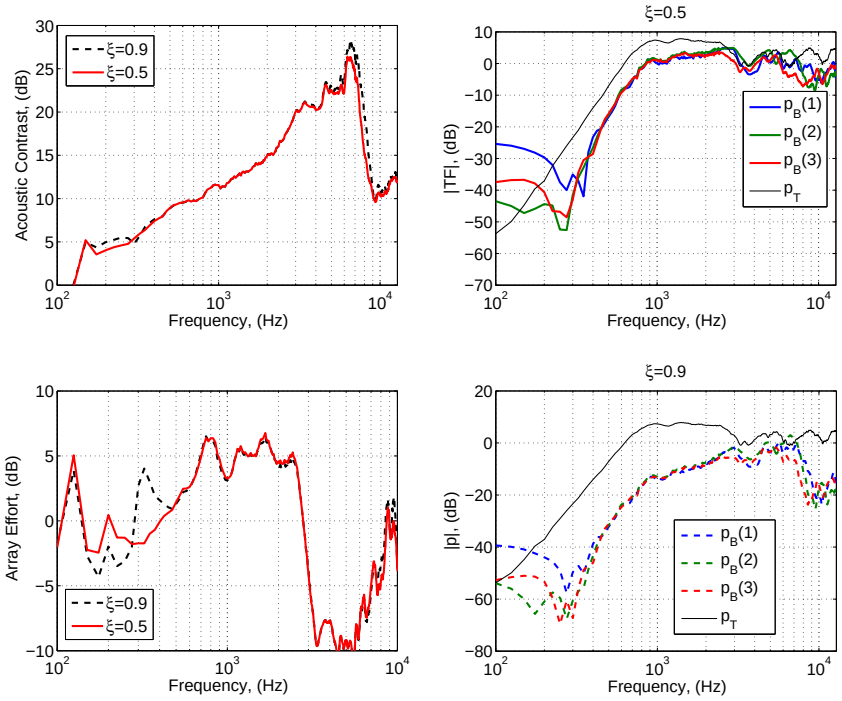


Figure 4.9: The left hand plots show the acoustic contrast and array effort for the array obtained with different values of ξ in the combined solution o Eq. 4.17. The right hand plots show the modulus of the frequency response of the three bright zone microphones for the two values of ξ, in comparison with the pressure target, p_T.

between 6 kHz and 8 kHz. Conversely, from the upper right hand side plot it can be seen that the differences in the bright zone are limited to be below 5 dB. This result is similar to that obtained by comparing the least squares method with the acoustic contrast maximisation method in the frequency domain, where the acoustic contrast maximisation method is able to obtain a higher contrast at the expense of a more uneven reproduction in the listening zone [53]. The array effort needed for the filters is similar for both values of ξ. In this case the larger acoustic contrast obtained by selecting ξ =0.9 is given at frequencies at which the condition number of the matrix is low, and hence the array does not need a larger electrical power to obtain that solution.

4.4.3 Minimising processing delay

One requirement of the practical array considered here is that, in order to maximise the speech intelligibility, it should radiate at the same time instant as the TV sound system. If the array is directly connected to the output of the TV, the processing time needed to convolve the input signal with the filters of the array has to be minimised, since otherwise a comb-filtering effect will be present in the bright zone audio. This effect has previously been studied for the processing introduced by wearable hearing aids in hearing impaired listeners [147]. In this

case, the acceptable delays were lower than 5 ms, but it was observed that in order to obtain the clearest speech the processing delay had to be lower than 1 ms at 4 kHz [147]. When the filters are created using the frequency domain optimisation method, the filters are non-causal, and need a modelling delay of $I/2 + 1$ samples to be made causal.

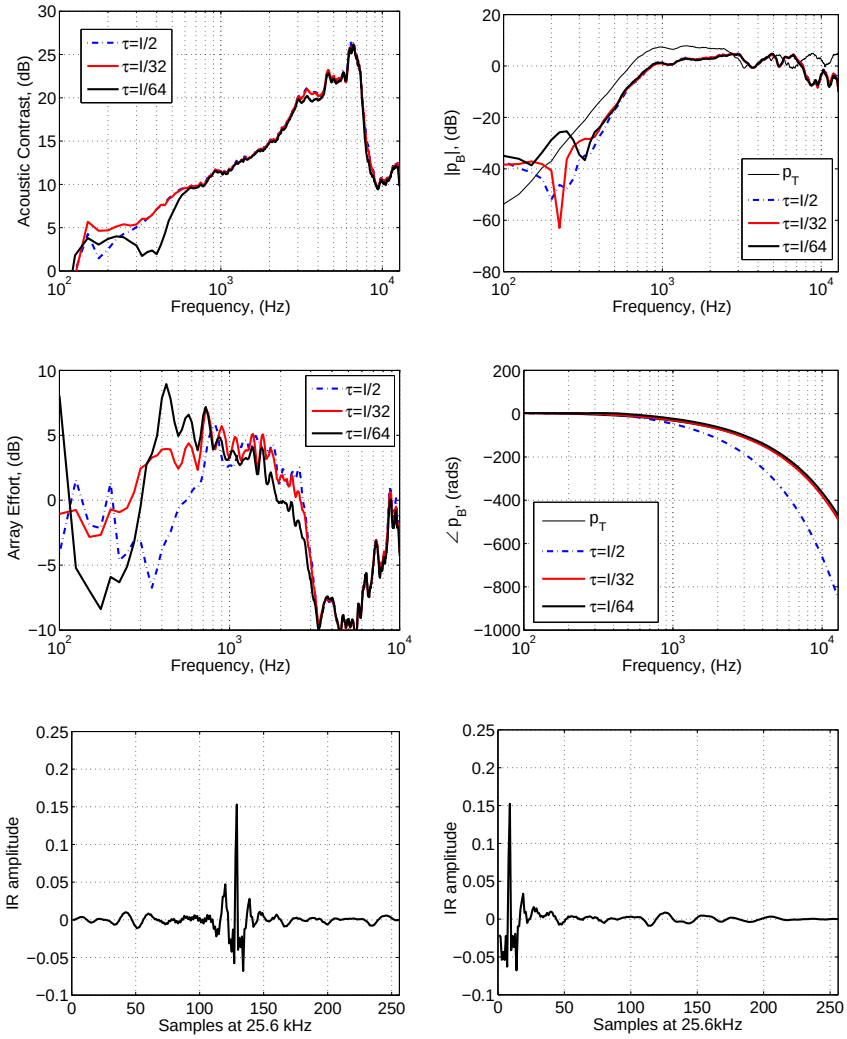


Figure 4.10: Performance results for different modelling delays, τ. The upper plots show the acoustic contrast and the modulus of the frequency response at the centre microphone of the listening zone of the array, compared against the pressure target, p_T. The middle plots show the array effort needed by the filters and the phase of the frequency responses, and the lower plots show the impulse responses of the 4-th source of the array when using modelling delays of $I/2$ (left hand side plot) and $I/32$ (right hand side plot) taps.

By using the time domain formulation, it is also possible to select the amount of modelling delay desired for the array filters, so that longer filters that at the same time satisfy the acceptable delays as given by [147] can be created. Previous

studies in the literature suggest that the modelling delay should be about $I/2$ to accommodate for the propagation time of each source and also to include in their responses part of the acausal components of the least-squares inverse [148]. Fig. 4.10 presents the results for sets of filters of I=256 coefficients with modelling delays of $I/64$, $I/32$ and $I/2$ coefficients, which corresponds to 0.15, 0.31 and 5 ms at $f_S = 25.6$ kHz respectively, with ξ in Eq. 4.17 being equal to 0.5. The acoustic contrast performance is similar for the three sets of filters. For the set with a modelling delay of $I/64$ the performance is slightly lower below 600 Hz. The frequency response in the bright zone is also quite similar for the three filter lengths. The predictions of acoustic contrast and array effort below 600 Hz are believed to have low accuracy, as the frequency response of the sources rolls off below 600 Hz. It is also possible to observe how the phase of the frequency response is closer to that of the target pressure when using short modelling delays. The impulse responses of the filter of the 4-th source of the array for modelling delays of $I/2$ and $I/32$ are also shown in Fig. 4.10. Filters can thus be directly optimised in the time domain with short modelling delays without a significant effect on frequency domain performance.

4.5 Conclusion

A formulation has been developed which allows the design of the filters for personal audio problems in the time domain. Using this theory, the "combined solution" algorithm of Chang and Jacobsen [67] has been formulated to minimise the error of an impulse-response based cost function. This formulation has then been employed to create filters for a personal audio device designed to increase the speech intelligibility of hearing impaired TV viewers.

In order to gain an understanding of how the time domain formulation works, initial simulations have been performed with a two-source endfire array. These results have shown that the performance of the filters optimised in the time domain can be larger than the windowed filters optimised in the frequency domain, when the filter order is low. This is because the time domain formulation directly calculates the optimal filter impulse responses, whereas the frequency domain formulation requires truncation.

Filters of different orders have also been design for a practical planar array. In order to limit the magnitude response of such filters, and not damage the sources

of the array, frequency dependent regularisation has been applied in the form of a low pass filter.This regularisation limits the maximum power used by the array filters to be below 6 dB. The combined solution formulation can be used to give a directivity which maximises the acoustic contrast between both control zones, hence overcoming the problems that have been previously observed when maximising the acoustic contrast maximisation in the time domain [76].

In order to provide a clear reproduction of the sound in the bright zone and also minimise the interference with the dark zone audio, the application for the array to be used as a complementary TV sound system requires the minimisation of the modelling delay of the array filters. A solution which has been presented is that of using filters with a short modelling delay, which has shown to give frequency and directional responses similar to those obtained using a typical modelling delay of about $I/2$ samples.

Chapter 5

The effect of reverberation on personal audio devices

When an acoustical radiator is placed inside an enclosed environment, the reflection of the radiated pressure with the environment surfaces generates an extra pressure component, generally known as *reverberant* pressure. This extra pressure component adds to the pressure generated by a radiator, decreasing its directional response, and introducing small variations in its transfer function.

This chapter introduces an analysis of the behaviour of the 1×8 array and the 4×8 array when these are placed inside two reverberant environments, which are an audio listening room with a very low reverberation time, and a classroom with a higher reverberation time. The arrays use least squares inverse filters (LSIF), see Section 3.1.4, with the performance assessed in terms of acoustic contrast. A formulation is first introduced, that allows the acoustic contrast to be estimated in a reverberant environment, based on the free field directivity of the source and the acoustic characteristics of the environment. Finally, the behaviour of both arrays in a reverberant environment is discussed in terms of robustness. It is also shown by means of simulations and measured results how, even without spatial averaging, the reverberation acts similar to regularisation when the room responses are used for creating the filters. This constrains the amount of power used and makes the array robust to source mismatches.

5.1 Reverberant performance

Although the final application of personal audio systems is in a normal, everyday room, the behaviour of these systems has mostly been studied in anechoic conditions. Directional reproduction in a reverberant environment is complicated by

the reflections introduced from the room surfaces, which introduce a high amount of variability in the transfer functions of the source [78]. The performance of a single directional acoustical radiator in a reverberant field was studied by Beranek [82], who derived an expression for the total acoustic field depending on the acoustic characteristics of the room and the radiator. A similar formulation was used to obtain the directional characteristics of a source in a reverberant environment by Druyvesteyn and Garas [40]. Measurements of the performance of a line array in reverberant environments have also been presented [81]. In order to provide a flat response over a certain spatial region, some researches have averaged an ensemble of transfer functions from such region [79], however, these techniques do not give enough accuracy for sound reproduction inside a room. One study has used the modes of the room to define the transfer functions of a source in the room [80], obtaining a good simulated performance. Other recent work has analysed the performance of different transfer response models for creating the filters of a superdirective array in a reverberant environment [83]. It is clear that the directional characteristics of an acoustical radiator are changed when it is placed in a reverberant room, but it is not so clear what the effect of the reverberation is on the robustness of a superdirective beamformer.

The robustness of these radiators is defined in terms of how much the directional response is affected by mismatches in source sensitivity and position. In superdirective antennas, for example, it has been observed how the practical response is far from the theoretical superdirective value if random errors are introduced in the antenna weightings [89]. The characteristics of these errors and how to obtain robust beamforming in their response has been studied for microphone arrays, assuming that the errors have a certain probability factor [90]. Work on loudspeaker arrays has also studied the effect of errors in source sensitivity and position, and how they affect the ratio of mean square pressures between two spatial zones [91, 92]. Generally, these acoustical radiators are made robust to changes in their transfer responses by applying regularization [64, 91], which also controls the power used to drive the array sources. Recent work has suggested that the space-averaged uncertainties introduced into the transfer responses of an array by reverberation can act in a way that is analogous to regularisation [91].

5.2 Experimental procedure

The directive performance of the two arrays described above has been assessed inside an audio listening room [149] and inside a classroom. The listening room has a low reverberation time and an homogeneous distribution of absorption along its surfaces, whilst the classroom has a larger reverberation time and a more irregular distribution of absorptive façades. The arrays were placed in the centre of the room, at a height of 1 m. For the measurement of each source array, these are placed in a 2D circular control zone, formed by 48 microphones, as shown in Fig. 5.1. For the experiments presented here, the observation distance, r, is 2 m. A "bright zone" [46], is defined so that it extends from an angle, θ, of 0 to 15°; whilst the "dark zone" extends from an angle of θ of 22.5 to 352.5°. This control geometry has been hence used for both measuring the transfer functions of the array and to measure the directional performance. Both reverberant environments and the set up used to form the control microphone array are shown in Fig. 5.2.

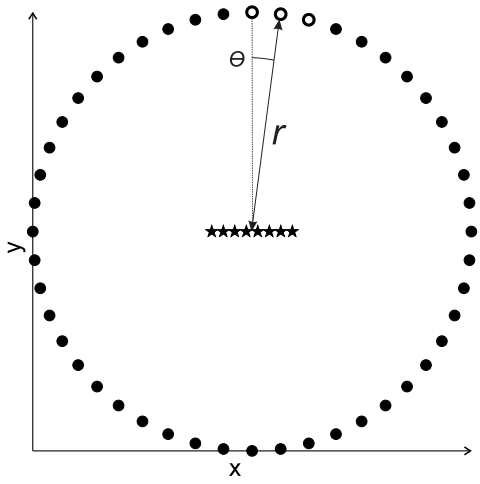


Figure 5.1: 2D control geometry used for creating the filters and for measuring the directive performance of both arrays. Black dots represent the dark zone control microphones and open white circles represent the bright zone control microphones. The stars represent the 8 sources of the array, and r the observation distance.

The reverberation times of both rooms are shown in Fig. 5.3. The reverberation time was measured by recording the impulse response (IR) of the room at 12 different positions according to the standard ISO 3382 [150] and performing backward integration. The IRs were measured using the sine swept technique [151]. The classroom is a considerably reverberant environment, with a reverberation time between 125 Hz and 1 kHz higher than 1 second. The audio room,

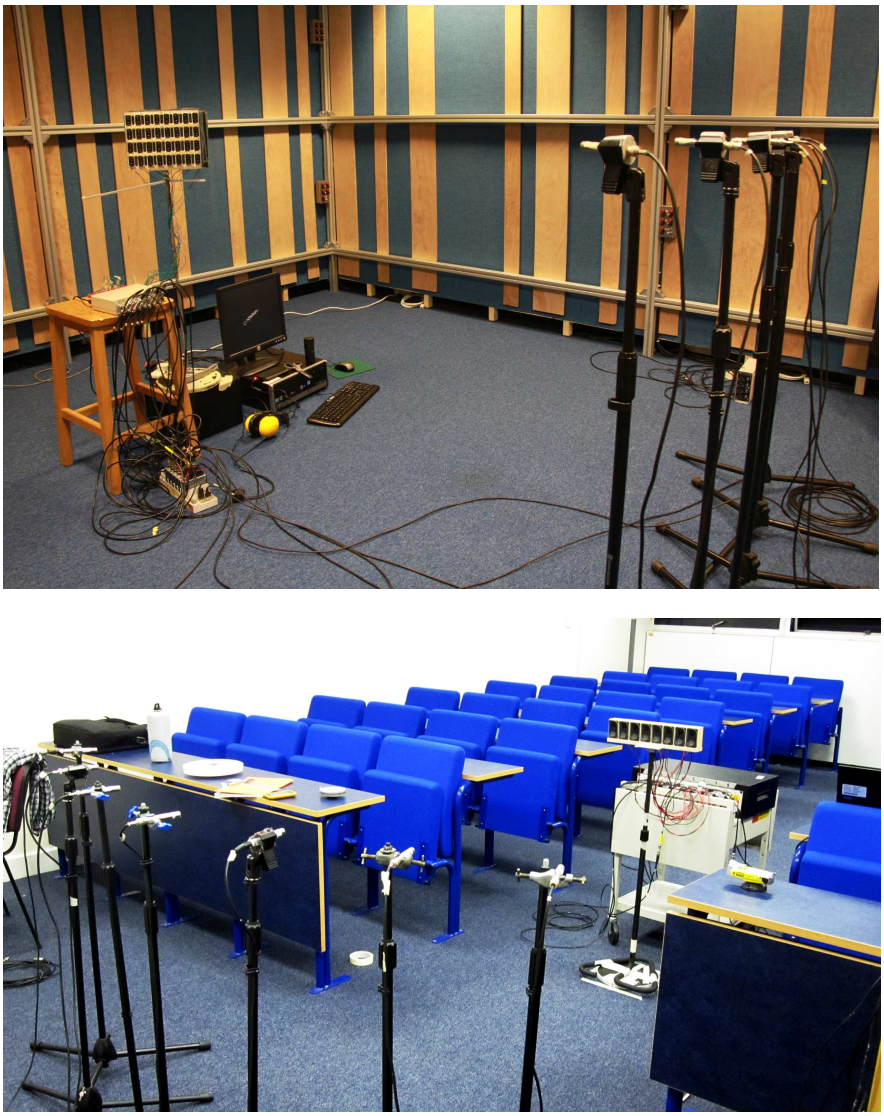


Figure 5.2: Set up for the measurement in the reverberant environments. Audio room (upper plot) and classroom (lower plot).

instead, presents a very low reverberation time, which is of about 0.2 seconds at 1 kHz, as it is intended to be used for testing high quality audio systems. The geometrical data and dimensions of both rooms has been measured and is shown in Table 5.1 together with the Schroeder frequencies of both rooms. Based on this data, the formulation presented in the standard ISO 354 [152] has been used to estimate the absorption of the rooms surfaces. Similar results for the absorption coefficients have been obtained if the absorption is calculated using Eyring's formulation [153]. The average absorption coefficient of the enclosure surfaces, α, is obtained at each frequency by

$$\widehat{\alpha} = \frac{55.3V}{cST}, \tag{5.1}$$

where S represents the surface area of the enclosure walls, V is the volume of the

enclosure T is the reverberation time. The obtained absorption coefficients for both environments are shown in Fig. 5.3. As expected, the estimated absorption coefficient for the audio room is much larger, as the walls of the audio room are covered with curtains.

Table 5.1: Geometrical data and Schroeder frequency for the two environments where the reverberant directivities of the array have been studied.

	V, (m^3)	S, (m^2)	f_{SCH}, (Hz)
Classroom	86.7	115.3	203
Audio Room	73.4	112.2	111

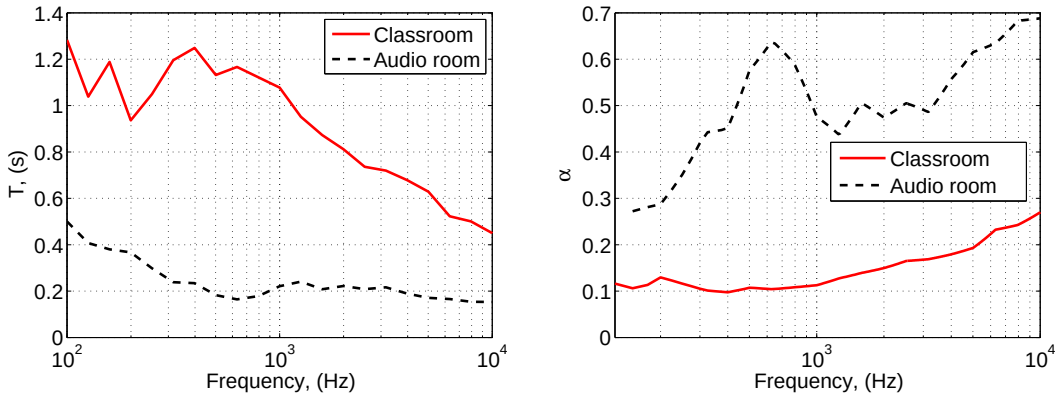


Figure 5.3: Reverberation times (left hand side plot) and mean absorption coefficients $\widehat{\alpha}$ (right hand side plot) for the two environments where the reverberant directivities of the array have been studied.

5.3 Estimation of reverberant directivity

5.3.1 Reverberant directivity formulation

When an acoustical radiator operates inside a reverberant environment, the pressure field it generates will have two components. The first component will be the direct field generated by the source, which is the same as the pressure field created under anechoic conditions. The other component is the reverberant pressure, which is due to all the reflections from the enclosure. In order to estimate the pressure contribution that a source makes to the reverberant field, a knowledge of the acoustic power that it produces is required. If the source is not a point monopole source, the total acoustic power radiated can be determined by integrating the intensities over the surface of a sphere [82]. Assuming the pressure is measured in the far field, the total radiated acoustic power under free field

conditions is given by

$$W = \frac{r^2}{\rho_0 c_0} \int_0^{2\pi} \int_{-pi/2}^{\pi/2} |p(\phi, \theta, r)|^2 |\cos\theta| d\theta d\phi \tag{5.2}$$

where r is the radius of the sphere, $c_0 = 343$ ms^{-1} is the speed of sound in air, p is the acoustic pressure and ϕ and θ are the polar and azimuthal angles. If the pressure is only known at a discrete set of control points, however, the integral can be approximated by the summation

$$W = \frac{r^2}{\rho_0 c_0} \sum_{m=1}^{2\pi/\Delta_\theta} \sum_{n=1}^{\pi/\Delta_\phi} |p(\phi_n, \theta_m, r)|^2 |\cos\theta_n| \Delta_\theta \Delta_\phi, \tag{5.3}$$

where $\Delta_\theta = 2\pi/N_H$ and $\Delta_\phi = \pi/N_V$ represent the angle in radians between each horizontal and vertical measurement point, where N_H is the number of horizontal measurements and N_V is the number of vertical measurements.

Under steady-state conditions the power input of a source into a diffuse field is balanced by the absorption of the room walls. The space-average squared reverberant pressure, $\langle |p_R|^2 \rangle$, is related to the power radiated by the source, W, by [82]

$$\left\langle |p_R|^2 \right\rangle = \frac{4\rho_0 c_0}{R} W, \tag{5.4}$$

with

$$R = \frac{S\widehat{\alpha}}{1 - \widehat{\alpha}}, \tag{5.5}$$

where $\langle\rangle$ denotes spatial averaging. This equation allows us to calculate the reverberant pressure that any source produces inside a reverberant environment, once the radiated acoustic power and the absorptive characteristics of the room are known. The radiated power in a diffuse field is assumed to be the same as that radiated into a free space, as originally shown for a monopole [154]. Once the reverberant pressure is calculated, this can be combined with the direct pressures from the source, allowing the calculation of the acoustic contrast. The spatially averaged mean square pressure at each control point at a distance r is then written

$$\left\langle |p_{B_R}(n_b, r)| \right\rangle^2 = |\mathbf{Z}_B(n_b, r)\mathbf{q}|^2 + \left\langle |p_R|^2 \right\rangle, \tag{5.6}$$

and

$$\left\langle |p_{D_R}(n_d, r)|^2 \right\rangle = |\mathbf{Z}_D(n_d, r)\mathbf{q}|^2 + \left\langle |p_R|^2 \right\rangle, \tag{5.7}$$

where $n_b = 1 : N_B$ and $n_d = 1 : N_D$ refer to the individual microphones of each of the control zones.

5.3.2 Experimental results

In order to check the validity of the above formulation, its accuracy is compared with measurements of contrast in normal environments. The mean square reverberant pressure in each room is calculated from the power output of the array, using Eq. 5.4. The performance that both arrays achieve in the rooms using least squares filters calculated with free field transfer responses has been measured. The performance has also been measured in an anechoic chamber along with the horizontal and vertical transfer responses, which have been sampled at 48 points horizontally and vertically. Values at coordinates different from $(\theta_1 = 0°, \phi_1 = 0°)$, have been linearly interpolated, so that each value at $\phi_n \neq \phi_1$ is equal to the values of the horizontal measurement slide, θ_1, normalized by ϕ_n/ϕ_1. This leads to a spherical measurement geometry consisting of a total of 1106 measurement points, which is shown in Fig. 5.4. The interpolation is of limited accuracy since the size of the sphere is large compared with λ for a certain range of frequencies. By multiplying the transfer responses with the same set of filters as used for the performance measurements, the radiated power of the source can be estimated.

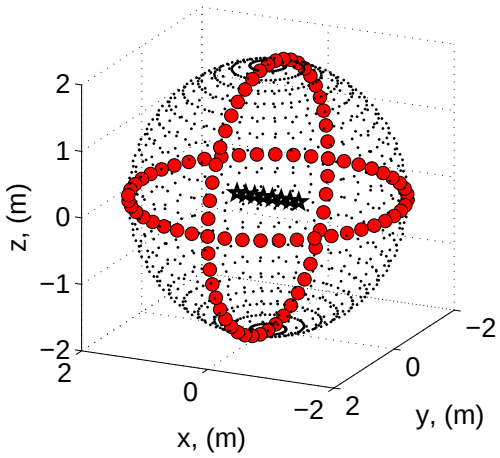


Figure 5.4: Measurement geometry used to estimate the power radiated by the two arrays. The closed circles represent the original measured control points and the dots represent the control points which have been obtained by interpolation of the original measured points. The stars represent the 8 sources of the array.

The plots of Fig. 5.5 show the measured acoustic contrast results for the two arrays in the free field and in reverberant conditions, together with the values of the contrast calculated using Eqs. 5.6 and 5.7. The array effort needed by the filters in free field conditions has been limited to be lower than 6 dB. For the 1×8 array, a large reduction in performance is observed when it is placed in the

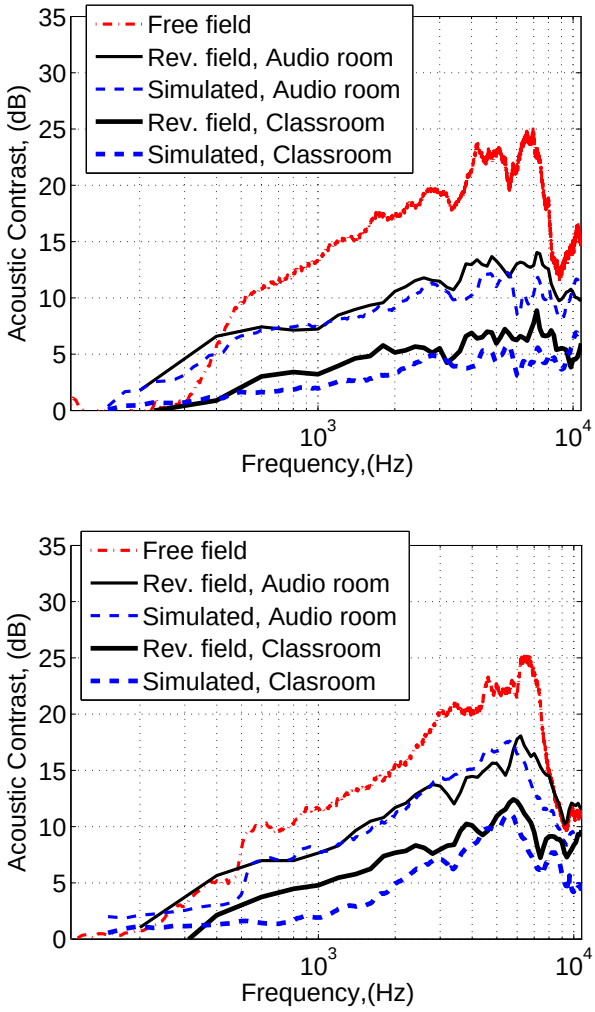


Figure 5.5: Acoustic contrast and array effort results for the 1×8 array (upper plot) and the 4×8 array (lower plot). The free field performance is shown by the dash-dotted lines. The measured reverberant performance is shown by the solid lines and the simulated reverberant performance by the dashed lines, being the audio room results plotted in thin line and the classroom results plotted in thick line.

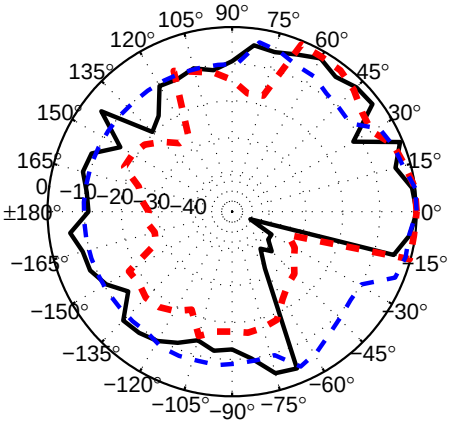
reverberant environments, which at 7 kHz is of about 12 dB when placed in the audio room and of about 18 dB when placed in the classroom. The performance of the 4×8 array is less affected by the reverberant field, due to the greater vertical directivity of the individual sources, even though its horizontal directivity in the free field is slightly less than the 1×8 array. The acoustic contrast for the 4×8 array in the audio room is about 18 dB at 6 kHz, which is around 5 dB higher than the 1×8 array. In the classroom the 4×8 array obtains an acoustic contrast of about 12 dB at 6 kHz, also around 5 dB higher than the 1×8 array.

The predicted reverberant performance of both arrays in the audio room is close to the measured reverberant performance. In this environment, the results show a mean difference between 500 Hz and 8 kHz of 1.2 dB for the case of the 1×8

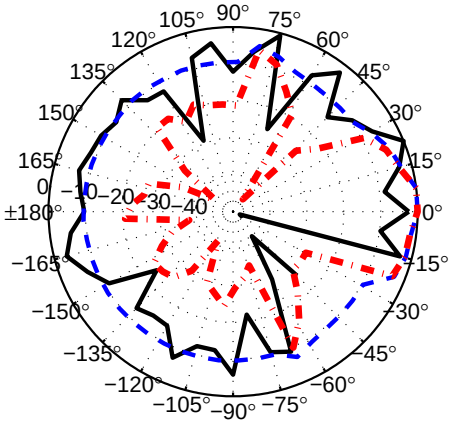
array, and of 0.3 dB for the case of the 4×8 array. The predicted reverberant performance in the classroom is, however, not very close to the measured results, with a mean difference of 1.8 dB for the 1×8 array and 1.7 dB for the 4×8 array. Contrary to the audio room, which has more uniform absorption all along its surfaces, the absorption of the surfaces of the classroom is much more uneven. This contributes to a less diffuse reverberant field, in which measurements of performance at a single point have a greater deviation from the space-averaged result.

The resultant directivities of the 4×8 array are shown in Fig. 5.6 at a number of discrete frequencies, for free field and both measured reverberant and simulated reverberant results in the listening room. The directivities measured in the reverberant environments represent the total magnitude of the pressure which is obtained by the addition of both reverberant and direct pressure fields. In the free field case it is possible to observe how the secondary lobes are small and how the main lobe is pointed towards the listening zone, from 0°to 15°. The measured reverberant results follow the same pattern, but the level of the secondary lobes is much larger, and the nulls in the directional response are not at the same angles as in the free field. This is due to the influence of reflections from the room walls. The predicted reverberant directivities, which represent the space-averaged case, show a reasonable match with that of the measured reverberant directivities, being around halfway between the peaks and nulls of the measured responses.

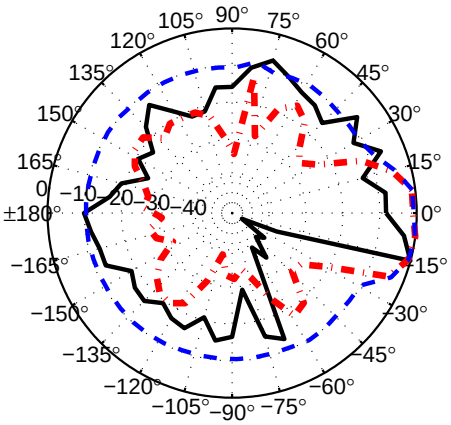
As observed, the directional response of an array is largely affected by the surfaces which surround the device. If the array is placed below a TV, the position of the TV with respect to a rear wall will affect the performance of the array. In order to study how the rear wall and the separation of the array with respect to it affect the directional response, an additional study is presented in Appendix F.



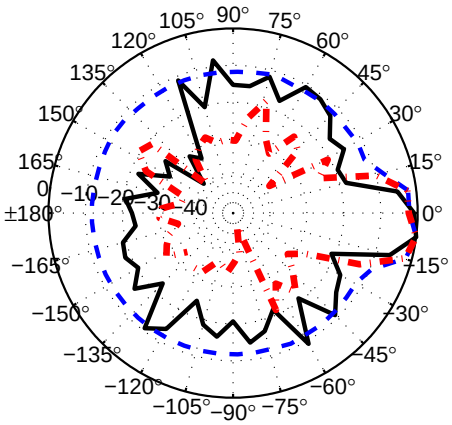
(a) 500 Hz



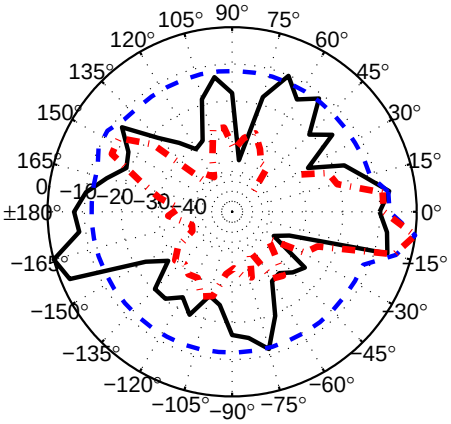
(b) 1 kHz



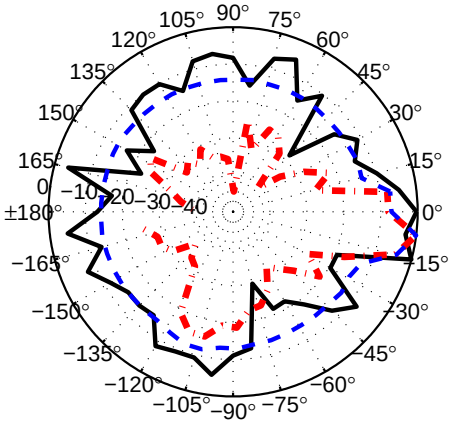
(c) 2 kHz



(d) 4 kHz



(e) 6 kHz



(f) 7.5 kHz

Figure 5.6: Directivities at discrete frequencies for the 4×8 array. The dash-dotted line represents the free field results, the solid line represents the reverberant results measured in the listening room and the dashed line represents the simulated reverberant results for the listening room.

5.4 Effect of reverberation on system robustness

5.4.1 Sensitivity of personal audio systems to mismatches in acoustic environment

Above the Schroeder frequency, the reverberant field is considered diffuse and is made of a large number of contributions from waves travelling in all directions [77], becoming a random function of the excitation frequency and the spatial position at which this pressure is measured. Considering the direct and reverberant components of the pressure field inside a reverberant volume, an element of the transfer matrices $\mathbf{Z}_B$ or $\mathbf{Z}_D$ may thus be written as

$$Z = Z_F + Z_R, \tag{5.8}$$

where Z_F is the deterministic, free field component related to the direct field and Z_R is a reverberant component. The reverberant pressure component amplitude and phase will vary between the response measured at each control point, and can be considered as a random variable whose mean square value is determined by the properties of the room. The reverberant components of the transfer response can thus be viewed as uncertainties in their nominal values, which are given by the direct field components. The matrix of transfer responses $\mathbf{Z}$ can be written to account for these uncertainties as

$$\mathbf{Z} = \mathbf{Z}_0 + \Delta\mathbf{Z} \tag{5.9}$$

where $\mathbf{Z}_0$ is a matrix of nominal, free field values and $\Delta\mathbf{Z}$ is a matrix of the uncertain, diffuse field, components. Supposing that the array is positioned in a room and the measured transfer responses are used to obtain a set of filters, then this is analogous to substituting Eq. 5.9 into Eq. 3.18, where the optimal set of source strengths is now given by

$$\mathbf{q} = \left[\mathbf{Z}_0^H\mathbf{Z}_0 + \Delta\mathbf{Z}^H\mathbf{Z}_0 + \mathbf{Z}_0^H\Delta\mathbf{Z} + \Delta\mathbf{Z}^H\Delta\mathbf{Z}\right]^{-1}(\mathbf{Z}_0 + \Delta\mathbf{Z})^H\mathbf{p}_T. \tag{5.10}$$

If the transfer responses are measured in the same control geometry but in a series of different places inside a room, and the mean value of the estimation is taken, the space-averaged results are obtained. If it is assumed that the uncertainties are uncorrelated with the nominal values, so that

$$< \mathbf{Z}_0\Delta\mathbf{Z} >= \mathbf{0}, \tag{5.11}$$

where $\mathbf{0}$ is the null matrix. The space-averaged properties of the uncertainties are defined as

$$< \Delta\mathbf{Z}^H \Delta\mathbf{Z} >= \Delta_{\mathbf{Z}}, \tag{5.12}$$

so that

$$< \mathbf{Z}^H \mathbf{Z} >= \mathbf{Z}_0^H \mathbf{Z}_0 + \Delta_{\mathbf{Z}}, \tag{5.13}$$

the optimal set of source strengths is given in this case by

$$\mathbf{q} = [\mathbf{Z}^H \mathbf{Z} + \Delta_{\mathbf{Z}}]^{-1} \mathbf{Z}^H \mathbf{p}_T. \tag{5.14}$$

If the uncertainties in the elements of $\mathbf{Z}$ were independent, as they would be well above Schroeder's frequency in a diffuse field, then $\Delta_{\mathbf{Z}}$ is proportional to an identity matrix, which causes the optimal, space-averaged, solution to be regularised even in the absence of any constraint on the effort [91]. This formulation provides an interesting limiting result for a radiator inside a room, which suggests that on average the reverberant field acts as a regularization term.

5.4.2 Simulation results

Instead of calculating the ensemble average of a series of measurements, a simulation is presented using an 8 point monopole array in a single control geometry, as that shown in Fig. 5.1. A single complex element of the transfer matrices which is contaminated with noise, Z_N, can be defined as

$$Z_N = Z_0(1 + SNR\gamma), \tag{5.15}$$

where Z_0 is the free field value, γ is a number from a Gaussian random noise generator and SNR is selected to control the RMS level of the added noise. The control matrices which are contaminated with noise are denoted $\mathbf{Z}_{B_N}$, $\mathbf{Z}_{D_N}$ and $\mathbf{Z}_N$. The noise can be added to the transfer responses *a priori*, before, or *a posteriori*, after, the calculation of the optimal set of source strengths, or at both times.

If the noise is added a priori of the calculation of the optimal set of source strengths, simulating the influence of the diffuse field, however, it should be taken into account that the correlation between different points in a three dimensional field follows a sinc function of kd_n [155], i.e,

$$R = \frac{\sin(kd_n)}{kd_n} = \mathrm{sinc}(kd_n), \tag{5.16}$$

where d_n is the distance between two points in such space. The reverberant sound field created by a single source at two different microphones will be correlated according to Eq. 5.16. By the same principle and using acoustic reciprocity, the correlation between the sound fields generated by two adjacent loudspeakers follows an identical pattern. Since the spacing between the microphones is at least 7 times greater than the spacing between the sources, the correlation between the elements of the uncertainty introduced in the impedance matrices, Eq. 5.12, is dominated by the source spacing. The correlation between the diffuse soundfields created by the sources of an array with $d = 35$ mm are shown in Fig. 5.7. It can be observed that the correlation of the soundfields created by two adjacent sources is still very high at 1 kHz, but it diminishes soon for sources which are more than a unit apart.

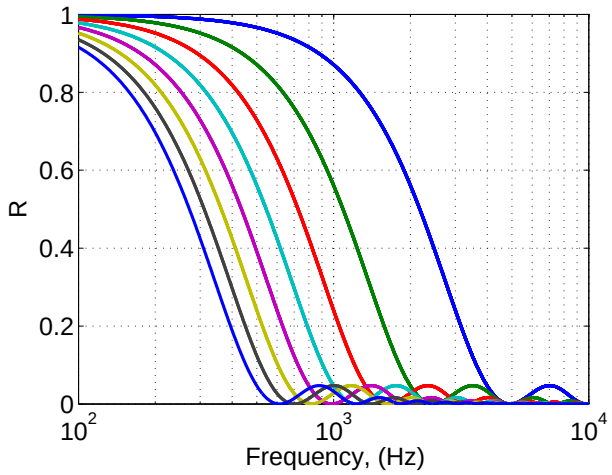


Figure 5.7: Correlation curves between the soundfields created by the 8 sources of a line array in a diffuse field.

The different correlation between each source can be taken into account by defining the matrix of uncertainties, $\Delta_{\mathbf{Z}}$, as

$$\Delta_{\mathbf{Z}} = \mathbf{Z}_N \mathbf{S} \tag{5.17}$$

where $\mathbf{Z}_N$ is an $N \times M$ matrix of uncertainties and $\mathbf{S}$ is an $M \times M$ matrix which causes the uncertainty added to one of the transfer impedances to be a weighted sum according to Eq. 5.16 of the uncertainties added to adjacent elements. According to models of partially correlated noise [156], $\mathbf{S}$ is hence

defined as

$$\mathbf{S} = w_S \begin{bmatrix} \mathrm{sinc}(kd_0) & \mathrm{sinc}(kd_1) & \cdots & \mathrm{sinc}(kd_{M-1}) \\ \mathrm{sinc}(kd_1) & \mathrm{sinc}(kd_0) & & \\ \vdots & & \ddots & \\ \mathrm{sinc}(kd_{M-1}) & & & \mathrm{sinc}(kd_0) \end{bmatrix}, \tag{5.18}$$

where d_0=0 is the distance between a loudspeaker and itself, d_1 is the distance between a loudspeaker and its adjacent, and d_{M-1} is the distance between a loudspeaker and that which is $M-1$ sources apart.The factor w_S is used to normalise the partially correlated obtained uncertainties, so that they have the same power as the original noise, and is given by

$$w_S = \sqrt{\sum_{n=1}^{M} \mathrm{sinc}^2(kd_N)}. \tag{5.19}$$

The optimal set of transfer functions can then be obtained using Eq. 5.10.

If the noise is introduced a posteriori of the calculation of the optimal set of source strengths, simulating mismatches in the transfer responses which are considered to be uncorrelated along the whole frequency band, the acoustic contrast is then given by

$$C = \frac{N_D}{N_B} \frac{\mathbf{q}^H \mathbf{Z}_{B_N}^H \mathbf{Z}_{B_N} \mathbf{q}}{\mathbf{q}^H \mathbf{Z}_{D_N}^H \mathbf{Z}_{D_N} \mathbf{q}}. \tag{5.20}$$

Fig. 5.8 shows the effect of noise in the transfer responses of an 8 point monopole sources array with a control zone as shown in Fig. 5.1. The following set of configurations are assessed:

- A free field case is shown as a reference, where anechoic transfer responses are used to calculate the vector of optimal filters and the acoustic contrast is calculated according to Eq. 3.9.

- "A priori" shows the performance when a single set of random variations are introduced in the transfer responses before the calculation of the optimal vector of source strengths, as in Eq. 5.10, analogously to placing the array in a room, which reduces the acoustic contrast with respect to the free field case. In this case the SNR variable in Eq. 5.15 is selected to include RMS random variations up to 5 times the nominal value of the transfer responses, which is representative of changes in response due to the reverberation. The array effort in this configuration is very similar to that obtained for the "A priori, Avg." case, but with a larger variance.

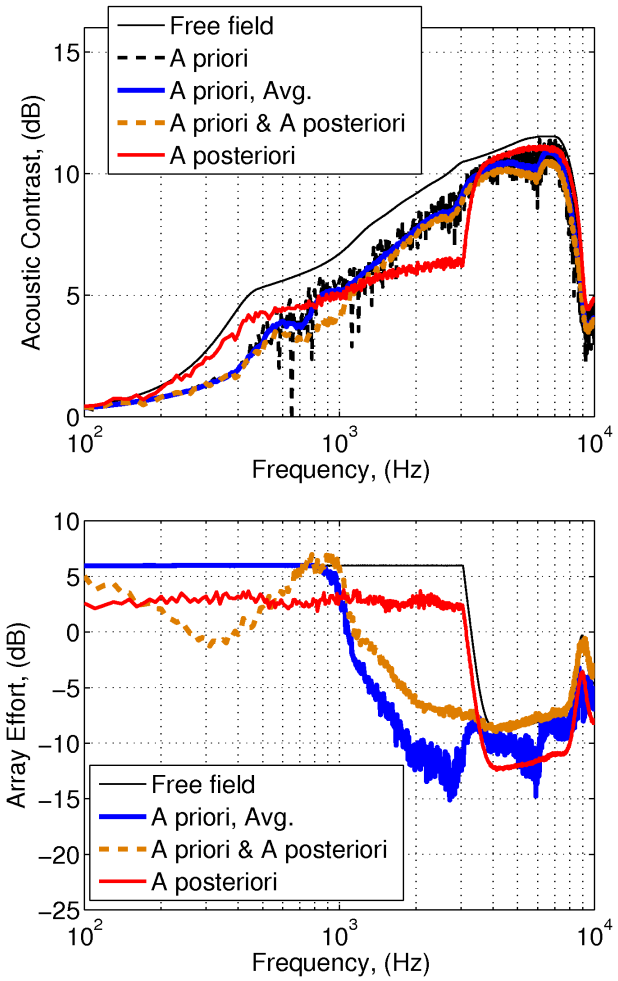


Figure 5.8: Acoustic contrast and array effort for an 8 point monopoles array where noise is added to the transfer responses before the calculation of the optimal set of source strengths, a priori, to simulate the effect of reverberation, or after the calculation of the optimal set of source strengths, a posteriori, to simulate mismatches in the transfer responses.

- "A priori, Avg.", shows the space-averaged result, as defined in Eq. 5.14, where the acoustic contrast is obtained by adding noise before the calculation of the transfer responses, as for the "A priori" case, but with the result averaged over 200 times. The space-averaged solution is quite similar to that obtained in the "A priori" case using a single set of random numbers, but with less variance. Above 800 Hz, the array effort is much reduced due to regularization implied by Eq. 5.14.

- "A posteriori" shows the effect of introducing noise in the transfer responses after the calculation of the optimal, free field, set of source strengths. The acoustic contrast is estimated according to Eq. 5.20, where SNR has been selected to include RMS random variations up to 0.25 times the nominal value of the transfer responses, which correspond to a sensitivity error of

about 1 dB. In this situation both the acoustic contrast and array effort are reduced. The reduction in acoustic contrast is larger at frequencies where the array effort is greater than 0 dB.

- "A priori & A posteriori " shows the space-averaged result of including *a different* random noise before and after calculating the optimal vector of source strengths. At some frequencies the performance is lower than for the " A posteriori" case, due to the random noise introduced before the calculation of the optimal set of source strengths, which has provided a strong regularization. At other frequencies, however, the performance is larger with respect to the "a posteriori case". The a posteriori introduced uncertainties cause the array effort to be slightly greater than that of the "A priori" case above 800 Hz.

At this point, two observations can be made from these results; first, that the space-averaged result is very close to that when only one average is used. Second, random errors introduced in the transfer responses before the calculation of the optimal source strengths vector make the solution robust to perturbations present after the calculation of the optimal vector of source strengths.

5.4.3 Measurements of robustness in a reverberant field

In order to assess the effect of reverberation on the robustness of personal audio devices, the performance of the two 1×8 array and the 4×8 array has been measured in the listening room defined in Section 5.3.2, using different set of filters. The different sets of filters are:

- "NO error": This case refers to filters created with transfer responses measured in free field.

- "1 dB error": In this situation the sensitivity of the 4^{th} source is assumed to be 1 dB higher when calculating the vector of optimal filters from the free field responses.

- "3 dB error": In this situation the sensitivity of the 4^{th} source is assumed to be 3 dB higher when calculating the vector of optimal filters from the free field responses.

- "GF": The filters have been designed using point monopole Green functions to model the transfer responses, as previously performed by [83]. The array effort is limited to be lower than 0 dB, as the point monopole Green functions are quite different from the measured transfer responses, and a

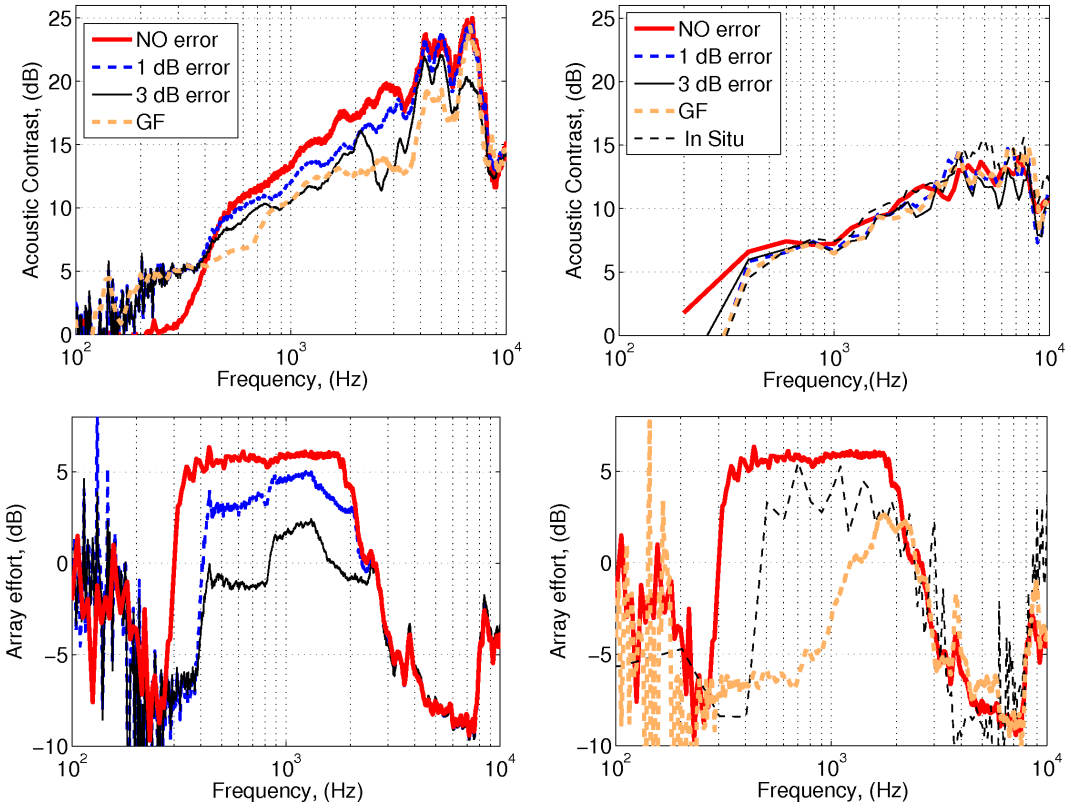


Figure 5.9: Measured results for the free field (left hand side plots) and reverberant (right hand side plots) acoustic contrast and array results for the 1×8 array for different assumed mismatches in the 4^{th} source of the array. "GF" stands for the result where point monopole Green functions have been used as transfer responses to create the array filters. "In Situ" stands for results where the array filters have been calculated with transfer responses measured in the reverberant environment.

large amount of regularization is used to overcome the differences in the response.

- "In Situ": For this case, the transfer responses have been measured in the room, so that the reverberant component is present in Eq. 5.10. Measurements of this kind have also been performed in [86], where the authors suggested that the room reflections could effectively increase the number of degrees of freedom of the array.

The performance measured with these sets of filters is shown for the 1×8 array in Fig. 5.9, and for the 4×8 array in Fig. 5.10. The contrast for the free field is reduced significantly if a mismatch of 1 dB or 3 dB is assumed when calculating the optimal vector of source strengths, or when this is obtained using point monopole Green functions. It can also be observed how the array effort is reduced when mismatches are present in the transfer responses, with the reduction in acoustic contrast and array effort being proportional to the magnitude of the mismatches.

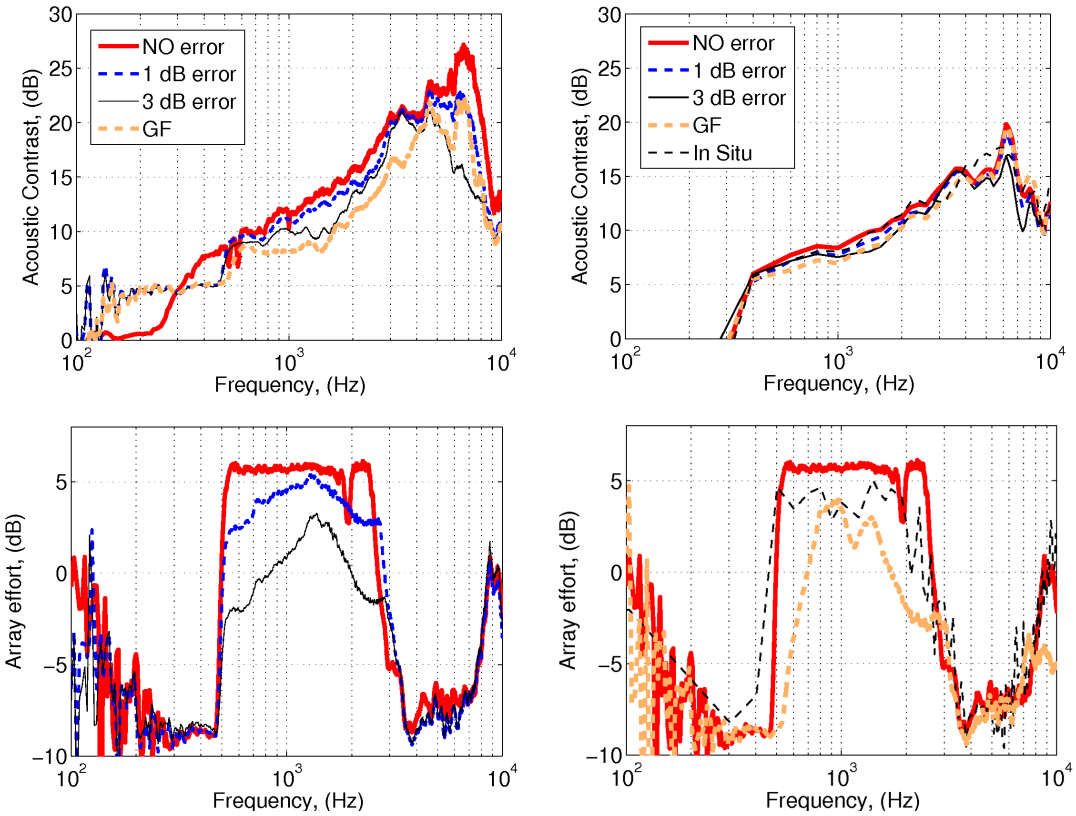


Figure 5.10: Measured results for the free field (left hand side plots) and reverberant (right hand side plots) acoustic contrast and array effort results for the 4×8 array when different mismatches exist in the 4^{th} source of the array. "GF" stands for the result where point monopole Green functions have been used as transfer responses to create the array filters. "In Situ" stands for results where the array filters have been calculated with transfer responses measured in the reverberant environment.

These differences are much smaller when the performance of the different filters is measured under reverberant conditions. For the case of the control geometry used here, the circular arrangement constrains the radiation towards the dark zone, which makes the solution of the "In Situ" filters very similar to that of the filters created based on free field transfer responses, however, slightly greater at some frequencies. Although the array effort is allowed to be up to a maximum of 6 dB, this is however reduced, as the room itself introduces constraints in the effort, which is also observed in the monopole simulations.

Similar effects of noise and reverberation on the performance are also observed in the results for the 4×8 array, shown in Fig. 5.10, where the results obtained in a reverberant field again present a close performance. Alike results are obtained when the performance of both arrays with the different set of filters is estimated inside the classroom. In this case the acoustic contrast performance of the array, as expected, is further reduced from that obtained in the classroom; the 1×8

array obtains a peak of contrast of about 8 dB around 6-8 kHz, whilst that the 4×8 array peaks at 6 kHz with a contrast of about 12 dB. This is much reduced than in the classroom, whilst both arrays obtain peaks of 14 dB and 20 dB respectively. The performance given by all the sets of filters is very similar, as expected from the regularisation introduced by the environment, that in this case is stronger due to the higher amount of reverberation. This can be observed also in the array effort for the "in situ" filters presented by both arrays, which is of about 1 dB on average for both arrays, about 2-4 dB lower than that for the filters created with measured impedances in the listening room.

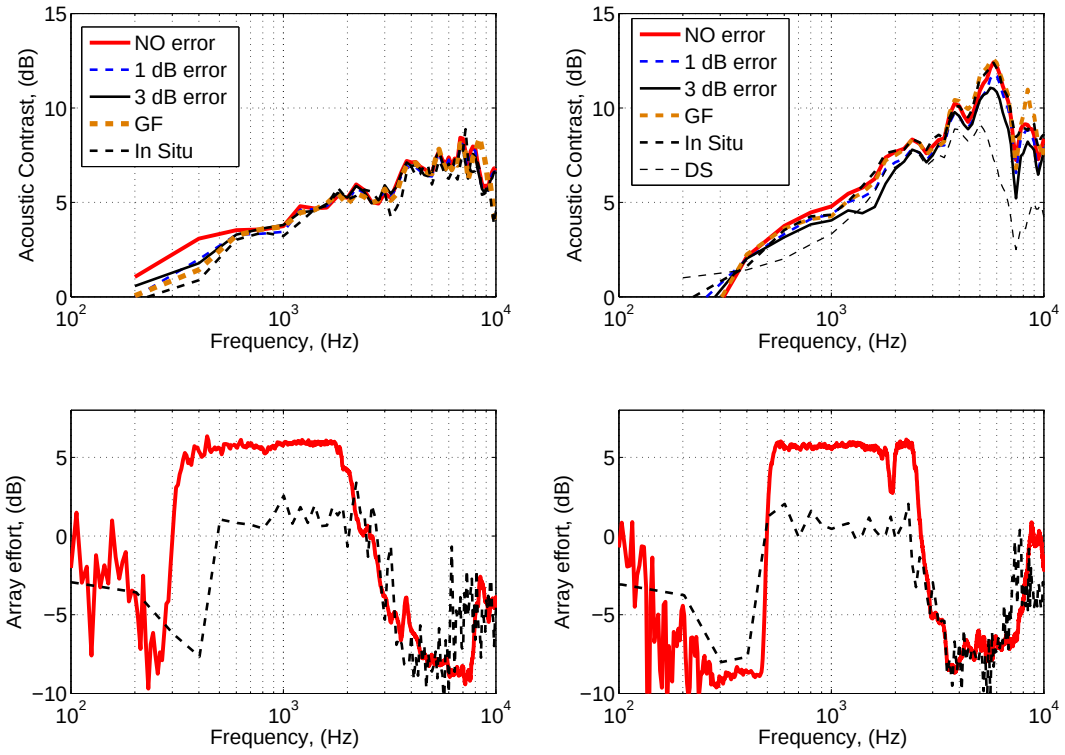


Figure 5.11: Reverberant acoustic contrast and array effort results for the 1×8 (left hand side plots) array and for the 4×8 (right hand side plots), as those presented in Figures 5.9 and 5.10 but measured in the classroom.

The effect that the a priori introduced mismatches due to the reverberation have on the robustness array can also be visualised by looking at the condition number with respect to the inversion of the matrix of transfer impedances. The condition number with respect to the inversion of the matrix $\mathbf{Z}_T^H \mathbf{Z}_T$ is shown for the 1×8 array in Fig. 5.12, being calculated with free-field transfer impedances, transfer impedances measured in the audio room, and transfer impedances measured in the classroom. Due to the absence of regularisation introduced by the environment, the free field results have the higher condition number. When the array

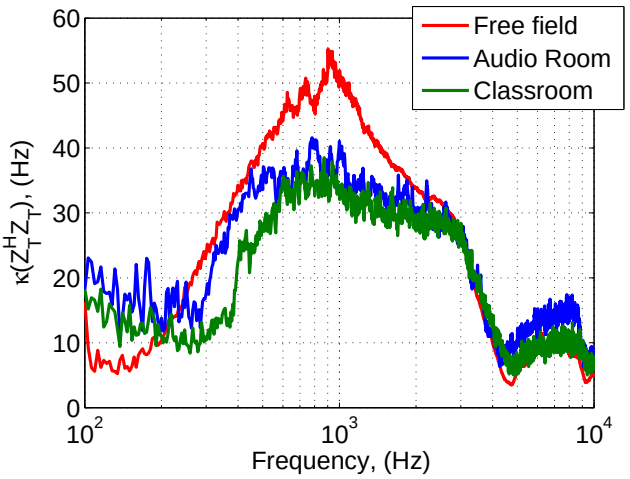


Figure 5.12: Condition number with respect to the inversion of the matrix $\mathbf{Z}_T^H \mathbf{Z}_T$ for the 1×8 array measured inside the three scenarios of study; the anechoic chamber, the audio room and the classroom.

is introduced in the audio room, the condition number decreases, as the transfer impedances matrix is regularised. The condition number is further decreased in the classroom environment. This fact is primarily observed at low frequency, where the condition number in the classroom is reduced for about 100 Hz more than in the audio room.

The fact that the reverberant field is more correlated at low frequency, as shown in Fig. 5.7, may lead to think that the condition number with respect to the inversion of the matrix $\mathbf{Z}_T^H \mathbf{Z}_T$ should increase at low frequency. It can be seen from Fig. 5.12 that, however, the condition number of the inversion decreases when the array is placed in a reverberant environment with respect to the condition number obtained when the array is in free field. An explanation to this fact can be found in Fig. 5.13, where there are shown the coherence functions for the transfer functions of the array at different angles. As observed, large drops in the measured coherences can be observed below 1 kHz which change between each source, which is believed to cause the condition number of the inversion to decrease for reverberant environments.

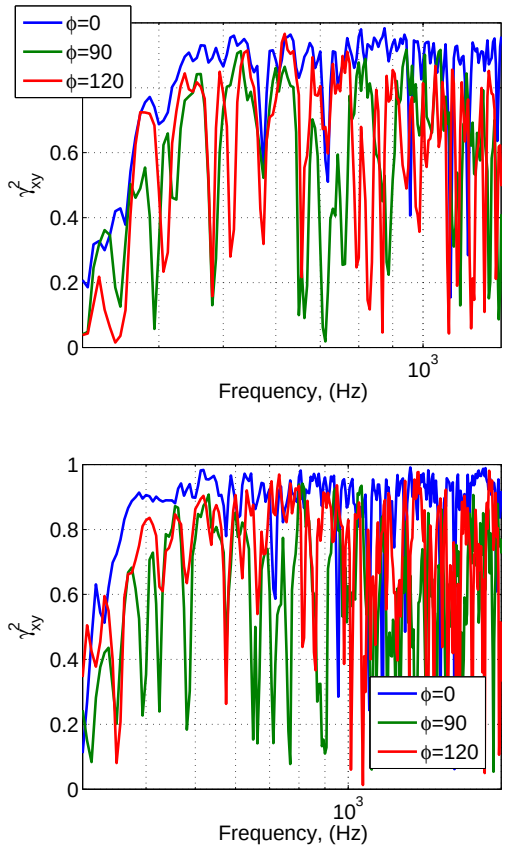


Figure 5.13: Coherences of the reverberant transfer functions of the 1×8 array 4-th (upper plot) and 5-th (lower plot) sources, measured in the classroom at different angles.

5.5 Conclusions

This chapter has investigated how the performance of a loudspeaker array for a personal audio device is affected when it is located in a reverberant environment. The two main effects of the reverberation are: First, the reverberation generates an extra pressure component, which on average, is of uniform level in the reverberant space and reduces the free field performance of the source, and second, the reverberation acts as regularization for the solution of the optimal vector of source strengths.

The extra pressure component due to reflections from the room surfaces is added to the radiation of the array in each direction, which diminishes the directional characteristics of the radiator. If a large level difference between two spatial zones is desired inside a room, the line array has to be very directional in a 3D sense. As an example, the reduction in the acoustic contrast produced by a superdirective array using 8 phase shift hypercardioid sources is considered inside an audio listening room. The acoustic contrast inside a reverberant space is reduced much less if each individual source is made more directive in the vertical plane, as in

the 4×8 array investigated here.

A formulation has also been introduced that allows the performance of an acoustic radiator to be estimated in a reverberant environment. The results of this formulation have been compared with measurements of the two arrays inside a reverberant environment, showing a good agreement. This theory can thus be used at the design stage of a line array to predict its characteristics so that the desired sound control is obtained in a room.

The effect that the reverberation imposes on the robustness of a line array has also been considered. It has been shown that small mismatches in the transfer responses lead to a large reduction in the free field directivity, but do not lead to such a large reduction in performance when the directivity is measured in a reverberant environment. Even a single measurement of the response in such a reverberant environment, when used to calculate the filters of the array, can lead to significant regularization of the system and enhanced robustness to uncertainties in the response between the loudspeakers and the microphones.

Chapter 6

Assessment of the speech intelligibility improvement introduced by the array

In order for the array to be used as an aid for the hard of hearing, the device has to provide a good balance between amplification to the hearing impaired TV viewers and annoyance to normal hearing TV viewers. The previous chapters have shown that by using the 4×8 superdirective array, it is possible to obtain a large directional response in a realistic environment, which in terms of acoustic contrast, satisfies these requirements.

Apart from an assessment of the acoustic performance, it is desirable to also quantify the effectiveness of the array in improving speech intelligibility. This is considered in this chapter through answering the following two research questions:

1. Can the speech intelligibility of TV programmes be improved through the use of the loudspeaker array for listeners with age-related hearing loss?

2. To what extent is the sound quality of the TV programmes affected for normal hearing people sitting in the dark zone when the array is being used?

To address these questions, two different approaches have been taken. A fist approach has been to use an objective speech intelligibility measurement, using the speech transmission index (STI). These simulations have served as an initial benchmark for the performance of the array.

In order to appraise the second research question, and confirm the results obtained with the STI simulations, two subjective listening tests have also been

performed. Most of the information related with the subjective assessment of the array performance has been obtained from the MSc project of Arthur Marker [106]. The author has also been involved in the design and experiments of these tests [146] .

This chapter first establishes the procedures that have been followed to equalise the output of the array, compared with the output of a simulated TV sound system. This information is contained in Section 6.1. After that, Section 6.2 introduces the formulation of the STI and considers the improvement of intelligibility by the use of this metric. The subjective assessment of speech intelligibility improvement in the hearing impaired, and the audio quality in the dark zone via listening tests are described in Section 6.3.

6.1 Design of filters for the array

To provide a personalised amplification for the hearing impaired, the array requires a very narrow beam, which only affects the audio reproduction near the hearing impaired listener. The creation of such a beam, in terms of directional performance, has been presented along Chapter 3 and Appendix 4 of this document. For practical use, it is also important to equalise the output of the array, so that the sound produced by the array in the dark zone does not affect the quality of the normal TV audio.

In order not to affect the audio reproduction at the normal-hearing listeners in the dark zone, the output of the array in this zone has to produce a sound pressure level (SPL) lower or equal than that produced by the TV speaker. In an initial equalisation approach, the output of the array has been equalised so that this is producing the same RMS pressure as that produced by the TV speaker in the dark zone. Although this needs knowledge of the transfer functions to calculate the equalisation, the results are not much different from adjusting the levels of both the array and the TV in the dark zone by listening to them.

6.1.1 Control zones disposition and array performance

A typical listening situation can be represented by 4 TV spectators, with one of them assumed to be hearing impaired and in a need of an extra amplification. This listener position forms the bright zone, whilst the positions of the group of the other 3 listeners constitute the dark zone. The control geometry representing such a situation is illustrated by Fig. 6.1, which is formed by four dummy heads,

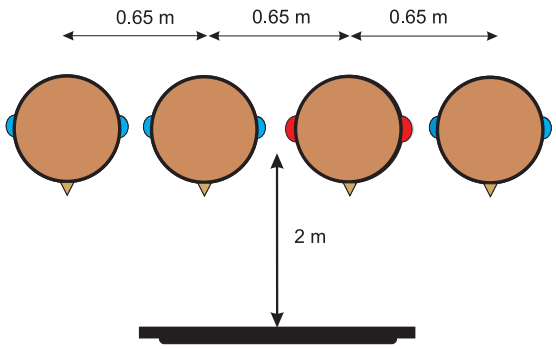


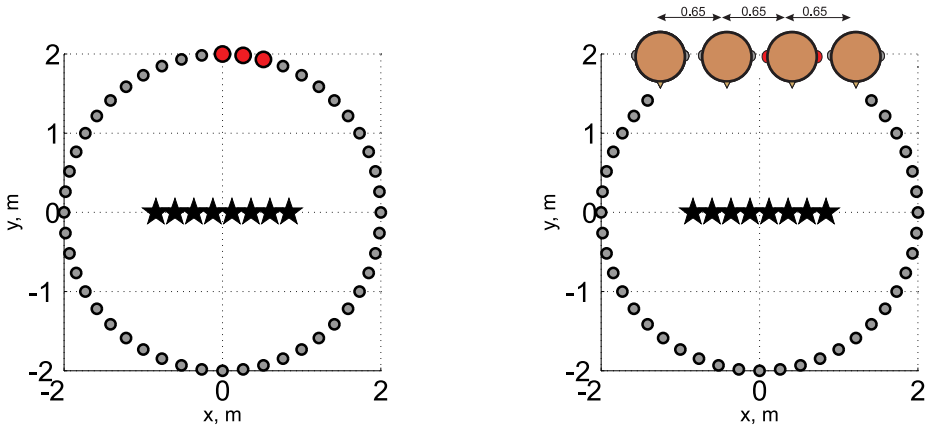
Figure 6.1: Control zone used to simulate an everyday scenario. The blue ears viewers have normal hearing, setting the dark zone, whilst the red ears viewer represents a spectator with hearing loss, which forms the bright zone.

spaced 0.65 m apart.

To maximise the reproduction inside a reverberant space in a control zone as shown in Fig. 6.1, the main beam of the array needs to be steered towards the *hearing impaired* dummy head. Two control geometries have been considered for creating the array filters according to this specification, one only formed by microphones, and a second one using the dummy heads as control points. Both control geometries are shown in Fig. 6.2. Although the contrast can be maximised for a 4 control microphones situation as for the case of Fig. 6.1, more dark control points are needed in order to minimise the excitation of the reverberant field. This is obtained by placing a microphone array of dark control points around the source array, hence reducing the radiation to other directions. The filters designed using the circular microphone array, as in Fig. 6.2a, are denoted the *circular array filters*. The filters designed according to the circular microphone array plus the dummy heads representing the 4 listeners, as in Fig. 6.2b, are denominated *dummy head filters*. Both sets of filters have been obtained with free-field transfer impedances using the LSIF formulation.

For simplicity, it can be assumed that the acoustic pressure level in the dark zone is constant for all control points. Therefore, the dark zone can be reduced from 6 points to 2 points, as shown in Fig. 6.3. In this case, the healthy hearing listener is sat at the left hand side of the array, constituting the dark zone. The hearing impaired listener is sat at the right hand side of the array, in order to receive the main beam as created with the control zone of Fig. 6.2a.

In order to assess the acoustic contrast performance that the different set of filters produce, the reverberant acoustic contrast performance has been calculated off-line by multiplying the transfer impedances for the control zone of Fig. 6.3



(a) Circular array filters **(b)** Dummy head filters

Figure 6.2: Control zones used to create the filters of the array. The red close circles represent the bright zone control microphones, whilst that the grey closed circles represent dark zone microphones, with the stars representing the array sources.

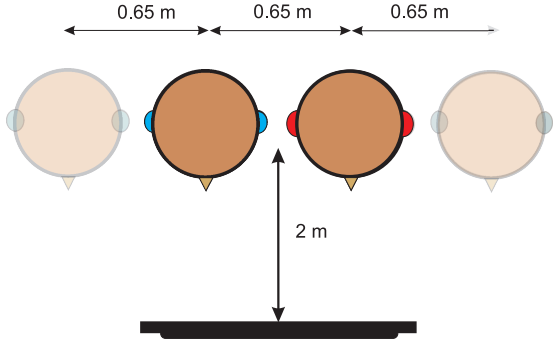


Figure 6.3: Simplification of the dummy head control scenario of Fig.6.1, in where only two listeners are considered. The dummy head with the red colour ears represents the hearing impaired listener, forming the bright zone. The dummy head with the blue colour ears is forming the dark zone, representing a healthy hearing listener.

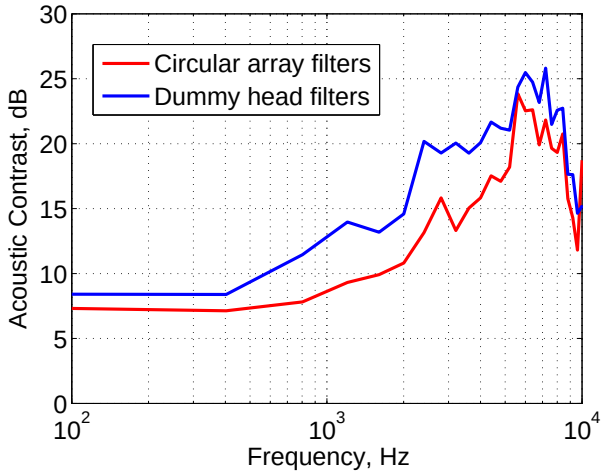


Figure 6.4: Acoustic contrast obtained in the control zones shown in Fig. 6.3 inside the listening room shown .

Table 6.1: RMS contrast, C_{RMS}, obtained using different filter configurations and different dark zones using babble noise as the test signal.

Dark zone	C_{RMS}
Dummy head filters	2.01
Circular array filters	1.48

with both set of filters, the "circular array" filters and the "dummy head" filters. These results are shown in Fig. 6.4, where it can be seen that the performance of the "dummy head" filters is greater than that presented by the "circular array" filters along the whole of the frequency band.

The performance of both set of filters is further checked by using the RMS contrast ratio, C_{RMS}. This is defined as the ratio between the RMS levels in bright and dark zones with the array ON and OFF, so that

$$C_{RMS} = \frac{\frac{RMS_{B_{ON}}}{RMS_{B_{OFF}}}}{\frac{RMS_{D_{ON}}}{RMS_{D_{OFF}}}} = \frac{RMS_{B_{ON}} RMS_{D_{OFF}}}{RMS_{B_{OFF}} RMS_{D_{ON}}}. \tag{6.1}$$

The C_{RMS} results that both set of filters provide in the control zone of Fig. 6.3 are shown in Table 6.1. These results have been calculated using babble noise as the test signal, and show that the best C_{RMS} can be obtained using filters created with dummy heads. This result is supported by the higher acoustic contrast that this configuration obtains, as seen in Fig. 6.4.

Finally, however, the filters created using the dummy control zone are not used, since they introduce a large amount of colouration in the dark zone, which is not desired. This is caused by the fact that the least squares filters try to equalise for the frequency response of the HRTFs. Instead, filters created using the circular control zone are used. Although this set of filters provides a lower C_{RMS}, they interfere very little with the TV audio in the dark zone, which is more desirable, as the quality of the dark zone audio is preserved.

6.1.2 Adjusting the gain of the array

As a first design stage, a small loudspeaker has been used to represent the sound system of the TV. In order to monitor the levels produced by the TV and hence equalise the array based on those, the transfer functions from the TV loudspeaker to the control points of Fig. 6.3 have been measured inside the listening room. Although this operation in a real system would be needed to estimate the transfer functions of both array and TV sound system in the listening environment, it has

been performed here as an initial equalisation procedure. Both the loudspeaker representing the TV and the 4×8 array in its earliest version are depicted in Fig. 6.5.

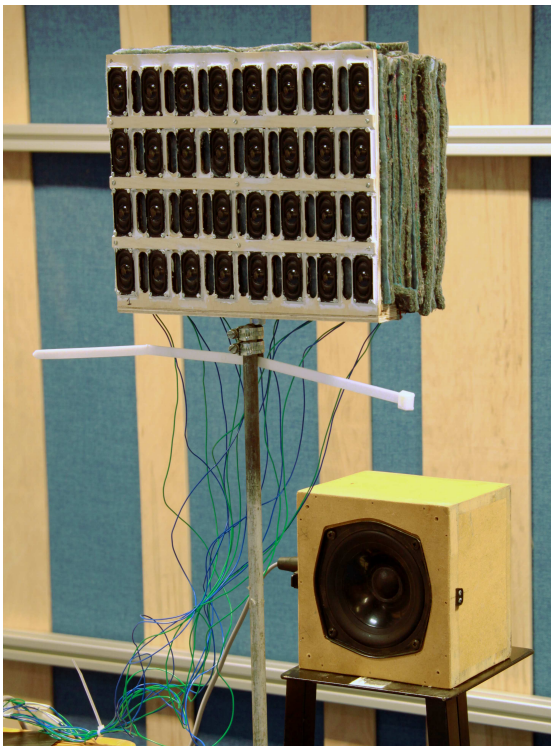


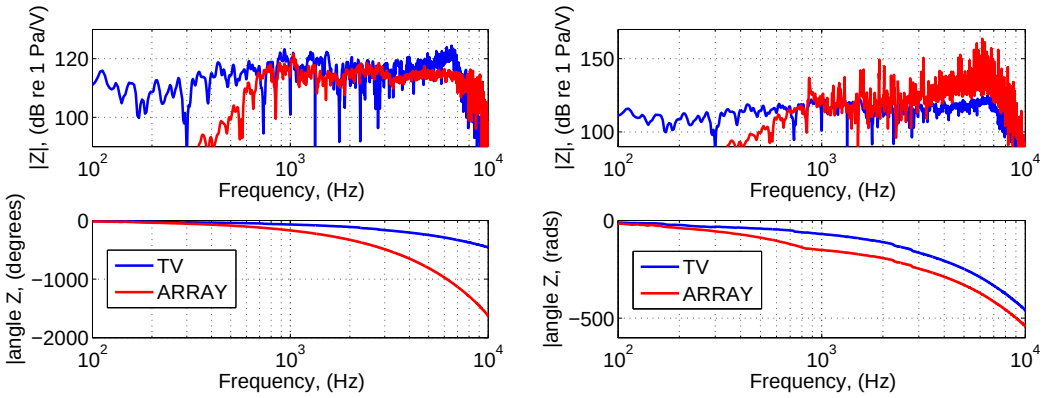
Figure 6.5: Array and loudspeaker simulating the TV.

As the bright zone represents the environment where the complementary amplification is performed, it is important that the radiated signals due to the array and the TV loudspeaker arrive in phase. The array and TV bright zone transfer responses are shown in Fig. 6.6, and these two present different amplitude and phase, and hence, need to be equalised so that a constructive interference is formed. This is obtained at each frequency by calculating

$$Z_{Array,EQ} = \frac{Z_{TV}}{Z_{Array}} Z_{Array}, \tag{6.2}$$

where EQ stands for equalised. The transfer functions of the array, before and after equalisation, are shown in Fig. 6.6. It can be observed how the action of Eq. 6.2 has changed the phase of the signal produced by the array which is now closer to that produced by the speaker. The amplitude response of the array has also been increased.

It is assumed that the maximum amplification applied to the bright zone without annoying the users of the dark zone is obtained when the frequency responses



(a) Before Eq. **(b)** After Eq.

Figure 6.6: Average transfer functions of the array and of the TV loudspeaker to the bright zone before (left hand side plot) and after (right hand side plot) Eq. 6.2 has been applied.

produced by the array and the TV in the dark zone are equal in amplitude, which is given by

$$Z_{Array,EQ_2} = \left(\sqrt{\frac{|Z_{TV}|^2}{|Z_{Array,EQ}|^2}} \right) Z_{Array,EQ}. \tag{6.3}$$

After Equations 6.2 and 6.3 are applied to the original array filters, the responses shown in Fig.6.6 are obtained. The equalisation has been performed so that Eqs. 6.2 and 6.3 are only applied above 850 Hz, which has shown to maximise the RMS contrast between both control zones, C_{RMS}, see Eq. 6.1.

As this equalisation is performed in the frequency domain, the inverse Fourier transform has to be applied to obtain the time series of the arrays filter. This creates excess of pre- and post- ringing in the array filters, as observed in Fig. 6.7, which decreases the quality of the audio. A truncation, as performed in Section C, has therefore been applied to give a better audio quality. Applying a hamming window of 1500 samples, filters have been created which provide a good compromise between audio quality and C_{RMS} ratio. The truncated impulse response is also shown in Fig. 6.7.

Results using this equalisation method are shown in Fig. 6.8. The pressures created by the array and the TV loudspeaker in the dark zone are identical, however, the total pressure in the dark zone is increased when both TV and array radiate at the same time. This can be further reduced by turning down the gain of the array by 3dB, as shown in the right hand side plot of Fig. 6.8. The pressure gain in the bright zone with the array ON with respect to when this is OFF is

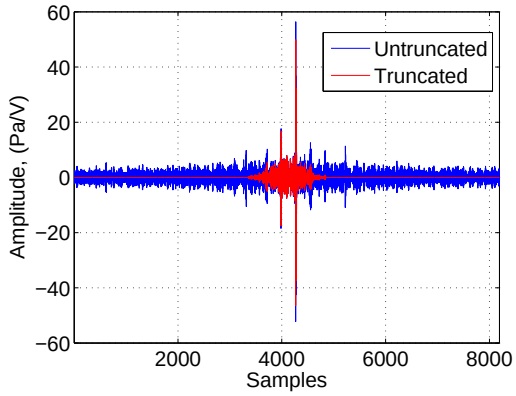


Figure 6.7: Equalized impulse response of one of the array filters before and after the truncation.

identical to the acoustic contrast obtained in the control geometry of Fig. 6.3. Although the total acoustic pressure is still increased with the array ON, the gain of the array could be further turned down to avoid any disturbance.

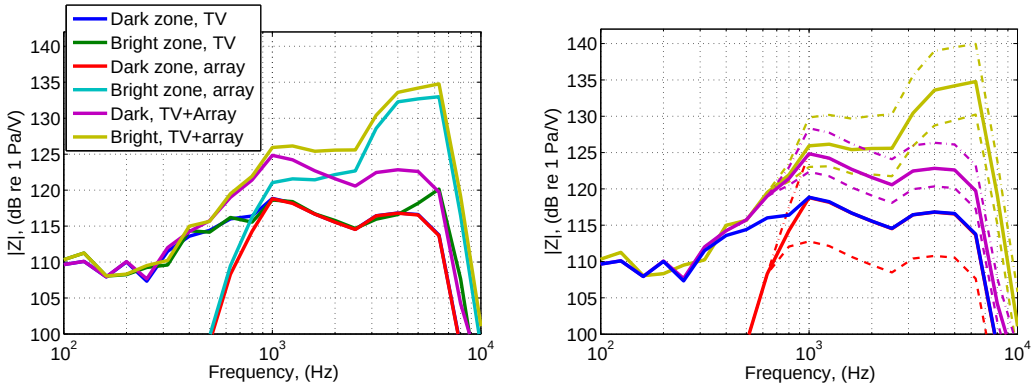


Figure 6.8: The left hand side plot shows the frequency responses of dark and bright zones with the array ON and OFF, averaged in third octave bands. The right hand side plot shows the variation in the dark and bright responses when the gain of the array is selected to be -3 dB (dashed lines) and +3 dB (dashed-dotted lines) with respect to the response of the TV speaker.

Based on the responses shown in the right hand side plot of Fig. 6.8, three sets of filters have been created to test the improvement in speech intelligibility that the array introduces with respect to the audio of the TV sound system. These sets of filters are:

- A set of filters in which the array produces the same SPL as the TV speaker in the dark zone, "0 dB" filters.
- A set of filters in which the gain of the array is turned 3 dB down with respect to 0 dB case. This set of filters is known as "-3 dB" filters.

- Another set of filters in which the gain of the array is increased by 3 dB with respect to the 0 dB case. This set of filters is termed as "+ 3 dB" filters.

Hence, the decibel relation for each set of filters is calculated as $10 \log_{10} \left(\frac{p_{D,Array}}{p_{D,TV}} \right)$.

6.2 Objective assessment of speech intelligibility enhancement

It is important to determine the quality of the speech transmitted via a certain channel. In an auditorium for example, the transmission of speech information to all listeners should be maximised, whilst in an open-plan office, the speech that is transmitted between two colliding desks should be minimised. For the loudspeaker array presented in this thesis, it is important to determine the change in the speech intelligibility of the hearing impaired listeners.

The first objective approaches for the assessment of speech intelligibility were based on an average of the number of words or nonsense syllabus that a number of different listeners were able to understand from a random speaker, however, these tests were time consuming and needed a large number of speakers and listeners to obtain a good characterization of the channel. This led to the creation of objective measurements based on the physical characteristics of a channel. French and Steinberg published a method in 1947 [157] which established a metric known as articulation index (AI), which was reconsidered by Kryter [158] and led to the standard ANSI S3.5-1969. This standard was modificated in 1997, to produce the speech intelligibility index (SII) [29], which is nowadays, together with the STI, one of the most used methods to predict and assess speech intelligibility through a transmission channel.

6.2.1 The speech transmission index (STI)

The STI as originally presented by Houtgast and Steeneken [159], used a different formulation from the one it has nowadays, and a speech-shaped noise as the test signal. In the original formulation, the quality of a transmission channel depended on the reduction in level between input and output at different frequency bands.

The shortcoming of this method was that it could not cope with effects as reverberation, noise or echoes, which made it not suitable to assess the quality of channels inside of enclosures. In order to account for effects of reverberation, the concept of modulation transfer function (MTF) was introduced in 1973, also by Houtgast and Steeneken [160]. The MTF measures the difference of modulation of a signal that occurs through a certain channel. The reduction of modulation can be seen as a a reduction in the speech-envelope spectrum, acting as an attenuation filter applied upon the original speech envelope spectrum. The characteristics of this attenuation filter are defined by the distortions (signal-to-noise ratio (SNR), degree of reverberation) of a given channel, and hence apply to any signal [161].

In Fig. 6.9 it is presented the optic analogue of the MTF, where the lines can be distinguished until their size is very small. The size of the last line that can be distinguished is hence affected by the MTF of this optical transmission channel (paper-air-eye).

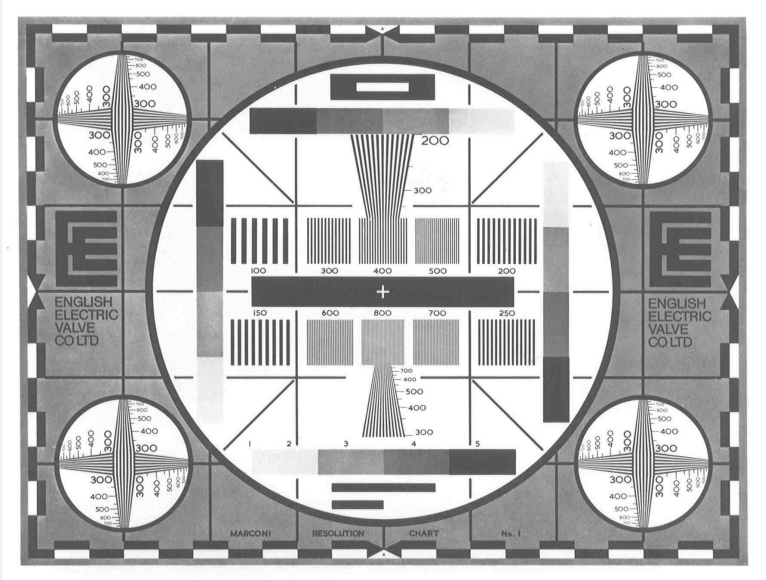


Figure 6.9: Optical equivalent of the modulation transfer function. Reproduced from [161].

The effectiveness of the STI in predicting the performance of transmission channels in reverberant environments, which is of special importance for the evaluation of the array, has been shown previously [162, 163]. The current formulation of the STI is presented in the standard IEC 60268-16 [42].

The modulation reduction

The first step is to determine the effective signal to noise ratio within 7 octave bands, from 125 Hz to 8 kHz. For this purpose the input speech signal is replaced by a test signal with the same spectrum as the LTASS. A graphical example is presented on Fig. 6.10, in where interfering noise is added to the original signal.

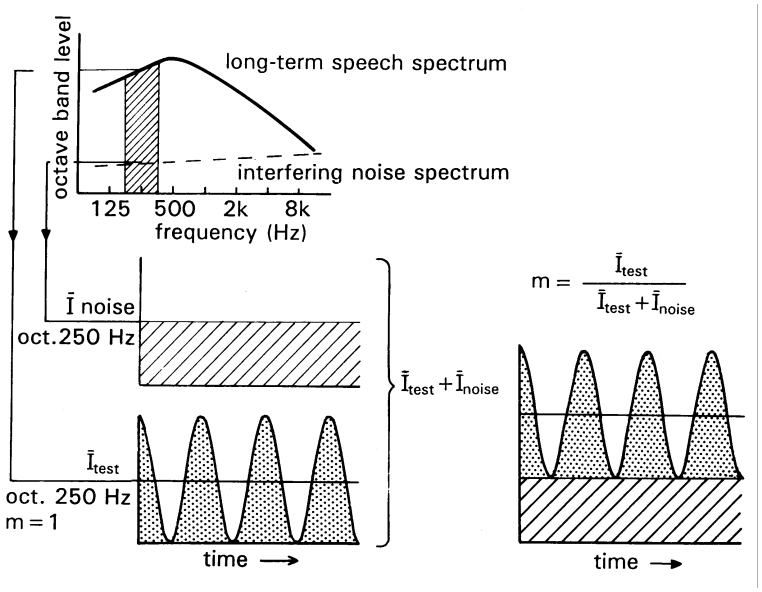


Figure 6.10: Relation between the modulation index of a transmission channel m_k and the signal-to-noise ratio. Reproduced from [161].

If the input signal is sinusoidal, as in the case of Fig. 6.10, the SNR of a transmission channel can be derived from the observed decrease in the modulation index. If a signal with modulation index equal to 1 is described as

$$I_k(t) = \overline{I}_k(1 + \cos 2\pi f_m t), \tag{6.4}$$

where $\overline{I}_k$ is the octave mean intensity and F is the modulation frequency, the signal at the end of the channel will be

$$I_k(t) = \overline{I}_k(1 + m_k \cos 2\pi f_m t), \tag{6.5}$$

in which the modulation index m_k depends on the intensity ratio between the test signal and this plus an interfering noise

$$m_k = \frac{\overline{I}_{TEST}}{\overline{I}_{TEST} + \overline{I}_{NOISE}}. \tag{6.6}$$

The signal to noise ratio can be then derived by performing

$$SNR = 10 \log \left(\frac{\overline{I}_{TEST}}{\overline{I}_{NOISE}} \right) = 10 \log \left(\frac{m_k}{1 - m_k} \right). \tag{6.7}$$

The values of m_k at each band can be obtained by correlating the original envelope function with the received envelope function.

The calculation of the STI

The STI is based in the contribution of a number of frequency bands within the frequency range of speech. This contribution is determined by the *effective* SNR. This SNR is called effective because it may be determined by several factors; background noise which contributes directly to lower SNR, but also effects as non-linearities and distortions in the time domain are considered as noise.

The envelope function of a fluctuating speech is formed by a succession of speech events from short speech items, as explosive sounds, up to words and sentences. Due to the time domain distortion (reverberation, echoes, automatic gain control) these fluctuations may be affected, leading to a reduction in intelligibility [161]. This is modelled in the STI by determining the MTF for the range of relevant frequencies present in the envelope of natural speech signals. In the STI formulation these frequencies are selected from 0.63 Hz up to 12.5 Hz [162], separated in 1/3-octave steps, which yields to a whole of 14 bands. This bears to a whole of 98 measurements needed to calculate the full STI. Fig. 6.11 shows the measurement set-up for one octave band, where a signal with the frequency spectrum of the long-term speech spectrum is amplitude modulated as in Eq. 6.4. At the receiving end the signal is octave band filtered and envelope detection is applied. From the resulting envelope function a Fourier analysis determines the modulation index reduction, due to the reduction by the transmission channel [161].

By using this measurement procedure, distortions as band-pass limiting, noise masking and distortion in the time domain can be deal with. Non-linear distortions have to be taken into account differently, as otherwise intermodulation components in other octave bands can be introduced, leading to a wrong STI estimation. For the case of non-linear distortions a special test signal is needed, wherein the frequency band being tested is replaced by a sine wave, which will be uncorrelated with respect to the speech at other frequency bands [162].

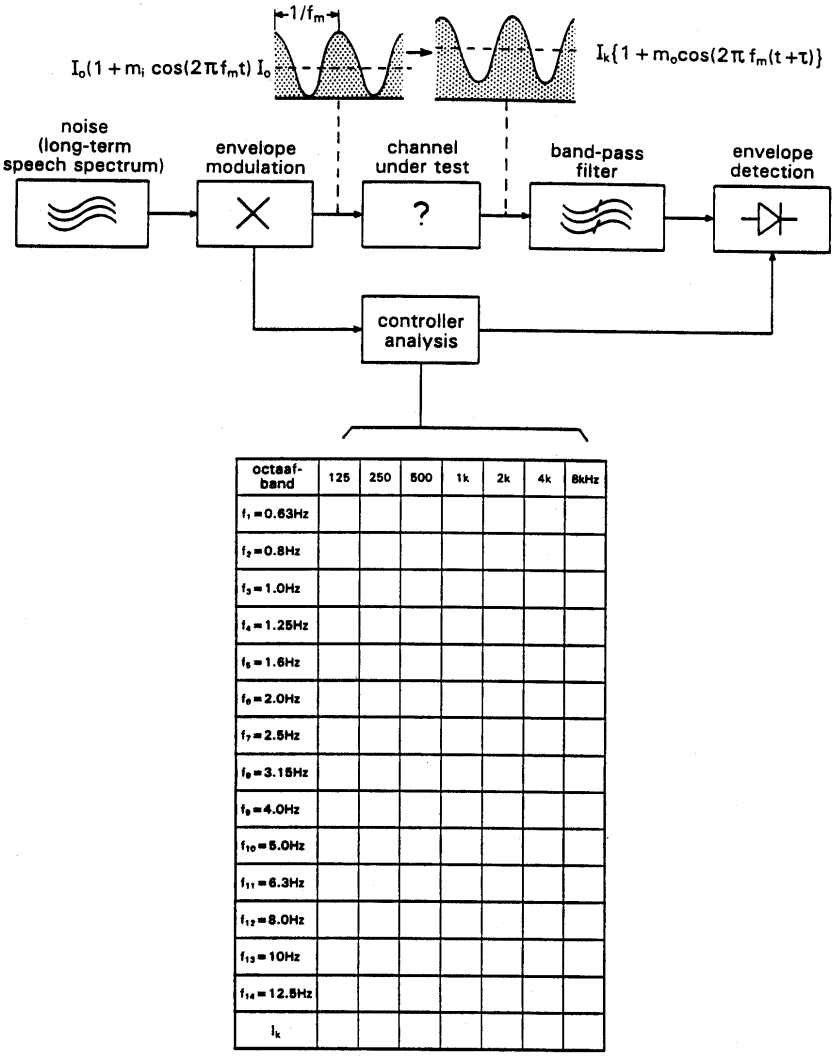


Figure 6.11: Block diagram of the measurement set up for one modulation frequency in one octave band, and the complete set of measurements needed to calculate the full STI. Reproduced from [161].

A method based on the squared impulse response of the transmission channel can also be used for obtaining the matrix of MTF [115]. This approach is known as the *indirect* method. Each MTF is obtained by

$$m_k(f_m) = \left| \frac{\int_0^\infty h_k(t) e^{-j2\pi f_m t} dt|}{\int_0^\infty h_k(t) dt} \right| \left[1 + 10^{-SNR_k/10}\right]^{-1}, \tag{6.8}$$

where h_k represents the squared impulse response for the octave band k, and SNR_k the SNR on the same octave band. This method is also limited to linear, time invariant systems.

Once the whole matrix of MTFs is obtained, it is needed to take into account the masking between adjacent octave bands and if supratreshold levels exist. In

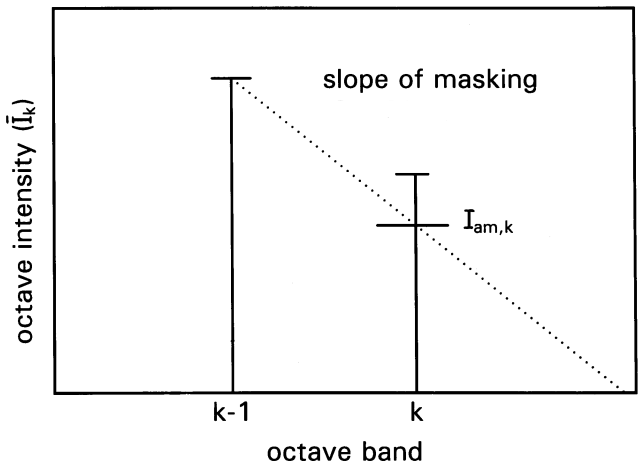


Figure 6.12: Auditory masking of octave band $k-1$ upon the next higher octave band k. Reproduced from [42].

the STI, these effects are modelled as an imaginary masking noise that leads to a decrease of the effective SNR. Fig 6.12 shows an example of how a lower frequency band masks a higher frequency band.

The MTFs are modified according to the auditory masking induced by the lower octave frequency band, as shown in Table 6.2.

Table 6.2: Auditory maskings as a function of octave band levels as given by the standard IEC-60268-16 [42].

SPL L of octave band $k-1$, (dB)	L_{k-1} <63	63≥ L_{k-1} ≤67	67≥ L_{k-1} ≤100	L_{k-1} >100
Auditory masking $amdB$, (dB)	0.5L-65	1.8L-146.9	0.5L-59.8	-10

The masking intensity, $I_{am,k}$ is given by

$$I_{am,k} = I_{k-1} amf, \tag{6.9}$$

where

$$I_{k-1} = 10^(L_{k-1}/10), \tag{6.10}$$

being L_{k-1} the overall sound pressure level of speech and noise for the octave band $k-1$. amf is obtained from the values of Table. 6.2 by doing

$$amf = 10^{amdB/10}. \tag{6.11}$$

Once that $I_{am,k}$ is obtained, this is used together with the absolute hearing threshold, $I_{rs,k}$. The absolute speech threshold intensity is modelled in the STI as an

intrinsic noise floor which reduces the effective signal to noise ratio when speech levels are low. Hearing loss (HL) can be added to the level of the absolute speech threshold to simulate a hearing impaired listener [161]. Both of these quantities are summed together to lead to a reduction of the modulation transfer indices so that

$$m_{k,f} = m_{k,f} \frac{I_k}{I_k + I_{am,k} + I_{rs,k}}. \tag{6.12}$$

A matrix of SNRs is created corresponding to each octave band and each modulation frequency using Eq. 6.7. The STI works by assuming that a SNR between -15 dB and 15 dB corresponds to an intelligibility of 0 and 1 respectively. Therefore the SNRs of each octave band k and modulation frequency f are normalized and converted to transmission indices

$$TI_{k,f} = \frac{SNR_{k,f} + 15}{30}, \qquad 0 < TI_{k,f} < 1. \tag{6.13}$$

This transmission indices are averaged for each modulation frequency to lead to the modulation transfer index, MTI_k

$$MTI_k = \frac{1}{14} \sum_{f=1}^{14} TI_{k,f}. \tag{6.14}$$

Finally, the STI is obtained by a weighted summation of the modulation transfer indices of each octave band and a corresponding redundancy function.

$$STI = \sum_{k=1}^{7} \alpha_k MTI_k - \sum_{k=1}^{6} \beta_k \sqrt{MTI_k MTI_{k-1}}, \tag{6.15}$$

where

$$\sum_{k=1}^{7} \alpha_k - \sum_{k=1}^{6} \beta_k = 1, \tag{6.16}$$

where α_k is the octave band frequency weightings factors and β_k the redundancy factors. Both of these factors are different for males and for females. They can be observed together with the absolute hearing threshold in Table 6.3.

STI relation with PB-word tests

The great accuracy and reliability of the STI and the fact that it is one of the most widely used intelligibility ratings, resides in its robust relation with a variety of subjective measures, as phonetically balanced (PB) word scores and consonant vowel consonant (CVC) word methods. An example of the relation can be observed in Fig. 6.13.

Table 6.3: Octave band specific weighting and redundancy factors and the absolute reception threshold level.

Octave band, (Hz)		**125**	**250**	**500**	**1k**	**2k**	**4k**	**8k**
Males	α	0.085	0.127	0.23	0.233	0.309	0.224	0.173
	β	0.085	0.078	0.065	0.011	0.047	0.095	-
Females	α	-	0.117	0.223	0.216	0.328	0.25	0.194
	β	-	0.099	0.066	0.062	0.025	0.076	-
Absolute reception threshold, (dB)	$I_{rt,k}$	46	27	12	6.5	7.5	8	12

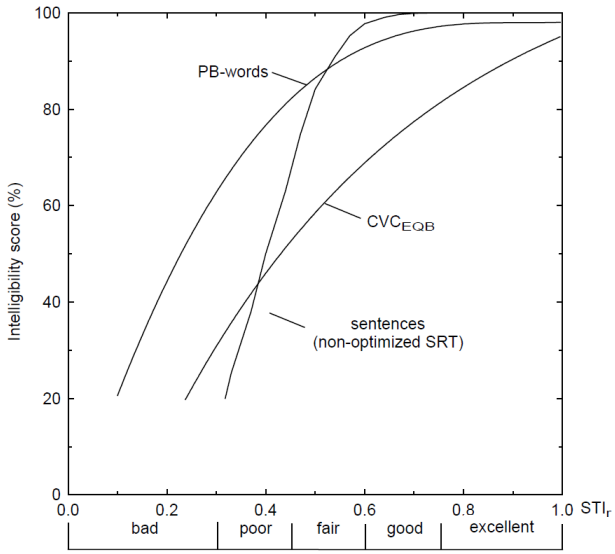


Figure 6.13: Qualification of the STI and relation with various subjective intelligibility measures for male speech [161].

A rating can be associated with the intelligibility according to the standard IEC-60268-16 [42]. This is shown in Table. 6.4, where ratings for normal hearing listeners and hearing impaired listeners are given. According to the standard IEC-60268-16 [42], hearing impaired listeners required 4.5 dB higher SNR for 20 dB HL to reach an intelligibility of the 50%.

[1]dB HL refers to the the HL in dB, defined as the Pure tone average (PTA) relative to 18-year old normal listeners. The PTA approximates the speech reception threshold, and is made of the average of the thresholds of pure tones at 500 Hz, 1 kHz and 2 kHz.

Table 6.4: Adjusted intelligibility qualification tables for normal listeners and people over 60 years old with hearing loss, reproduced from [42].

STI label	**Normal listeners**	**Older listeners, PTA=15 dB**	**Older listeners, PTA=20 dB**	**Older listeners, PTA=30 dB**
bad-poor	0.30	0.42	0.47	0.51
poor-fair	0.45	0.57	0.62	0.66
fair-good	0.60	0.72	N.A.	N.A
good-excellent	0.75	N.A.	N.A.	N.A

6.2.2 Simulated performance of personal audio systems

A first insight into the STI can be performed using the transfer functions of a free-field monopole line array. It is possible, hence, to calculate the transfer response of the whole array working at a time and determine the quality of the signal it produces. Similar simulations as the one introduced here have been presented previously to assess how large a spherical microphone array should be to obtain certain intelligibility characteristics in a certain kind of room [164].

The simulations introduced here test the intelligibility in terms of STI against observation distance and background noise. The STI figure that both the acoustic contrast maximisation (ACM) and lest squares inverse filters (LSIF) obtain is compared between each other.

The MTFs are obtained by using Eq. 6.8 together with h_k, being h_k obtained by inverse Fourier transformation of the impulse response of the whole array to a microphone of the bright zone

$$h_k(n) = \Delta \sum_{-1/2\Delta}^{1/2\Delta} (\mathbf{Z}_{B_i}\mathbf{q}) e^{j2\pi f n \Delta} df, \tag{6.17}$$

where $\mathbf{Z}_{B_i}$ represents the vector of transfer impedances to the centre of the bright zone in the free-field, and $\mathbf{q}$ represents the vector of optimal volume velocities.

One of the elements which affects the intelligibility prediction is the external noise present on the environment where the STI is being measured. If the sources that contribute to the noise is assumed to be further from the STI estimation point than the reverberant radius, it is possible assume that the noise sources are in the reverberant field. The excitation of the reverberant field due to the noise sources

is given by

$$< |p^2_{N_{REV}}| >= \frac{4\rho_0 c_0 W_N}{S\alpha}, \tag{6.18}$$

where W_N is the total acoustic power of the noise sources. The reverberant pressure due to noise is used to modify the actual MTFs using the definition of SNR_k as in Eq. 6.8. SNR_k is calculated at each octave band by

$$SNR_k = 10 \log_{10} \left(\frac{p}{p_{N_{REV}}} \right), \tag{6.19}$$

where $p = \mathbf{Z}_{B_i}\mathbf{q}$ is the direct pressure due to the array.

STI variation with observation radius and SNR

Considering a room with a given reverberation time, it is possible to predict the STI which an array obtains at a certain distance from the source. If an external noise source is present in the room, such source is going to produce an excitation of the reverberant field which will be constant all around the extension of the room. When two arrays with different number of sources radiate with the same acoustic power, the direct pressure of the more directional array will arrive until a longer distance, contributing to a larger STI when the array is further from the STI estimation point.

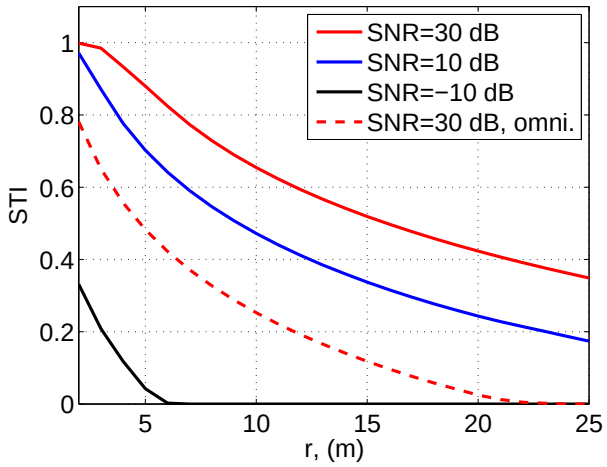


Figure 6.14: STI vs observation radius for different SNRs and for an omnidirectional source. Simulation calculated for an array of 8 sources and for a point source (omni). The STI is predicted at the centre of a bright zone as that of Fig. 6.2a, with a constant 70 dB SPL all along the frequency bandwidth.

Fig. 6.14 shows the STI given by an 8 point hypercardioid array plotted against the observation radius, together with results using an omnidirectional source. A

background noise source is introduced, with the SNR calculated as a function of the total power radiated by the array and by the external noise source. If comparing the STI at 30 dB SNR, the array is able of obtaining a performance about 0.4 larger than the omnidirectional source at various distances, which according to Table 6.3 can change the speech intelligibility from "bad" to "excellent". If the SNR is reduced, the STI is also decreased and the same figure of intelligibility is obtained at positions much closer for the array. The extreme situation happens when the SNR is -10 dB. In this case the array gives a very low STI, even at distances very close to the estimation point.

STI vs. control algorithm

It has been previously observed that the filters created with the ACM perform worse than the LSIF in terms of uniform pressure coverage of the listening zone. This becomes a problem in some situations, wherein the audio quality of the reproduced speech, and hence the perceived speech intelligibility could be reduced. The effect of the algorithm on the STI has been studied, and it has been seen that both beamforming methods perform in a very similar way, as shown in Fig. 6.4, where the STI obtained for the ACM and LSIF methods for different SNR is presented. In this case the array transfer functions obtained with both methods are equalised so that they produce a constant level with frequency in the dark zone, which gives a very similar STI figure for both approaches. Above 10 dB of SNR the LSIF performance is slightly better than that of the ACM method.

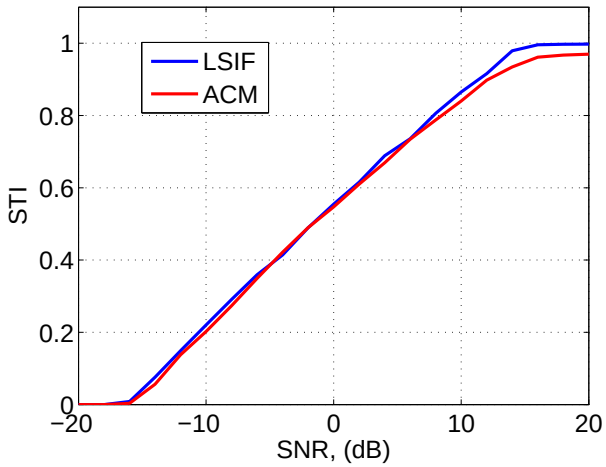


Figure 6.15: STI as a function of SNR when using the LSIF or the ACM method for creating the array filters.

6.2.3 Optimum array structure and dimensions for maximum STI

An iterative script has been implemented which checks the STI for a given separation of the sources of the array. The simulations are performed with a low pass filtered input, so that the higher input frequency to the array is that of the array aliasing limit.

Two array configurations are considered here; a classical uniform spaced line array and an array with 5 high frequency central sources and 7 medium frequency sources, 7×5, so that the spacing of the medium frequency sources is the double of that of the high frequency sources.

The estimated STI, which is shown in Fig. 6.16 has been calculated under an SNR of 20 dB, for a 70 dB SPL and for a 70 year-old woman, based on the hearing loss given by the standard ISO 7029 [12]. The elevated hearing threshold for a 70 year-old listener is created by summing the hearing loss to the absolute threshold of speech [161], however, this has shown not to have a great effect in the final estimated STI.

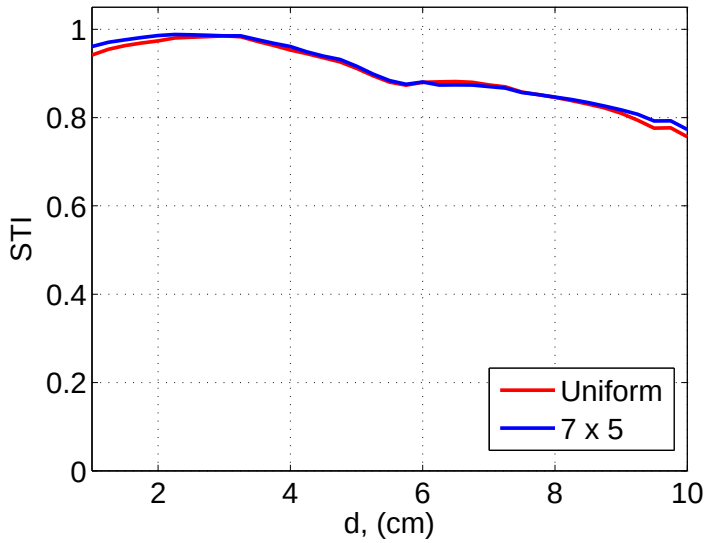


Figure 6.16: STI *vs.* inter element distance, d, between the array sources. Results calculated using LSIF.

It is observed that both arrays give similar results. The optimum distance is placed between 20 and 30 mm, as this places the aliasing limit so that the array output includes the 8 kHz band in the calculation of the STI. The maximum STI is obtained with the separation between the array sources which presents the

lowest reverberant field pressure injection at a certain frequency range, and hence maximise the SNR in the STI weightings which are more relevant.

6.2.4 Improvement of speech intelligibility in the TV

The STI formulation can be used to assess the improvement in speech intelligibility that is gained when the array is working at the same time as the TV loudspeaker. To this end, the STI has been calculated using the responses shown in Fig. 6.6, for a control geometry as shown in Fig. 6.3. The responses have been measured in the listening room of the ISVR, at a distance of 2 m. In order to equalise the array and the TV loudspeaker, the responses of the array and the TV loudspeaker in the dark zone have been equalised based on the RMS levels of both of them when driven with babble noise. The STI has been calculated using Eq. 6.17 with the IRs of the TV loudspeaker or that of the TV loudspeaker combined with the output of the array in the bright zone at different gains. The STI response is hence assessed for various SNR configurations which are different from typical SNR values when watching the TV, but which represents a common way to assess speech intelligibility [165].

The results of such prediction are shown in Fig. 6.17. When the array is OFF, the response of the TV speaker in the bright zone produces a STI of about 0.5 at 0 dB SNR. At the same SNR configuration, the STI is increased by approximately 0.1 if the array is working in conjunction with the TV speaker with a gain of -3 dB. If the gain of the array is further increased until 0 dB or +3 dB, the improvement of speech intelligibility is about 0.18 or 0.26 respectively. In terms of SNR this means that for the same STI score, the SNR is reduced 3 dB if the gain of the array is -3 dB, 5 dB if the gain of the array is 0 dB, or 8 dB if the gain of the array is + 3 dB. According to Table 6.3, an increase in the STI of 0.15 leads to an improvement of the intelligibility score from fair to good and from good to excellent.

The predicted performance shows how the joint use of the array and the TV loudspeaker can increase the intelligibility of speech in normal hearing listeners, thanks to the boost in high frequencies. Although hearing loss can be introduced in Eq. 6.12 and hence calculate the STI for a hearing impaired person, the results obtained are not very different from that when calculating the STI without hearing loss, which suggests that this method is not very effective, as also commented in [42].

Instead, in order to assess the performance of the array when improving the in-

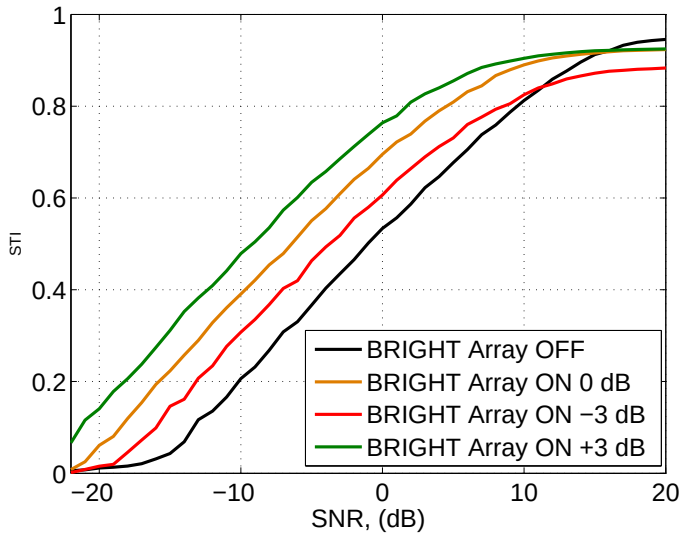


Figure 6.17: Prediction of STI as a function of SNR for the TV loudspeaker and for the array being used with gains of -3 dB, 0 dB and +3 dB. The estimation has been performed inside the ISVR listening room at a distance of 2 m.

telligibility of hearing impaired listeners, subjective tests dealing with hearing impaired listeners are needed. Such assessment is contained in the next section, in where behavioural tests using 30 individuals are carried out to test the performance of the array.

6.3 Subjective assessment of speech intelligibility

Generally, the majority of the acoustic measurements to characterise an audio system are based in objective estimations such as total harmonic distortion (THD), impulse response (IR) or frequency response. These methods, however, do not give direct conclusions on the subjective performance perceived by human listeners. Hence, it is needed in many cases to perform a subjective assessment to provide a better insight into human perception, and back up the observations from the objective results. The subjective evaluation of audio systems is generally performed by individual listening tests. On the other hand, listening tests are complex and time consuming. Nevertheless, although other objective equivalents or models have been tried, listening tests still represent the best approach to judge the performance of a system [166].

In the context of the evaluation of the loudspeaker array to be used as a hearing aid, listening tests can serve to backup the STI estimations presented in

Section 6.2. Furthermore, the STI estimations are based on normal hearing listeners. Concerning the intelligibility enhancement in hearing impaired listeners, only listening tests taking into account this hearing condition can be performed. Furthermore, listening tests are also used to assess the annoyance caused by the array to the listeners of the dark zone when this is operating. For this reason, two separate listening tests need to be carried out, focusing on each of these two aspects.

In order to test the assessment of intelligibility in hearing impaired listeners, a hearing loss simulator has been used. The simulator is used to process audio cues so that when these are listened by healthy hearing individuals they experiment the same impairments as those which hard of hearing listeners suffer.

6.3.1 The modelling of hearing loss

The effect of cochlear hearing loss (CHL), introduces a series of changes in the auditory system, that as described in Section 1.1.1, directly affect the intelligibility of speech. In order to test the performance of the superdirective arrays when enhancing the audio of the TV in the hearing impaired, it is needed to perform subjective tests with hearing impaired users. Another simpler possibility, is that of using a hearing loss simulator to process audio cues. Thus, it is possible to allow a normal hearing listener to experience the perceptual effects of cochlear hearing loss (CHL).

Several studies have examined methods to simulate the perceptual effects of CHL [167, 168, 165, 169, 170]. These methods perform the hearing loss simulation either by using a fast Fourier transform (FFT) or a wavelet based approach. For processing the audio cues used in the subjective assessment of the intelligibility, a FFT simulator kindly provided by Brian C.J. Moore and his team [171] has been used. The operation of such system is presented in Section 6.3.1.

Hearing loss simulators based on the FFT

Another existent approach to model the diminished sensory abilities of people with CHL is by using a FFT based simulator. This technique was first applied by Baer and Moore [168] and by Moore and Glasberg [167]. While the former study was focused on the reduction of frequency selectivity, also known as *frequency smearing*, the second paper studied the effect of threshold elevation and non-linear loudness shift, which is also termed as *loudness recruitment*. The im-

portance of these effects has been subject of study, but it is still not completely understood how the different psychoacoustic factors affect speech intelligibility [2]. Nonetheless, the studies in frequency smearing and threshold elevations showed that they contributed directly to a reduction in speech perception.

These two effects of hearing loss in the cochlear filters were later combined in a simulator presented by Nejime and Moore [165]. The block diagram of the processing performed by the simulator is shown in Fig. 6.18.

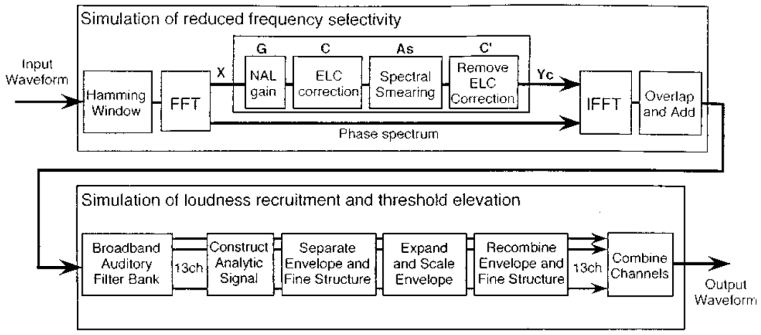


Figure 6.18: Block diagram of the FFT hearing loss simulator developed by Nejime and Moore [165]. Reproduced from [165].

In the block diagram depicted in Fig. 6.18 it can be observed how the processing stage is divided into two submechanisms; first, the simulation of frequency selectivity loss is performed, and then the simulation of loudness recruitment and threshold elevation is carried out. The reduction of frequency resolution is based on the work performed by Baer and Moore [168]. The authors use the procedure described in the block diagram of Fig. 6.19. For each channel or sub-band of analysis, the short-term spectrum is calculated using a windowed FFT. The size of the FFT is of 256 coefficients, to reduce computational complexity and be still sufficiently long to encompass a typical pitch period, so that an acceptable frequency resolution is obtained. Spectral smearing was performed by replacing each component of the power spectrum with a weighted sum of the surrounding components. The weighting function was similar to the shape of the broadened impaired auditory filter centred on the component. The authors used broadened *roex* (rounded-exponential) filters [172] as the weighting function.

Baer and Moore [168] also tried the effectiveness of their model by performing subjective tests on normal hearing listeners, where they simulated several degrees of smearing. The authors observed that speech intelligibility was slightly varied for cases with high SNR, however, the intelligibility of smeared speech decreased

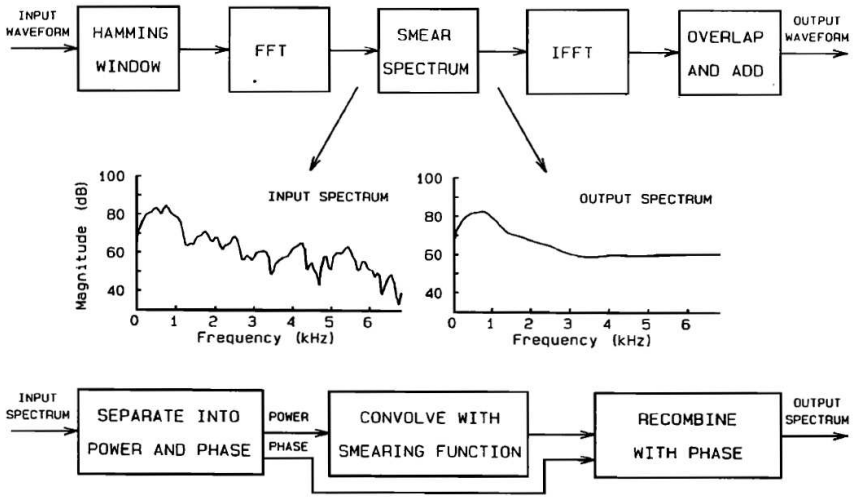


Figure 6.19: Block diagram of the sequence of operations used to perform the spectral smearing by Baer and Moore [168]. The upper part shows the overall sequence, and the lower part shows the sequence of operations within the "smear spectrum" block. Reproduced with permission from [168].

greatly as the SNR was diminished.

Once this process is finished, the time waveform is obtained via inverse Fourier transformation and overlap added. Following this stage, the simulation of non-linear loudness shift combined with the elevation of absolute threshold is performed. The loudness recruitment step is based on the model of Moore and Glasberg [167]. The authors model performs an expansion and scaling of the sound wave in the time domain. It assumes that the cochlear non-linearity acts as an automatic gain control circuit which acts within a few ms in a healthy cochlea. The lack of compressive non-linearity in an impaired cochlea is introduced by a fast-acting expansive non-linearity. As loudness recruitment is more noticeable at high frequency, the audio input is divided into a series of sub-bands, in which the expansive non-linearity is applied independently in each band.

The framework of the loudness recruitment processing can be observed in Fig. 6.18. The model processes the envelope of the signal waveforms in the time domain. For each input waveform, the model first obtains the squared envelope of the waveform, by means of adding the squared waveform and its squared Hilbert transform. The envelopes are hence enhanced to the power of N, where N is the slope of a hearing impaired loudness growth function[1]. After the envelopes are

[1]Loudness growth functions are obtained by plotting the sound pressure level which is needed in a normal ear to obtain a certain loudness as a function of the sound pressure level of an impaired ear. These functions are normally obtained in people with unilateral hearing loss.

processed, they are recombined and multiplied with a smoothed version of the enhanced envelope, to avoid time artefacts.

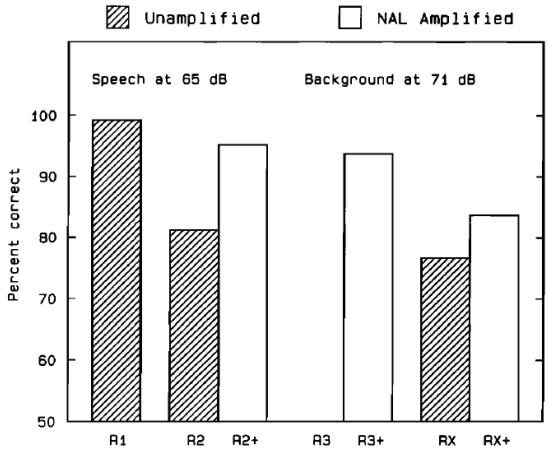


Figure 6.20: Correct percent scores for speech recognition with different hearing loss degrees for a SNR=- 6 dB. $R1$ corresponds to normal hearing, $R2$ corresponds to a moderate HL, $R3$ to a severe flat hearing loss and RX to a severe sloping HL. Reproduced with permission from [167].

After this processing, the speech cues can be used in normal subjects to simulate the elevated thresholds. Moore and Glasberg [167] tried the effectiveness of this simulator by performing trials in normal hearing subjects. The authors concluded that the model simulated both effects with an acceptable accuracy, and that threshold elevation combined with recruitment simulation was enough to decrease the audibility of speech in the presence of a single competing talker. An example is shown in Fig. 6.20, in where *percent correct score* results for speech recognition with some degrees of hearing loss are presented. The effect of the NAL amplification procedure [33] to recover speech intelligibility is also shown in the figure. It can be observed that by using amplification, speech intelligibility can be greatly recovered.

6.3.2 Subjective assessment of speech intelligibility enhancement

Experiment planning

The evaluation of speech intelligibility has been performed by comparing the averaged speech reception threshold (SRT) in noise, i.e., the signal to noise ratio giving the 50% intelligibility for each test condition. The test conditions used in the experiments can be observed in Fig.6.21. The test conditions are based in the filters created in Section 6.1 and are defined as

- DZ Array OFF: Normal hearing person in the dark zone.
- BZ Array OFF: Hearing impaired person in the bright zone with the array OFF.
- BZ Array ON 0 dB: Hearing impaired person in the bright zone with the array ON.
- BZ Array ON +3 dB: Hearing impaired person in the bright zone with the array ON, gain of the array turned +3 dB.

The SRTs for each test condition have been obtained by fitting a probit function to the 50% points of intelligibility scores recorded at different SNRs. The fixed SNRs used in the experiments have been obtained by performing an initial pilot study. These values are shown in Fig.6.21.

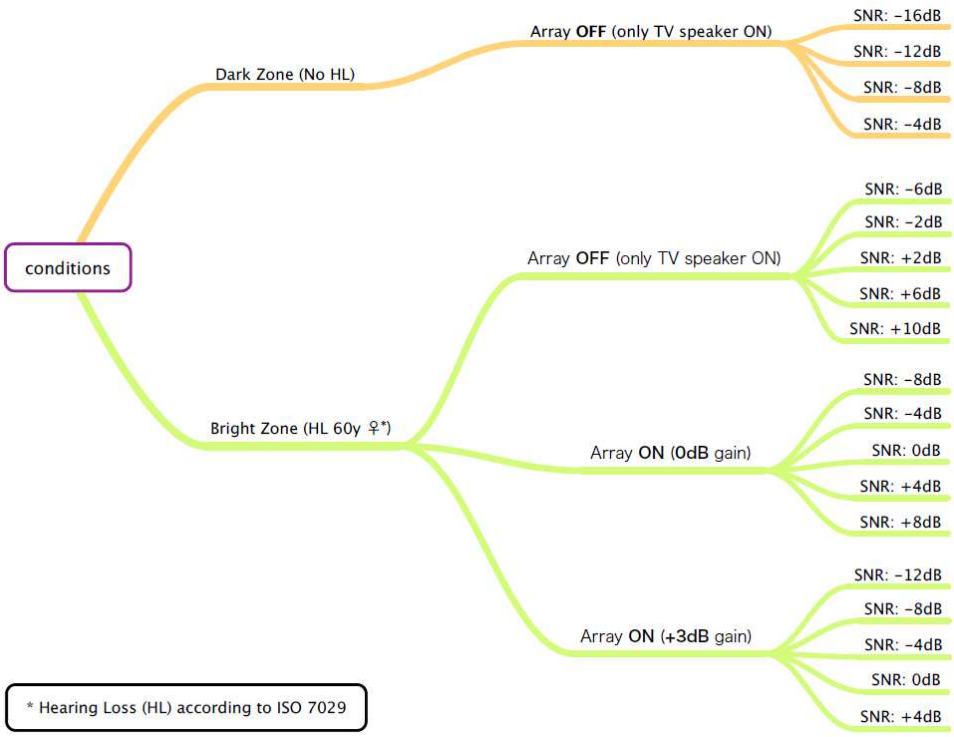


Figure 6.21: Independent variables defined for listening tests for the evaluation of the speech intelligibility improvement introduced by the array. Also shown are the fixed SNRs used at each condition. Reproduced from [106].

The subjective tests were performed using 30 participants plus an additional 3 who were used in an initial pilot study. All the participants were students from the University of Southampton. The participants were 15 males and 15 females,

[1]The dB relation is calculated according to $10\log_{10}\left(\frac{p_{D,Array}}{p_{D,TV}}\right)$.

with all the individuals aged between 19 and 28, with a mean age of 23 years old. All the subjects presented a normal hearing, which was checked before the realisation of the experiment via audiometric check. As intelligibility was tested, all the subjects had to be native English speakers.

To evaluate the intelligibility, the target signal was comprised by Bamford-Kowal-Bench (BKB) sentence lists [173]. The material is formed by 21 sentence sets, each set containing 16 meaningful sentences spoken by female and male English speakers. Only sentences spoken by male talkers were used, however, as to exclude the variable of gender differences. Each set is reported to produce equivalent intelligibility scores when played under equal conditions. Consequently, each condition can be represented by a set of sentences, so that a maximum of 21 different conditions can be tested. According to the SNRs contemplated in Fig.6.21, a total of 19 sentence sets are required to cover all the conditions. The remaining two lists were used to make the subjects familiar with the test.

In order to vary the SNR of each sentence, a masker noise was introduced. The masker consisted of babble noise made out by adding together eight female speakers talking at the same time. The reference condition (SNR=0 dB) was defined by scaling the masking according to the RMS value of the whole set of 336 BKB sentences.

Although it may be more realistic to preform listening tests on people with hearing impairment, in the other hand, the hearing thresholds of impaired people present a large amount of variability. Instead, the intelligibility tests have been performed by modifying the BKB sentences with the hearing loss simulator of Nejime and Moore [165]. The hearing loss introduced by the simulator has been based in the median hearing loss for a 70 year old woman given by the standard ISO 7029 [12]. The processing of the hearing impaired BKB signals was performed off-line.

To test the sentence comprehension, percentage scores were made based on the number of key words correctly repeated by the individuals out of 50, which is the total number of keywords in each sentence set. A keyword was considered to be correctly recognised if the subject repeated back the exact word with same tense, respectively whether it was singular or plural [146]. As the order in which the programme material is presented can affect the subject response, a Latin square design was implemented to assure that the means are not correlated with a specific position of the BKB sentence lists [174].

The more realistic approach to perform the listening tests would be to use the array to present the stimulus. However, in order to avoid the variances introduced by the listening environment, background noise, and head user rotation, the presentation of stimulus has been performed via headphones. In pursuance of a realistic presentation, the recorded head related transfer functions (HRTF) from the array and the TV loudspeaker to each of the control points of the listening zone of Fig. 6.3 were used. The binaural signals presented to each user were formed by convolving the processed BKB sentences with the HRTFs for left and right ear of dark zone or bright zone, for the TV loudspeaker in the case of "Array OFF" or combined with the array signal in case of "Array ON".

The listening tests were performed using a personal computer fitted with Matlab [175]. The material was presented to the participants using circum-aural headphones. The equipment was calibrated with a Brüel& Kjæer type 2260 sound level meter. The volume was adjusted so that the participants were below the sound exposure limit, as defined in [176]. The experiment was controlled using a guided user interface (GUI), as shown in Fig. 6.22.

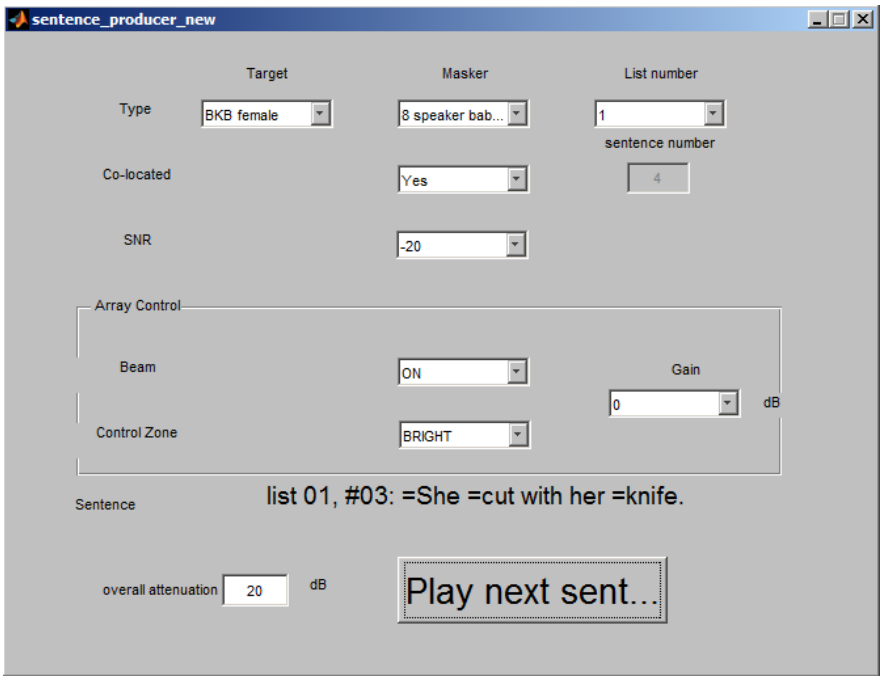


Figure 6.22: Guided user interface employed in the subjective intelligibility tests by the experimenter. Reproduced from [106].

Intelligibility results

After performing the experiments, the obtained data from the 30 subjects was analysed. Fig. 6.23 shows the psychometric functions for each condition averaged over the 30 participants. As observed, a clear distinction can be made between each condition, which proves the suitability of the array to be used as a hearing aid.

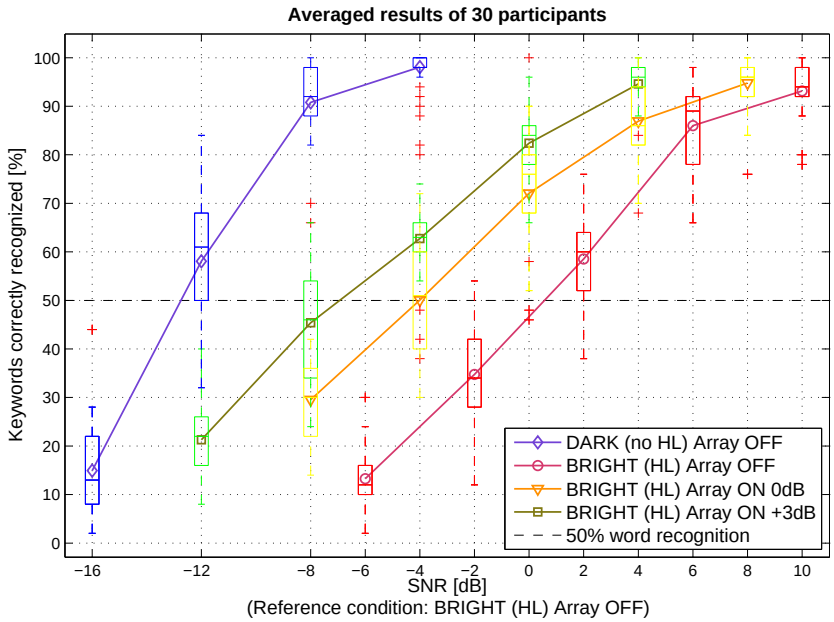


Figure 6.23: Mean psychometric functions for the four different conditions. Reproduced from [106].

For the statistical analysis only the data from the bright zone, corresponding to hearing impaired listeners, was used. In order to test the statistical significance of the obtained results, a one-way repeated measures analysis of variance (ANOVA) was performed to compare the mean between related groups. The participants with the SRT as continuous dependent variable, under three different conditions, "BZ Array OFF", "BZ Array ON 0 dB" and "BZ Array ON +3 dB". These conditions were the levels of the independent variable, the gain. The measured dependent variable, the SRT, is obtained through probit analysis [177] based on the subjects performance on correctly repeated key words. The condition "DZ Array OFF" was not included in the analysis, since the main focus of the experiment was to evaluate whether there are statistically significant improvements in terms of intelligibility scores for a hearing impaired person in the bright zone through the use of the loudspeaker array.

In order to perform the ANOVA analysis, it was necessary to check that the data was normally distributed and that sphericity was not violated. Sphericity was checked via Mauchly's test $\chi^2(2) = 2.5246$, $p = 0.28$. A Shapiro-Wilk ($p > 0.05$) test was used to check that the data sets were normally distributed. The obtained SRTs are statistically significantly different for different levels of the independent variable (gain). $F(2, 58) = 427.089$, $p < 0.05$, *partial* $\eta^2 = 0.936$. The SRT mean and standard deviation for each gain level are shown in Table 6.5. The statistical analysis has shown that the SRTs have improved in terms of statistical significance ($p < 0.0005$) for a person with simulated hearing loss through the use of the loudspeaker array [106].

Table 6.5: SRT mean and standard deviation (SD) for the gain configurations in the bright zone. Reproduced from [106].

	Array OFF	**Array ON 0 dB**	**Array ON +3 dB**
Mean SRT, (dB)	0.45	-4.07	-6.66
SD of SRT, (dB)	1.30	1.79	1.75

Another way of observing the results is to look at the improvement of speech intelligibility obtained at a given SNR. It can be seen in Fig. 6.23 how at -4 dB of SNR a hearing impaired listener would only understand the 24% of the words with the array switched OFF. If the array is operating, the percentage of recognised words grows to be of 50.1% when the gain is 0 dB or to 62.7% when the gain of the array is +3 dB. These results are still significantly lower than that of a person with normal hearing, as it has been shown that a total recovery of speech intelligibility is not possible in persons with hearing loss [2]. Nevertheless, the use of the array would be very beneficial towards enhancing speech intelligibility for a hearing impaired person.

6.3.3 Annoyance caused by the array outside the listening area

Experiment planning

A listening test has also been carried out to quantify how much the audio quality in the dark zone is modified when the array is turned ON. The evaluation of the audio quality perceived in the dark zone has been performed using the standard ITU-R BS.1534, "MUSHRA" [178]. This standard has previously shown to provide accurate and reliable results for the evaluation of audio quality [179, 180].

The evaluation of the quality is based on the perceptual impression of the individuals to the audio material, defining the different conditions of the experiment. The amplification conditions for the four conditions of the dark zone are: "DZ Array OFF", considered as a reference, and three configurations for the array gain: "DZ Array ON -3 dB", "DZ Array ON 0 dB" and "DZ Array ON +3 dB". The standard MUSHRA also requires of a hidden reference plus a hidden anchor, which is constituted by a low pass filtered version of the reference signal. The anchor is required to compare the system under test in a more absolute scale [178]. For the evaluation, a test subject has to compare the three conditions when the array is ON with each other as well as with the reference, and give an opinion on the perceived quality through a user interface.

For the evaluation of the dark zone audio quality 17 participants (12 male and 5 female) have been used, aged between 23 and 29 with a mean age of 26. All listeners were checked to have normal hearing prior the experiment. The participants set was formed by a conglomerate of native and non native English speakers.

According to the standard ITU-R BS.1534, the test material for the audio quality evaluation should be formed by typical broadcast material [178]. To this end, the previously used BKB sentences have been used to perform the quality evaluation. Three, different, randomly selected sentences were chosen. Each test sentence was formed by the BKB signal convolved by the TV loudspeaker HRTF plus the same BKB signal convolved by the array HRTF corresponding to each gain configuration. The different test sentences were presented by headphones, in order to get rid of the variability due to position non-alignment and room response. It is assumed however, that the audio quality experiment will not be affected by the fact that the test subjects have different HRTFs from those measured with the dummy head, as this difference will be identical for all the scenarios being tested.

The tests were performed using a GUI in a provided laptop. Before the realisation of the experiment, the test individuals were trained on how to perform the main experiment. The quality was evaluated in a scale from 0 to 100. The subjects were told to rate the reference signal as a primary task, and to assign a 100 out 100 to it as it represents the best quality available. The perceived quality of the rest of the test sentences could be divided into 5 equal intervals: 0-20, bad; 21-40, poor; 41-60, fair; 61-80, good and 81-100, excellent.

Experiment results

The dark zone audio quality evaluation results are depicted in Fig. 6.24. As expected, the plot shows how the perceived audio quality is decreased as the gain of the array is increased.

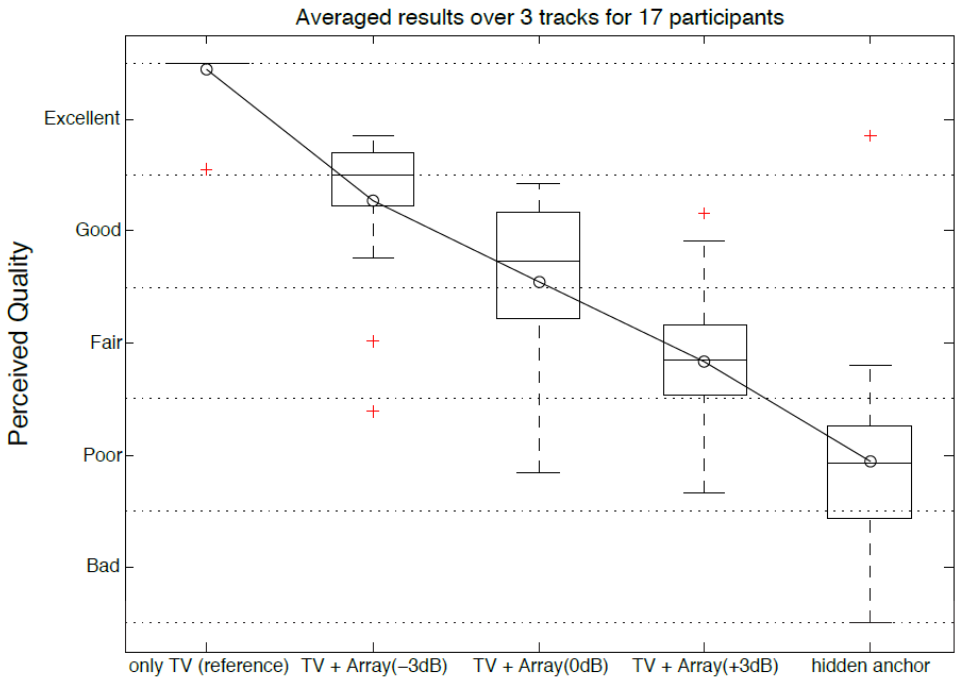


Figure 6.24: Perceived audio quality as a function of the array gain. Results averaged over all 3 tracks presented to the 17 participants. Reproduced from [106].

A repeated measures ANOVA was conducted to determine whether there was a statistically significance in the perceived sound quality for the five different levels. The data sets for each of the conditions were tested to be normally distributed by a Shapiro-Wilk test ($p > 0.05$). The assumption of sphericity had been violated as assessed by Mauchly's test, $\chi^2(9) = 17.205$, $p = 0.048$, therefore, a Greenhouse-Geisser correction was applied ($\epsilon = 0.661$). The result of the repeated measured ANOVA showed that the perceived sound quality was statistically significantly different for different levels fo the independent variable [106]. The means and standard deviation for each conditions are presented in Table 6.6.

The evaluation of the quality showed that the perceived sound quality is statistically significant($p < 0.0005$), decreasing as the gain gets increased. Furthermore a linear relationship between increased gain and perceived reduction of the quality was found. From "Array ON -3 dB" to "Array ON +3 dB" the perceived sound quality drops almost linearly, with the quality reduced a 5% for each dB

Table 6.6: Mean and standard deviation (SD) of the perceived audio quality for the different gain configurations in the dark zone.

	Array OFF	**Array ON -3 dB**	**Array ON 0 dB**	**Array ON +3 dB**	**Anchor**
Mean, (dB)	100.0	79.9	64.9	50.4	27.7
SD, (dB)	0.0	5.9	11.4	10.9	9.7

that the gain is increased. The quality is on average rated as "excellent" with an array gain of -3dB, dropping to "good" when the array gain is 0 dB and to "fair" at an array gain of +3 dB.

6.4 Conclusions

This chapter first considered the equalisation of the array with respect to the TV sound system. This equalisation has been performed since the array has to enhance the speech content of the bright zone, where the hearing impaired listeners are, without disturbing other healthy hearing listeners present on the dark zone. A reference set of filters has been created, in which the pressure produced by the array in the dark zone is equal to that produced by the TV speaker also in the dark zone. Two additional sets of filters have been created, one set with the array gain being -3 dB lower than the reference set, and a second set in which the gain is increased +3 dB from the reference set. The equalisation performed requires of the a priori knowledge of the array and the TV loudspeaker transfer functions. This allows to maximise the accuracy of the equalisation and hence the speech intelligibility enhancement. Looking towards a commercial application, this can be performed by using a remote control with a microphone which transmits the information of both responses back to the array via bluetooth.

Using these sets of filters, the assessment of the speech intelligibility enhancement provided by the array has been performed via the speech tranmission index (STI). The speech intelligibility performance of personal audio systems has been modelled and it has been analysed how sensitive the response is to background noise and observation distance. The effect of the control algorithm has shown that both acoustic contrast maximisation (ACM) and least squares inverse filters (LSIF) offer a similar STI figure. It has been also observed that the STI is maximised when the array is able of achieving control in the 8 kHz octave band, which places the inter element distance to be around 30 mm. The STI has then

been used to assess the intelligibility enhancement introduced by the array when this works together with the audio system of the TV. For a given SNR, the results predict an improvement on the STI of 0.1 if the gain of the array is selected to be -3 dB, 0.18 if the gain of the array is 0 dB and 0.26 if the gain of the array is 3 dB. This corresponds to a change in speech intelligibility from "fair-good" to "good-excellent" [42] when the array is turned ON with 0 dB gain.

The objective predictions of intelligibility enhancement are based in normal hearing systems. In order to check the potential use of the array to enhance the intelligibility of hearing impaired listeners, subjective listening tests have been carried out. Listening tests have also been used to assess how the quality of the audio perceived in the dark zone is modified when the array is switched ON. The evaluation in hearing impaired listeners has been performed by processing the received audio signals using a hearing loss simulator. It is found that for the same SNR, the percentage of recognised words can be increased a 25% if the gain of the array is selected to be 0 dB or a 38% if the gain is selected to be +3 dB. This suggests that a hearing impaired person sitting in the beam will be able to understand considerably more content of a TV programme using the loudspeaker array rather than with a normal TV speaker. The obtained results match very closely those predicted with the STI, which additionally corroborates to the effectiveness of the STI as an intelligibility prediction method. On the other hand, the evaluation of the audio quality based on the standard ITU-R BS.1534-1 (MUSHRA), showed that the perceived sound quality in the dark zone could be compromised through the use of the loudspeaker array. Nevertheless, it was found that when the array is equalised to produce the same level as the TV speaker in the dark zone (0 dB gain) the quality is still perceived as "good" using a five point scale divided into "Excellent", "Good", "Fair", "Poor" and "Bad".

It has thus been seen that it is possible to find a reasonable trade-off between the speech intelligibility enhancement for a hearing impaired person in the bright zone, and the perceived sound quality for a normal hearing person in the dark zone.

[1] levels in the dark zone between TV and array, $10\log_{10}\left(\frac{p_{D,Array}}{p_{D,TV}}\right)$.

Chapter 7

Towards improving speech intelligibility for the hearing impaired listeners

Despite the fact that the STI gives a reasonable measure of the speech intelligibility for the normal-hearing listener, it does not take into account the multiple effects of hearing impairment. Although behavioural, psychoacoustic, data can be used to quantify some of these effects, it may be more efficient to model them in terms of their physiological origin. A model of the hearing system which is based on the physiology and dynamics of the cochlea can reproduce many of the effects of hearing impairment. Using such a model, it is possible to begin to understand how diverse factors as sensitivity, frequency selectivity and upward masking thresholds are modified with hearing loss. By understanding how the nature of the hearing loss is generated, it may then be possible to design a better amplification procedure to recover speech intelligibility.

This chapter introduces a cochlear model which is based on the physiology of the organ of Corti. The formulation corresponding to that model is detailed on Appendix H. The cochlear model is then used to produce psychoacoustic factors, such as tuning curves and upward masking patterns. The model is also used to simulate the hearing impaired cochlea, by adjusting the threshold velocities of the BM along the length of the cochlea. By obtaining hearing impaired masking patterns, it may be possible to find the amplification which minimises the auditory masking and maximises speech intelligibility in a hearing impaired listener. A study of how masking can be beneficial for personal audio applications is also presented.

7.1 Reproducing cochlear responses

7.1.1 Cochlear damaging

Since speech intelligibility is reduced with increasing hearing loss, it is important to understand the physiological processes that lie behind hearing impairment. If the source of the hearing loss and its various pyschoacoustic effects are better understood, this may lead to improvements in the way that speech intelligibility can be recovered.

Hearing impairment can occur due to various mechanisms. In the peripheral and outer hearing systems, it is *conductive* hearing loss which involves a reduction in the transmission of the sound through the outer and/or middle ear. The hearing loss which accompanies ageing is mostly related with the degeneration of the amplifying properties of the cochlea. This is known as *cochlear hearing loss* and involves the damage of the structures inside the cochlea [2]. Degeneration in the neural pathways and of cognitive function will also make a contribution to hearing loss.

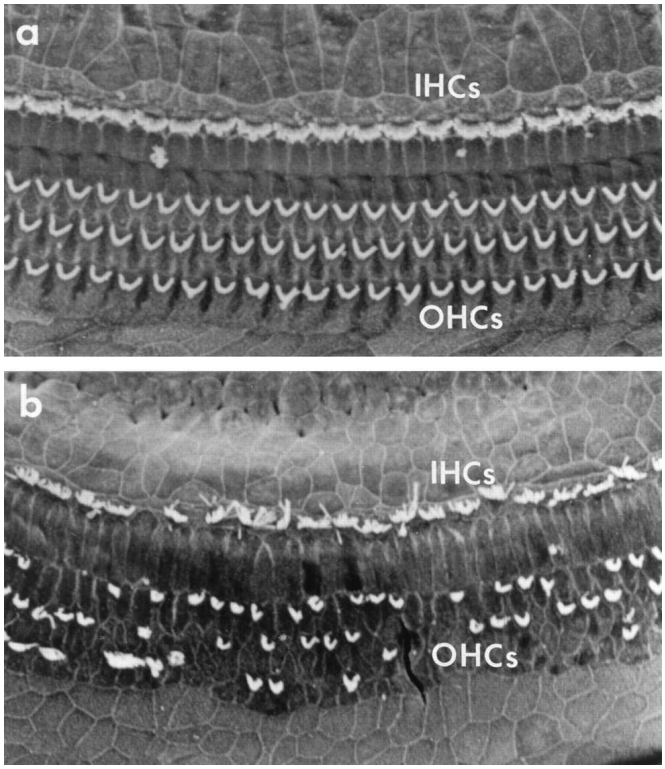


Figure 7.1: Electron micrographs of the normal (a) and damaged (b) cochlear sensory epithelium. In the normal cochlea, a row of IHC and three rows of OHC are presented. In the damaged cochlea, hair cells are missing, and stereocilia are abnormal, leading to hearing loss. Reproduced from [181].

As explained more in detail in Appendix H, the cochlea has two types of hair cells that transform mechanical motion into electrical signals. Each of these hair cells experiment motion by the deflection of the stereocillia that protrude from the top of the cells, whose motion opens ion channels that allow a current to flow into the cell. The 3000 or so inner hair cells (IHCs) convert motion within the organ of Corti at different places along the cochlea into neural signals that are passed up the auditory pathways. The main function of the 12000 or so outer hair cells (OHCs), however, is to act as local positive feedback loops and amplify motion within the organ of Corti at low sound pressure levels.

Cochlear hearing loss can occur in two ways, due to damage to the stereocillia of the OHC or IHC. A micrograph showing damaged IHC and OHC is shown in Fig. 7.1, in comparison with healthy stereocilia. Damage to OHC affects the active cochlear amplification, which results in a lower BM vibration level for a given stimulus. Damage to IHC results in an inefficient transduction, as the IHC are the cells which trigger the nerves that send messages to the brain. A common aspect of both mechanisms is that the amount of BM velocity required to reach a certain threshold is larger than for a normal cochlea. Not only the level of perception is reduced, but the frequency discrimination of sounds is also affected.

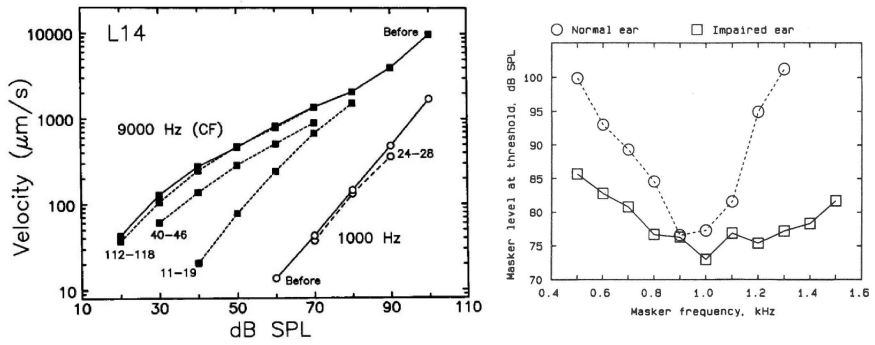


Figure 7.2: The left hand side plot shows the Input-output relation on the BM before (solid lines) and after (dashed lines) a furosemide injection. The right hand side plot shows the psychophysical tuning curves determined in simultaneous masking for the normal ear (circled and dashed lines) and the impaired ear (squares and continuous lines) of a subject with unilateral hearing loss. Reproduced from [2] and [182].

Damage to the OHCs mostly affects the non-linear cochlear amplifier (CA). Fig. 7.2 shows how the non-linear relationship between input and output is affected in a damaged chinchilla cochlea [182] for two different sinusoidal excitations. The solid curves represent a healthy cochlea (before), whilst the dashed lines are measurements after the animal has been injected with furosemide[1], which impairs

[1]Furosemide is a loop diuretic which is used for the treatment of hypertension and edema,

the action of the OHCs. Such loss of amplification is analogous to that found in aged listeners, where the degeneration of the stria vascularis tissues with age also involves a reduction in endocochlear potentials. The malfunctioning of the CA also involves reduction in frequency selectivity, causing the frequency responses at a single point along the cochlea to be broader than normal. The reduction in selectivity causes a point along the cochlea to respond to tones of frequencies which are different to that of the characteristic frequency of such point. Fig. 7.2 also shows the measured psychophysical tuning curves[2] [183], for normal hearing and hearing impaired subjects. For an impaired cochlea the curve is much broader, and the level of the masker is lower, as a single point along the cochlea responds to a broader range of frequencies around the characteristic frequency.

7.1.2 A model of the organ of Corti

Attempts to model the mechanics of the cochlea appeared soon after Von Békésy [184] showed that sound waves of different frequencies excited different nerves in the inner ear. Over the last few decades, a large number of models based on filter banks have been developed to analyse and study the psychoacoustical processes of hearing. These models are typically based in combinations of gammatone or all-pole gammachirp filters [185] to mimic the response of the processing stages of the auditory systems. Such models are able to replicate the key features of the auditory system, such as tuning, critical bands compression and frequency- and level-dependent responses [186]. These models, however, reproduce the overall psychoacoustic responses without directly linking these individual filters with the fundamental biophysical details and undergoing cellular sub processes.

Other physically-based modelling approaches try to reproduce the mechanisms of the actual processes which happen inside the cochlea, more exactly inside the organ of Corti. Such models are based on a mechanical or electroacoustical representation. Mechanical models of the inner ear or those based on electrical resonant systems work by tuning the parameters of a set of oscillators to match the frequency response of the real cochlea. As a mechanical system, the mammalian cochlea shows great resolution both in frequency and in time [187], which suggests that some active amplification or active tuning is required. The discovery of the OHC motility [188] led to the assumption that the OHCs were the active transducer of the CA.

however, it also reduces the endocochlear potentials.

[2]These curves show how much the level of a narrowband masker noise has to be to mask a given probe tone [2].

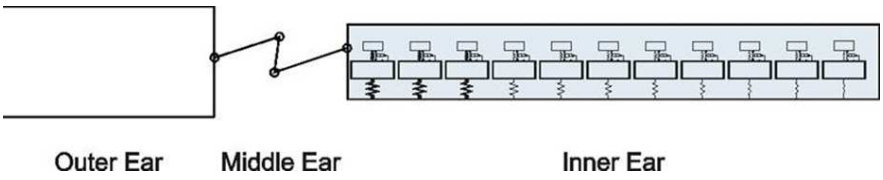


Figure 7.3: Lumped parameter representation of an uncoiled cochlea. The cochlea is divided into N resonant systems with a feedback loops emulating the action of the CA. Whilst the resonant systems are not coupled between them, they are coupled via the fluid of upper and lower cochlear chambers. Figure reproduced with permission from [189].

Models of the cochlear mechanics are often based on arrays of mass-spring damper systems with active feedback loops, all coupled through the fluid chambers of the cochlea, which are often assumed to be uncoiled, as shown in Fig. 7.3 [188]. These models often use lumped parameter models of the OC, so that its complicated structure is modelled by a simple combination of masses and springs. The drawback of many of these models is that the value of the parameters (mass, stiffness, damping) cannot be directly related to measured physical parameters, due to the simplification of the complex cell structure on top of the BM. More information about the underlying mechanism and simplifications of such models is found in Appendix H, from which the three degree of freedom lumped parameter model shown in Fig. 7.4 is taken. In this model m_1, m_2 and m_3 represent the lumped mass of the basilar membrane, reticular lamina and tectorial membrane at a position along the cochlea.

This lumped parameter model of the organ of Corti is thus physically realisable, and at the same time offers an accurate reproduction of the frequency responses and psychoacoustic factors estimated in measurements. The model has been adapted from that originally developed by Ramamoorhty *et al.* [190]. This model was originally developed using finite elements, and was tuned to model a guinea pig cochlea. The model has been converted into a lumped-parameter model by Simón Gálvez and Elliott [191]. The formulation of this model, and the tuning needed to reproduce human responses is presented in Appendix H.

The model introduced here is linear, but the effect of different levels of excitation can be reproduced using different values of the OHC feedback gain, γ, which drive the CA. The response of the model in terms of the velocity of the BM with various gains of the CA used here, at a single point along the cochlea, is shown in Fig. 7.5. The amplification of the model grows as the gain of the CA is increased. A small reduction in response is found just below the frequency at which the BM

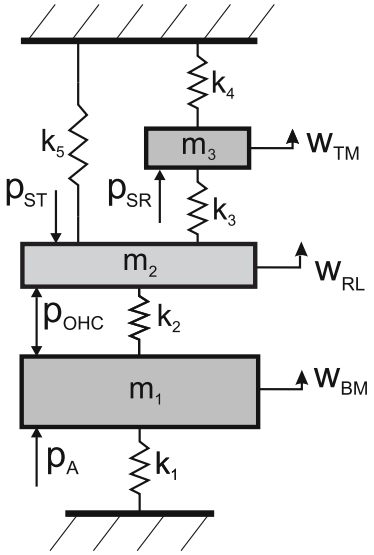


Figure 7.4: Lumped parameter representation of the organ of Corti model, which is detailed in Appendix H. The mass m_1 represents a cross section of the BM. The BM is driven via the fluid pressure difference, p_A, and the active pressure, p_{OHC}. The mass m_2 represents a cross section of the RL and the mass m_3 represents a cross section of the TM. The feedback pressure p_{OHC} is proportional to the hair bundle displacement, given by the relative motion between RL and TM.

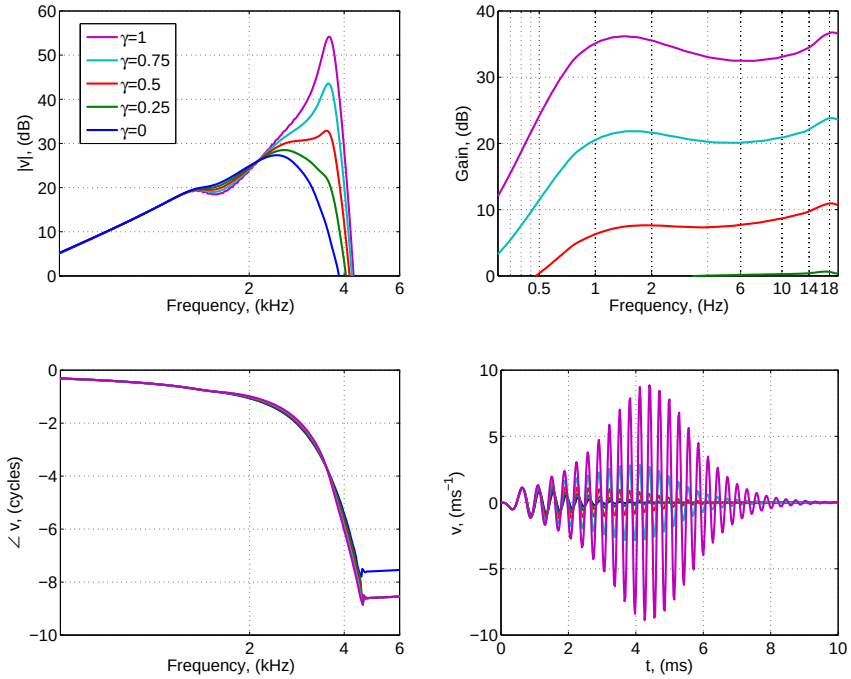


Figure 7.5: Frequency response of the cochlea, amplification with respect to a passive case and impulse responses for different gains of the cochlear amplifier. The frequency and impulse responses are shown for a position 15.3 mm along the 35 mm length of the cochlea.

velocity response starts to build up, as also observed in Fig. 7.6. Whilst in the passive impulse response (IR) the maximum of the envelope is found at shorter latencies, the energy tends to build up later as the gain of the CA is increased. At the same time, the BM fine-structure is near invariant with stimulus level, which

is in qualitative agreement with numerous studies [192, 193, 194]. As observed in the figure, the amplification with the gain of the CA set to a maximum, $\gamma = 1$, goes from about 30 dB at 600 Hz to 38 dB at high frequency, which is close to the results reviewed by Robles and Ruggero [187]. Both the passive and fully active frequency response of the model at various points along the cochlea can be observed in Fig. 7.6.

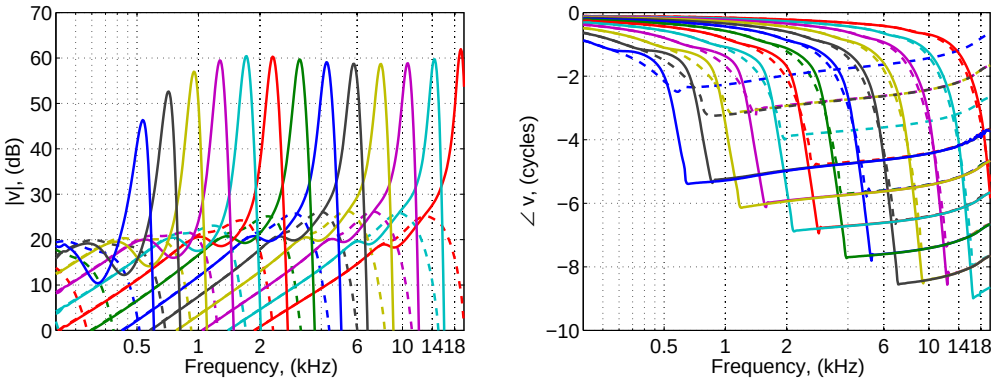


Figure 7.6: Frequency response of the coupled passive (dashed lines), $\gamma = 0$, and fully active (solid lines), $\gamma = 1$, cochlea. Response shown between x=2.1 mm and x=34.9 mm, in increments of 2.1 mm.

The tuning of auditory filters is normally expressed in terms of equivalent rectangular bandwidth (ERB), which is related to the critical bandwidth [195]. The ERB corresponds to the bandwidth of a perfect bandpass filter which has the same white noise power output as the specified filter. The results presented here are plotted as a function of Q_{ERB}, which is defined as

$$Q_{ERB} = \frac{f_C}{ERB(f_C)}, \tag{7.1}$$

where f_C represents the characteristic frequency, i.e, the frequency at which a point of the BM is more sensitive.

Fig. 7.7 shows the quality factor of the cochlear filters, in Q_{ERB}, of the model when working in fully active mode, $\gamma = 1$. The model gives a higher Q_{ERB} than that observed in psychoacoustic experiments using notched-noise masking [196], but lower than other estimates, based on measurements of phase-gradient delay of stimulus-frequency otoacustic emissions [197, 198].

Attempts to find parameters that give the model a sharper tuning, led to the response of the model becoming unstable. It may be that with longitudinally coupling in the BM and TM, as used by Meaud and Grosh [199] and Cantrel and Elliott [200], a higher stability limit can be obtained, and the responses can be

tuned to give a higher Q_{ERB}.

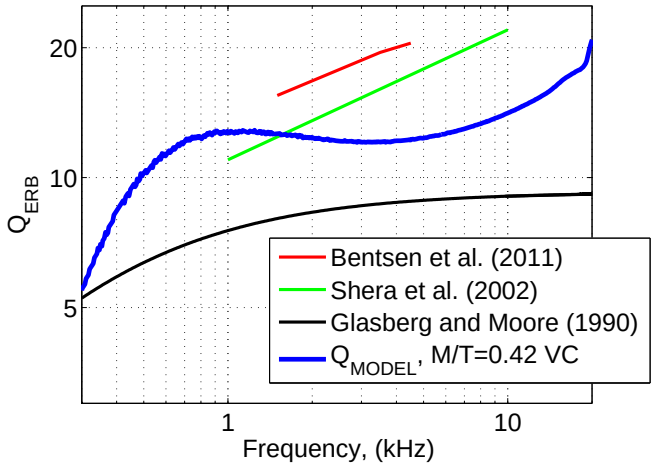


Figure 7.7: Auditory filter quality factor, Q_{ERB}, given by the model of cochlear mechanics used here. Estimated Q_{ERB} curves from experimental studies are also shown, including Bentsen *et al.* [197], Shera *et al.* [198] and Glasberg and Moore [196].

7.2 The modelling of cochlear hearing loss

Cochlear models were originally used to reproduce hearing impaired cochlear responses by the current author, by studying the variation of frequency selectivity of the cochlear filters with hearing loss [201] using the 1986 Neely and Kim model [202]. Previous research using models of normal hearing has also predicted upward masking patterns [203]. Verhulst *et al.* developed a model to reproduce transient stimulation and otoacoustic emissions [204]. The effect that hearing loss has on the distortion product otoacustic emissions (DPOE) was recently studied by Liu and Neely [205], who reduced the gain of the CA by a 50% to model a hearing loss. Other studies have used cochlear models to check the changes in temporal fine structure due to sensorineural hearing loss [206] and to improve hearing aid fitting [207]. Zilany and Bruce also considered the use of a model to look at auditory nerve (AN) responses under high level excitation for normal and impaired listeners [208]. The origins of cochlear hearing loss by studying how the reduction in endocochlear potentials (EP) affect the responses was recently considered by Saremi and Stenfel [209].

It has been demonstrated that the micromechanics of the stereocilia change during exposure to intense stimulation, which may recover shortly after the stimulation is ended [210]. If the exposure to sound is too intense or too long, however, it may lead to permanent damage or even death of the OHC itself through either apop-

tosis or necrosis processes [211], which is known as noise induced hearing loss. This medical condition appears commonly in construction and factory workers. Evidence of direct damage to OHC due to noise exposure has been observed in mice. Lim *et al.* [212] exposed a group of mice to a 122 dB SPL noise 3 hours per day during 3 consecutive days. After that, the authors performed a quantitative investigation of changes in the OHC, finding that those were very damaged at the base (about 40 % survival) and less damaged at apical positions (about 90 % survival).

Apart from exposure to high level sounds, cochlear hearing loss is also related with age. The amplifying properties of the OHC and the transduction properties of the IHC depend on a voltage which is intrinsic to the cochlea, known as endocochlear potential (EP) [213]. The *stria vascularis* is a vascularised tissue on the lateral wall of the cochlear duct, which injects potassium ions (k^+) into the scala media (SM). Such potassium ions are linked to the 89 mV EP along the cochlear duct [214]. Because of the age related changes in these cells, the k^+ recycling is impaired and the EP potentials are no longer maintained [215, 216, 217]. As the transduction properties of the IHC are reduced, they are less sensitive to sounds, increasing hearing thresholds. Various studies have observed the role of the IHC ions channels on the temporal structure of auditory information in the mice cochlea, observing that deficiency in the IHC mechano-electrical transduction (MET) channels decreased the IHC receptor current, with consequent reduction in spike timing and firing rate [218, 219]. These authors concluded that this may have a disadvantageous effect on the detection of signals in noisy environments. Apart from the reduction of current intensity through the transduction channels, the reduction in EP also affects the transduction of sound, due to the decrease in the amplification provided by the OHC. The reduction of EP due to the atrophy of the stria vascularis and its implication in reduced action of the CA has been shown in numerous studies. Gates *et al.* [220] measured the effect in the hearing threshold and in DPOE, showing an effective degeneration with age. Experiments in aged animals also showed a decrease of the cochlear amplification action [221].

A graphical example of the effect of different causes of cochlear damage is shown in Fig. 7.8, where the mean audiogram of a group of elder people is compared with the audiogram of young listeners with a noise induced hearing loss. In the figure it is possible to observe how the normal hearing listener presents a very small deviation from the 0 value in dB HL. The noise induced hearing loss listeners show almost intact audition at low frequencies, however as the frequency

is increased the hearing loss becomes greater, which is thought to be primarily due to damage to OHC. The listeners with presbyacusis additionally present a hearing loss at low frequency, due to the reduction of EP which affects the IHC, as well as a more significant hearing loss at high frequency, due to the reduction in amplification of the OHC.

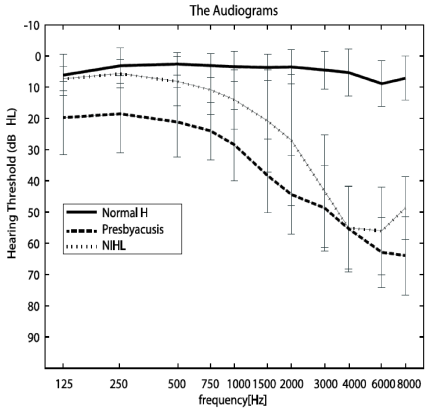


Figure 7.8: Averaged audiometric hearing thresholds for three groups of subjects: "Normal H", normal hearing subjects aged between 30 and 55, "Presbyacusis", subjects aged between 65 and 70, and "NIHL" listeners, aged between 43 and 57. Reproduced from [222].

7.2.1 The dynamics of the hearing impaired cochlea

In order to obtain cochlear responses which match those of hearing impaired subjects, an iterative procedure has been used, as previously developed in [201]. The program finds the active gain distribution along the cochlear length that minimises the error between the peak BM velocity, calculated with a given set of active gains, and a target velocity distribution of the cochlear partition, assumed to be corresponding to that of older subjects of a certain age. Such velocity distribution represents only the hearing loss (HL) due to malfunction of the OHC.

The distribution of HL over frequency given by the standard ISO 7029 *Statistical distribution of hearing thresholds as a function of age* [12] has been used to define the target loss of sensitivity. This standard gives an estimation of the HL assumed to occur in different percentages of the population due to ageing. The values of HL estimated to occur for 50 percent of the female population are used here, as shown in Fig. 7.9.

It is assumed here that, for each age, the HL at 125 Hz are caused by a reduction of the IHC transduction properties, and that this is constant all along the BM.

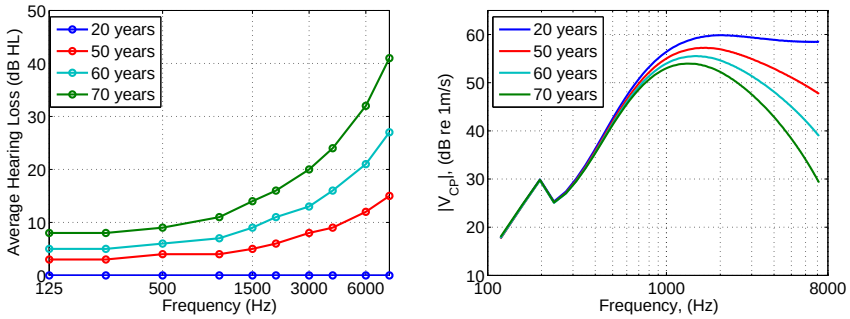


Figure 7.9: Median HL expected in 50 % of females at different ages, as given by the standard ISO 7029 (left), and the corresponding peak BM velocity envelope distributions, calculated from the envelope for the active cochlea (20 years), minus the HL assumed due to the OHCs at different ages (right).

The HL due to OHC are thus assumed to be that shown in the left hand side plot of Fig. 7.9 minus the amount at 125 Hz.

The changes to the envelopes of the peak BM velocity, are assumed to occur only due to OHC malfunction, as shown in the right hand side plot of Fig. 7.9. These velocities have been obtained by assuming the output of the fully active model to be that of a 20 years old cochlea, v_{20}. The velocities corresponding to other degrees of hearing impairment are reduced according to the dB HL which correspond to a certain age

$$v_{HL} = v_{20} 10^{-HL/20}, \tag{7.2}$$

where HL stands for the dB HL of each audiogram. The distribution of peak velocities is obtained up to 8 kHz, which corresponds to the highest frequency presented in the standard ISO 7029 [12].

Whilst the HL due to the OHC are clearly reflected in a reduction of the BM velocity, this is not the case for the HL due to IHC damage, where it is the measured response which is reduced [2]. Some studies have contemplated modelling IHC damage by changing the slope of the IHC MET function [208, 209]. The reduction of auditory nerve spike rate due to the IHC malfunction is modelled in these papers as a level-independent damping component, which is constant along the whole length of the cochlea, as previously used by other researchers [223, 207].

A first attempt at reproducing hearing loss with the model was made by adjusting the CA gain to match the HL at each frequency, assigning this gain value to the cochlear point with a characteristic frequency (CF) as that of the excitation. When the coupled response of the cochlea was then calculated, however,

the model was found to be unstable. A smooth decrease on the cochlear amplifier gains between adjacent elements is needed in order to assure stability [224]. This lead to the use of a polynomial function, for the OHC gain, γ, as a function of position along the cochlea, which provides a smoother variation. Fig 7.10 shows the reduction in BM velocity with the CA gains calculated to obtain the results of Fig. 7.9.

The gains for the active amplifier have been obtained by fitting a 4^{th} polynomial. The coefficients of such polynomial have been optimised until the envelope of the responses showed a reasonable agreement with that of Fig. 7.9, leaving mean errors of 0.2, 0.6 and 0.7 dB for the 50, 60 and 70 year old cochlea respectively. Other trials were also performed with polynomials of lower and higher order, but the most reasonable results have been obtained with a 4^{th} order approximation. The results are based on median HL up to only 8 kHz, but the gain curves continue to decrease towards more basal locations, as shown in Fig. 7.10. Such distribution is consistent with studies performed in other mammal species. Previous research in mice exposed to high intensity sounds have shown the OHC at basal conditions to be strongly damaged, with the OHC at apical conditions to be almost intact [212, 225]. The same pattern has been observed in studies with aged gerbils [221].

The results of Fig. 7.10 show how the higher the hearing loss, the greater the reduction in CA needed to replicate such losses. For the case of 70 years old, the gain has been set to be 0 for positions closer than 5 mm to the base, as is expected to happen in a human cochlea with a severe hearing loss. The figure also shows the responses for positions along the cochlea at x=6.8 mm, x=10.9 mm, x=15.0 mm, x=19.1 mm, x=23.2 mm, x=27.3 and x=31.4 mm. Whilst for the case of 50 years old there still some active response at 12 kHz, for the 70 years old case the response at positions close to the base is mainly passive.

7.2.2 Modelling of hearing impaired psychoacoustic factors

Frequency selectivity in the hearing impaired

As well as the sensitivity reduction, the decrease of amplification leads to a broadening of the cochlear responses, with a corresponding loss in frequency selectivity. The loss in frequency selectivity can be quantified by calculating the sharpness of the equivalent cochlear filters, generated for various degrees of HL.

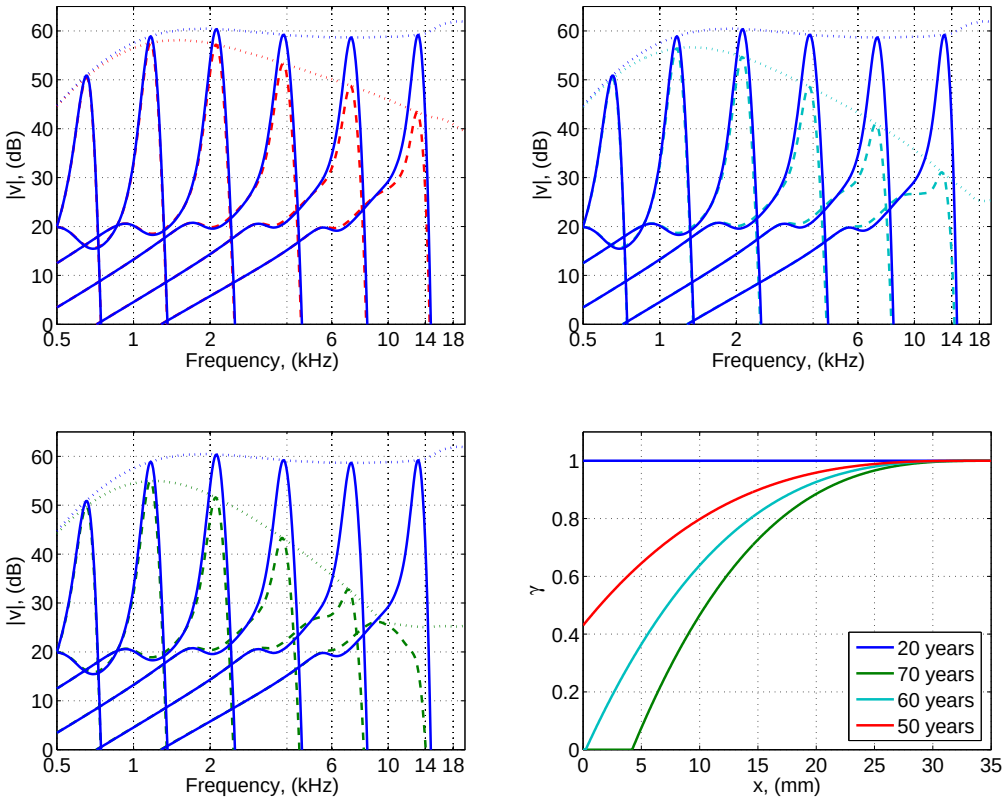


Figure 7.10: The solid lines show the BM velocity responses against frequency calculated for a 20 year old cochlea (solid lines) and adjusted to approximate a 50 year old cochlea (top left), a 60 year old cochlea (top right) and a 70 year old cochlea (bottom left). The responses are calculated at positions at x=6.8 mm, x=10.9 mm, x=15.0 mm, x=19.1 mm, x=23.2 mm, x=27.3 and x=31.4 mm from the base. The dotted line represents the envelope of the response, as in the right plot of Fig. 7.9. The optimised distributions of cochlear amplifier gains, used to obtain these results, are plotted as a function of position along the cochlea at the bottom right graph of the figure.

A common method of observing the cochlear selectivity is by plotting tuning curves, which show the sensitivity over frequency of a certain point along the cochlea [2]. Using the distributions of gains of Fig. 7.9, tuning curves have been calculated at a distance of 8.5 mm from the base. The BM velocity for the same stapes input of a 20 year old listener has been considered as the threshold. The tuning curves, shown in Fig. 7.11, present a reduction in sensitivity and selectivity which increases proportional to the age of the simulated listener. Whilst for the 20 year old listener the response shows a clearly defined peak, the 70 year old cochlea show a clearly passive response, in which the response at other frequencies rather than the CF is also reduced, due to malfunction of the IHC MET transduction. These results are in accordance with studies using ototoxic drugs [226] and with behavioural measurements [222].

Apart from assessing the tuning of the cochlear filter corresponding to a certain

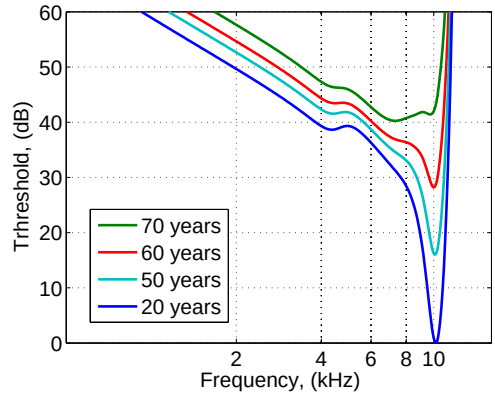


Figure 7.11: BM tuning curves associated with the HL given by the standard ISO 7029 at a position 8.5 mm from the base.

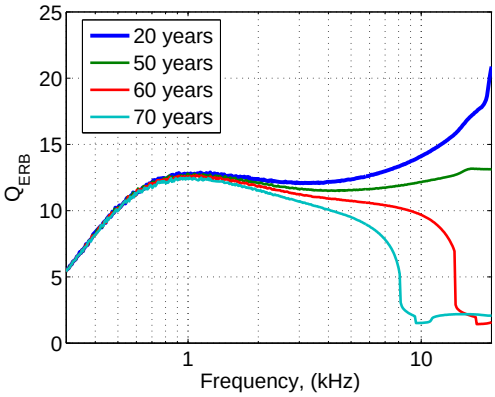


Figure 7.12: Estimated Q_{ERB} against frequency for different degrees of HL.

position, it is interesting to observe how this varies along the length of the cochlea. Fig. 7.12 shows the estimates of Q_{ERB} as a function of frequency that are obtained for various degrees of HL. The loss of frequency selectivity due to the increase in bandwidth of a cochlear filters and the corresponding reduction in Q is clearly seen in Fig. 7.12. Following the reduction in cochlear gain, the broadening of the responses is more severe at high frequency. For the 50 years old cochlea, the Q_{ERB} value is of about 12-14, whilst that for the 70 years old cochlea, the Q_{ERB} value above 8 kHz corresponds to that of a passive cochlea.

Prediction of upward masking thresholds

Masking is the process by which one sound alters the perception of another. Although many types of masking have been investigated, this thesis focuses on the simultaneous masking of one tone with another. Masking is generally thought to be due to two processes. The first, sometimes termed suppression, is caused by the non-linearity of the BM response within the cochlea. This is modelled in this chapter by incorporating a non-linear gain in the CA, that depends on the

amplitude of the relative displacement of the IHC, which is proportional to the input sound pressure level. The second process that contributes to masking is due to the non-linearity in the firing of the nerves excited by the motion of the BM, and is sometimes called swamping [5]. This second effect is not modelled here.

Such masking curves were measured in an early series of listening experiments by Fletcher [227]. In these experiments a set of reference masking tones were played at constant volume and frequency. For each of these tones, the listeners adjusted the level of a probe tone until this could just be heard. The amount of amplification needed for the probe tone to be heard is the threshold shift. Fig. 7.14 shows the result of the experiment conducted by Fletcher, for masker tones between 200 and 3500 Hz and levels from 20 to 100 dB SPL [227].

Although many other authors have performed similar experiments since Fletcher, as discussed in [228], the results are generally of the same form. Idealisation of the masking curves are used, for example, in the MPEG standard for MP3 coding [229] in order to calculate the minimum number of bits possible per sub band for a given background noise level, or in the calculation of the STI [42].

In the simulations presented here, the BM velocity at the threshold of hearing is used as a measure of audibility. This seems a sensible assumption, as the threshold of hearing denotes the point at where a tone becomes audible. The SPL for the absolute threshold of hearing in quiet is well approximated at each probe frequency by the non-linear threshold function [230]

$$T_Q(f) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5e^{-0.6\left(\frac{f}{1000}-3.3\right)^2} + 10^{-3}\left(\frac{f}{1000}\right)^4, \text{ (dB SPL)}. \quad (7.3)$$

The output of such formula is shown in Fig. 7.13a together with the minimum audible pressure (MAP) [2].

Although the model of the cochlea used here has a linear response over gain, its gain can be made level dependent. In this case, γ, the parameter which controls the gain of the CA, is made proportional to the IHC displacement, w_{ST}, which is proportional to the input level, L_{IN}. The active gain as a function of γ can be seen in Fig. 7.13b. The CA gain as a function of the L_{IN} has been adjusted with the results shown by Elliott and Shera [189]. To this end, a first order Boltzmann

function has been used so that γ and L_{IN} are related by

$$\gamma = \frac{1}{1 + 1.6e^{0.0941 L_{IN}}} \tag{7.4}$$

as shown in Fig. 7.13c.

The compressive non-linearity resulting from Eq. 7.4 is shown in Fig. 7.13d, which is calculated for a healthy cochlea and for hearing impaired cochleas according to the hearing loss of Fig. 7.9. The curves which correspond to the hearing impaired cochleas introduce a reduction in the low-level gain, in concordance with those shown by Ruggero and Rich [182].

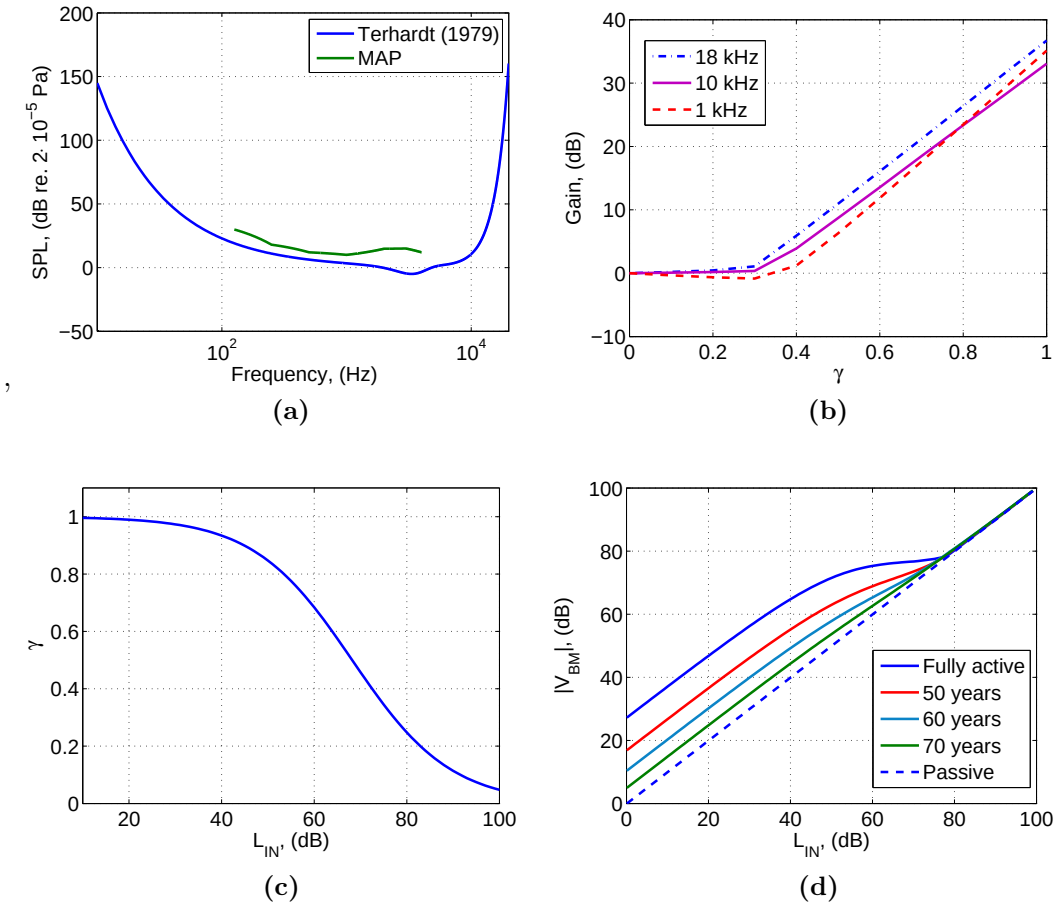


Figure 7.13: **(a)** Absolute threshold of Hearing, as given by [230],"Terhardt (1979)", and Moore [2], "MAP"; **(b)** The gain of micromechanical model as a function of γ; **(c)** The gain of the CA as a function of the input level for the fully active cochlea; **(d)** effect of reducing the gain of the CA simulating OHC malfunction on the compressive non-linearity in the BM response.

Before predicting the masking for the hearing impaired, the model has been tested

by calculating the masking curves obtained by Fletcher [227] for normal hearing subjects, as shown in Fig. 7.14. The model has been first run with a pressure at the ear equal to each of the masker tones, L_M. In order to obtain the corresponding stapes excitation, a combined outer-middle ear model has been used, which is defined in Appendix I. The stapes velocity at each masker tone L_M is denoted as $u_M(ST)$. For a given $u_M(ST)$, the basilar membrane peak velocity along frequency is recorded, denoted as $v_M(BM)$. The model is run again with a pressure input at the ear equal to that of the hearing threshold in quiet, so that it generates a certain velocity at the stapes, denoted as $u_Q(BM)$. The BM peak velocity is then recorded, denoted as $v_Q(BM)$. At each frequency of analysis, $u_Q(ST)$ is increased until the peak velocity of the BM is equal to that if being excited by $v_M(BM)$. The resulting stapes velocities are denoted $u_P(ST)$. Using the outer-middle ear transfer function defined in Appendix I, the obtained stapes velocities are converted back into acoustic pressure levels at the ear input, L_P. The level of the absolute threshold in quiet is subtracted from each of the L_P pressure levels, so that the threshold shift curves are obtained. The resulting threshold shift curves are presented in Fig. 7.15, and show a reasonably agreement with the original results given by Fletcher [227]. The fine structure of the predicted masking curves is different from that of Fletcher and its is probably due to various details in the model.

The incorporation of hearing loss in the CA gain allows the prediction of upward masking curves for the hearing impaired. This may be useful in the fitting of hearing aid devices or in the modification of the STI calculated for hearing impaired listeners. Fig. 7.17 shows the result of driving the cochlear model with masker tones at 2.4 kHz, 3.5 kHz and 4.5 kHz. The hearing impaired gain distributions have been included to recreate the masking in cochleas of 50, 60 and 70 years old.

Fig. 7.16 shows the masked BM velocity patterns. Such patterns show how for the hearing impaired cochleas the BM velocity peak is reduced for low-level inputs (L_M <80 dB), due to the lack of cochlear amplification. The curves for high level input sounds, $L_M = 80$ dB or $L_M = 100$ dB show an identical behaviour for all ages, as the cochlea works merely passive.

The predicted upward masking curves for cochleas of different ages are shown in Fig. 7.17. The masking curves show a very similar on-frequency value, but show elevated off-frequency (higher frequencies) thresholds, which is consistent with behavioural studies [231, 232]. The proxy used in the model assumes that the

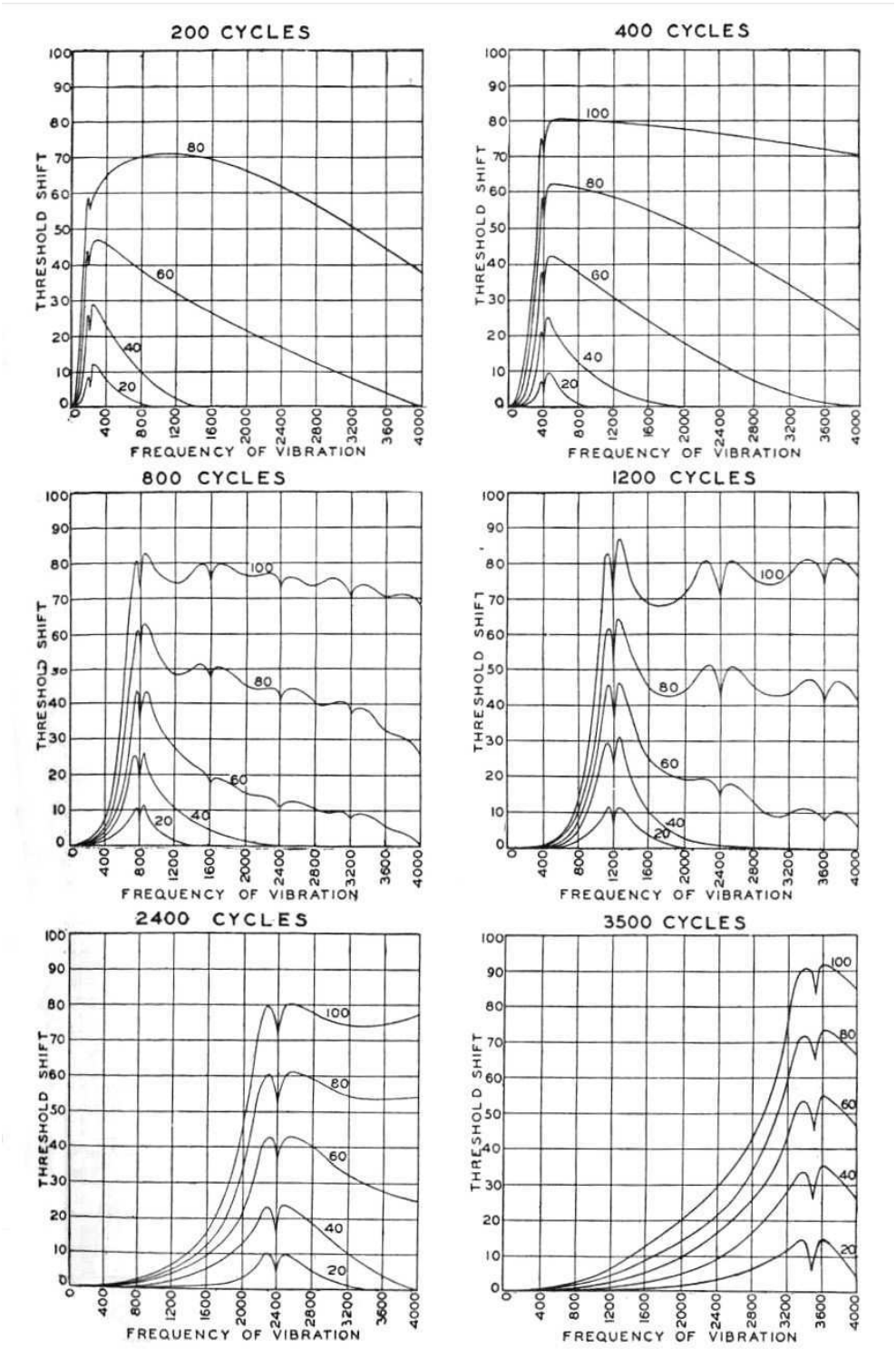
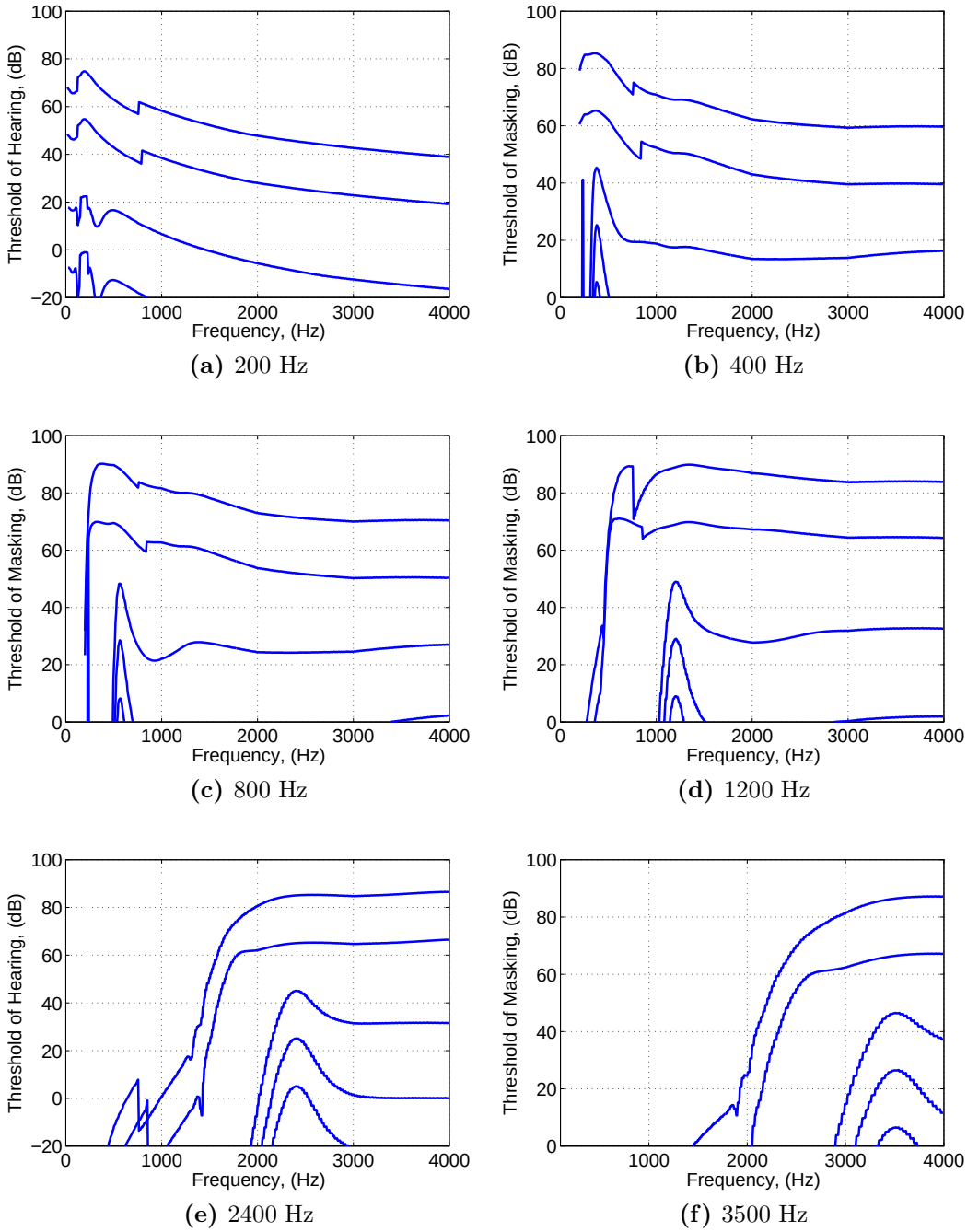


Figure 7.14: Threshold shift curves as measured by Fletcher. The reference masker tones vary between 200 and 3500 Hz, and are played at different volumes, from 20 to 100 dB. A probe tone is varied between 0-4 kHz, and is adjusted in volume until it just can be heard. The increased gain relative to the masker tone required for each case produces the threshold shift curves shown. Reproduced with permission from [227].

probe has to be as loud as the suppressor to be heard. Hence the BM velocity in an impaired cochlea, which a probe tone needs to be resolved when the BM is



(a) 200 Hz **(b)** 400 Hz

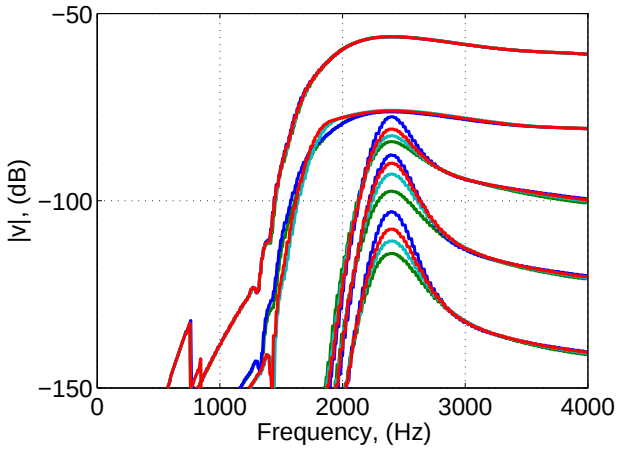
(c) 800 Hz **(d)** 1200 Hz

(e) 2400 Hz **(f)** 3500 Hz

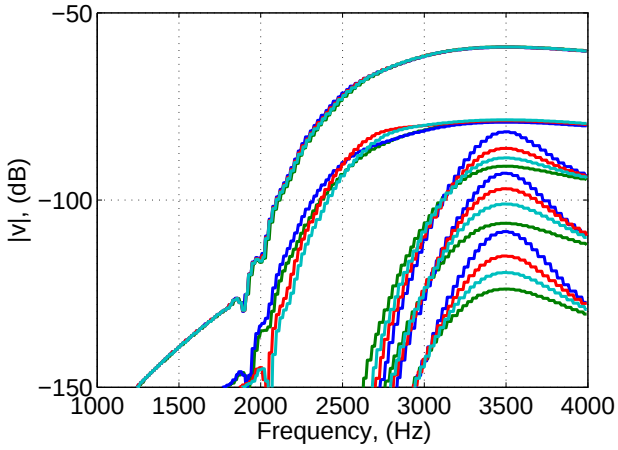
Figure 7.15: Output of the micromechanical model for the Fletcher threshold shift curves (see Fig. 7.14). The five increasing curves at each masker frequency correspond to masker tones of L_M= 20, 40, 60, 80 and 100 dB.

masked by a tone of the same frequency, is the same that the velocity needed in a healthy cochlea. At off-frequency, the masking produced by a hearing impaired cochlea is larger, as more upper frequencies are excited by the masker tone, due to the fact that the cochlear filter is broadened. Observing the results of f_M =2.4 kHz, it can be seen how the 70 years old cochlea has an off-frequency threshold increased about 15 dB more. For the masker tone of 4.5 kHz the results are

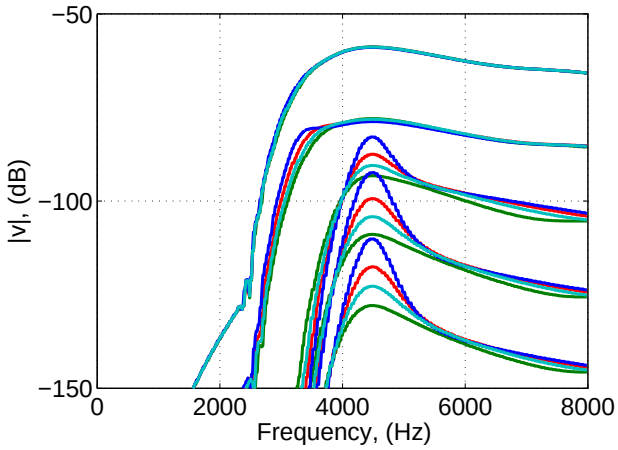
similar, with the off-frequency thresholds increased by about 20 dB. The greater the hearing loss, the greater the threshold elevation at frequencies lower than f_M, as observed in the three masker tones results, which has also been seen in behavioural studies [232].



(a) 2400 Hz

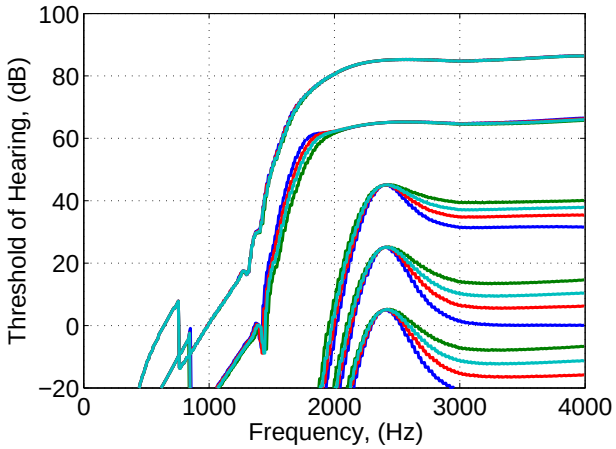


(b) 3500 Hz

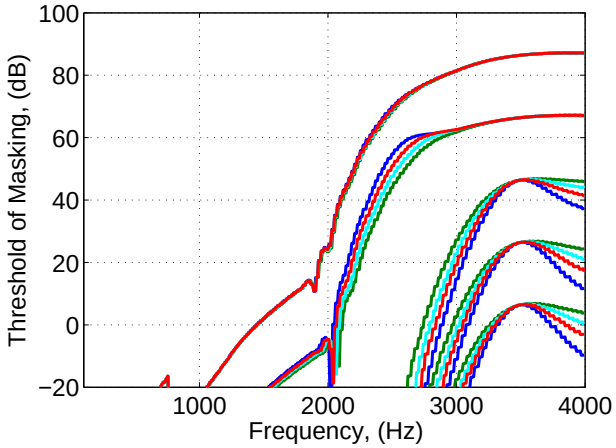


(c) 4500 Hz

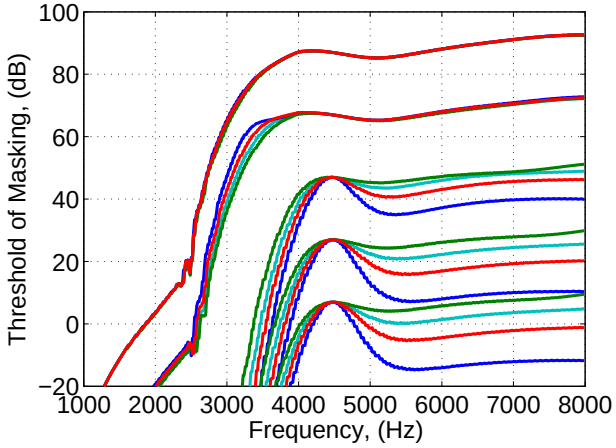
Figure 7.16: Masked vibration patterns of the BM for maskers placed at 2.4 kHz, 3.5 kHz and 5 kHz. The blue graphs show the results for a 20 years old cochlea, and the red, cyan and green show results for 50, 60, and 70 years old cochlea respectively. The five increasing curves at each masker frequency correspond to masker tones of L_M= 20, 40, 60, 80 and 100 dB.



(a) 2400 Hz



(b) 3500 Hz



(c) 4500 Hz

Figure 7.17: Hearing impaired masking thresholds with maskers placed at 2.4 kHz, 3.5 kHz and 5 kHz. The blue graphs show the results for a 20 years old cochlea, and the red, cyan and green show results for 50, 60, and 70 years old cochlea respectively. The five increasing curves at each masker frequency correspond to masker tones of L_M= 20, 40, 60, 80 and 100 dB.

7.3 The influence of masking in speech intelligibility and sound zoning

The perception of speech involves a number of acoustical, physiological and psychological processes. One property of the auditory system which is exploited in several audio and speech processing algorithm is upward masking. As shown in the previous section, one louder, low frequency sound can prevent another weaker, high frequency sound from being audible. This is desired, for example, when trying to reduce distraction with respect to external noises in an office [233], where random background noise may be introduced to improve the privacy of conversations. For other audio applications, however, noise is typically undesired. For speech intelligibility, it is clearly undesirable that the louder, low frequency background sound masks speech. Apart from that, if the program signal being reproduced has energy concentrated in the low frequency bands, this will conceal the energy concentrated on higher frequency bands from being heard. This is a situation that occurs commonly when listening to music; in this case the low frequency loud instruments mask the high frequency weak instruments, which are only heard when the low frequency instruments are not playing.

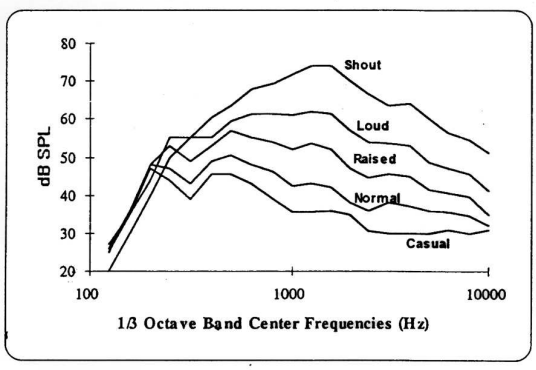


Figure 7.18: Long term average speech spectra (LTASS) at various local efforts. Reproduced from [234].

Fig. 7.18 shows the long term average speech spectrum (LTASS) presented at different levels [234]. For a normal hearing listener, the cochlea is working mainly actively for low vocal efforts, where the input of the sounds is below 50 dB SPL. If the level of the vocal effort is larger, the cochlea will be working more passively, increasing the amount of upward masking, as can be observed in Figs. 7.14 and 7.15. In this case the information introduced by lower speech formants will mask that introduced by higher speech formants. This contributes directly to the reduction of speech intelligibility, as observed for speech at large vocal efforts [29].

In a hearing impaired cochlea, this phenomena occurs at even lower input levels, as seen in Section 7.2.2.

This section studies this facet of masking, using the model of masking developed in the previous section. A set of simulations shows how by a careful equalisation of the array program spectrum, it is possible to reduce masking at high frequency and maximise speech intelligibility.

7.3.1 Spectral shaping of program for maximisation of speech intelligibility

It is known that the presentation of speech at high level reduces intelligibility [157]. Part of this effect is caused by a reduction in phonetic information due to the increase of vocal energy [235], but also because upward masking thresholds are increased and less high frequency information is processed. In a hearing impaired listener, this phenomena occurs at lower presentation levels, as the upward masking curves are largely increased.

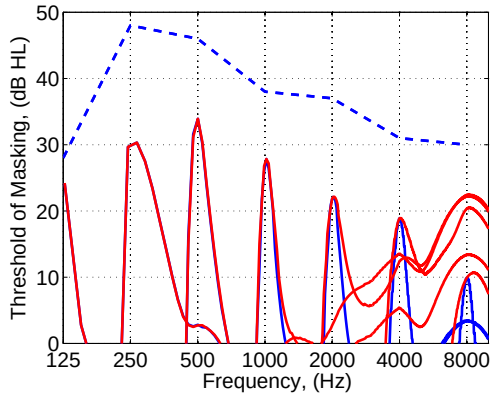


Figure 7.19: Predicted masking patterns in octave bands for the LTASS at normal effort for a 20 years old listener (blue graphs) and for a 70 years old listener (red graphs), assuming that the level in each octave band is due to a single harmonic in that band. The dashed line shows the input level of the masker in dB SPL.

The magnitude of this effect can be estimated using the hearing impaired masking model. Fig. 7.19 shows the threshold of masking for a 20 years old and a 70 years old cochlea, using the LTASS at normal effort as masker. The thresholds of masking have been calculated assuming the level at each octave frequency to be a pure tone. The figure shows how the high frequency tones are perfectly resolved for a 20 years old cochlea, and how the 8 kHz tone is at least 7 dB higher

than the masking thresholds due to other tones. For the case of the 70 years old cochlea this is not the case, and as the cochlear filters are broader, the 8 kHz tone is now masked by the vibration produced by the tones of 1 kHz, 2 kHz and 4 kHz.

The reduction of speech intelligibility in the hearing impaired by masking is observed in many behavioural studies. The use of binary masking in low frequency bands have shown to have greater benefits in hearing impaired listeners than in normal hearing listeners [236]. As the results of Fig. 7.19 show, if the 8 kHz high frequency tone is to be heard in a 70 years old cochlea, the input volume at that frequency has to be increased. Research in the past has shown that by high pass filtering amplified speech in a noise background, the intelligibility is increased [237]. This constitutes one of the main basis of hearing aid fitting procedures [238, 33, 36].

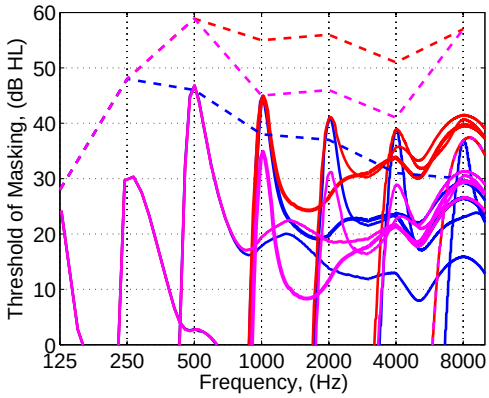


Figure 7.20: Predicted masking patterns in octave bands for the LTASS at normal effort plus the gain produced by the array on the bright zone, for a 20 years old listener (blue graphs) and for a 70 years old listener (red graphs). Also shown is the masking patterns for a 70 years old listener using an alternative equalisation (magenta graphs). The dashed line shows the input level of the masker in dB SPL.

The main application of the arrays introduced in this PhD is that of amplifying the sound of the TV in a certain spatial position. If the TV is playing speech with a LTASS as shown in Fig. 7.19, the sound programme in the bright zone will be amplified a certain amount. Using the cochlear model for predicting masking thresholds, it is possible to observe how the upward masking curves are modified by the amplification provided by the array. Fig. 7.20 shows the predicted threshold of masking for the amplified program provided by the array for a 20 years old listener and for a 70 years old listener. For a 20 years old listener the results are similar to that of Fig. 7.19. For the case of the 70 years old cochlea,

[2]See comments of Section 6.3, where it is explained how the speech is equalised for the objective and subjective intelligibility tests.

even with the amplification provided by the array, the tone of 8 kHz is masked by the tones of 1 kHz, 2 kHz and 4 kHz. This is, however, due to the fact that the amplification at 1 kHz, 2 kHz and 4 kHz is very similar to that at 8 kHz.

It has been shown that by spectral shaping the speech programme at low and mid frequency, it is possible to increase speech intelligibility [239]. This fact suggests that speech intelligibility is increased by the reduction of masking at higher frequencies, as also proposed in other studies [240]. In the same way, the amplification produced by the array can be equalised so that the masking release is reduced at higher frequencies. An alternative equalisation is also presented in Fig. 7.20. In this equalisation, the amplification produced by the array at 1 kHz, 2 kHz and 4 kHz has been reduced, in order for the cochlea to resolve the 8 kHz tone. In this case it can be seen how the tone at 8 kHz is resolved, with a margin of around 8 dB. The masking patterns shown in thicker line correspond to that of the 1 kHz tone, frequency at which the 2-nd formant is usually found in speech. The masking threshold curves at 8 kHz show that more than 10 dB of margin are still available with respect to the 1 kHz masking curve. By further increasing the gain at 1 kHz the speech intelligibility may be further improved, as it has been shown that the second formant contributes mainly to the speech information [241].

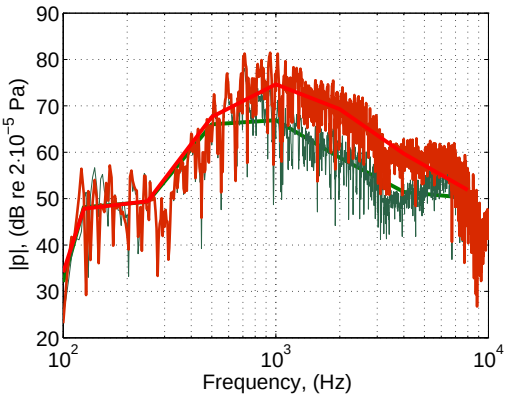


Figure 7.21: LTASS produced by the array in the bright zone when this amplifies full bandwidth (red line) and when the response has been reduced by 10 dB on the 1 kHz, 2 kHz and 4 kHz bands (green line).

In order to see if the amplification suggested in Fig. 7.20 is effective, a simulation using the STI [42] has been run. The STI does not simulate the increase in masking threshold due to hearing impairment, however, it does simulate the increase in upward masking release when the presentation SPL is high. Fig. 7.21 shows the presentation levels in the bright zone for the two amplifying configurations of Fig.

[2]The STI formulation is presented in Chapter 6.

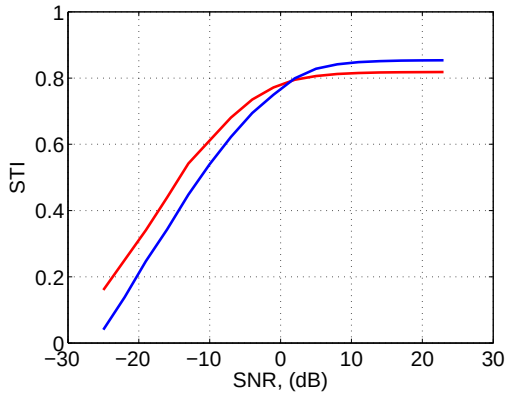


Figure 7.22: STI value for different SNRs for the amplification using the full bandwidth of the array (red line), red spectra in Fig. 7.21, and for the spectral shaped amplification (blue line), green spectra in Fig. 7.21.

7.20 which are similar to the shouted speech; when the array is amplifying the full bandwidth, and with the amplification reduced at 1 kHz, 2 kHz and 4 kHz. For the full bandwidth amplification, it can be seen how a difference of around 10 dB exists between 1 kHz and 4 kHz from the spectral shaped one. The STI results for both amplifications are shown in Fig. 7.22. The curves for the spectral shaped amplification are lower than that from the full bandwidth program for negative SNRs, however they are higher for positive SNRs. This is because for negative SNRs the equalised signal is below the noise value for some frequencies, and so the SNRs is reduced. For higher SNRs the masking reduction is taken in account and the equalised amplification gives an STI about 0.05 higher. This is the case of the amplification provided by the array, in where it is assumed to operate in positive SNRs. Although the STI is just a first design approach, and its use has to be validate with behavioural studies, it has been shown to correlate well with subjective intelligibility tests, and may be used as a first quality check.

7.4 Conclusions

A physiologically based model of the cochlea has been developed to study the potential equalisations that may increase speech intelligibility for the hearing impaired. The model uses a lumped parameter implementation of the organ of Corti, and it is a able to predict responses which reasonably match with behavioural measurements. By calculating the BM for a set of frequencies it is possible to predict upward masking thresholds and psychophysical tuning curves. The model allows the incorporation of hearing loss by modifying the gain of each micromechanical element.

Each micromechanical element is based on a third degree of freedom system, in which one mass models the BM, a second mass models the RL+TM bending displacement and a third mass models the shear displacement of the TM. The action of the CA is replicated in the model by a pressure which acts on the BM and RL and is proportional to the relative displacement between TM and RL. These duplicates the action of the OHC, which produce a force proportional to the bending displacement of the stereocilia. Such feedback enhancement provides a realistic intensity variation of the cochlea time response, with an almost invariant placement of the IRs zero crossings. The model is able of giving an active enhancement of about 35 dB, which is in agreement with that observed in measurements.

The model has been ran iteratively to find a distribution of active gains along the cochlea similar to that assumed to be the median hearing loss of people of a certain age. These cochlear amplifier gains give an idea of how the performance of the cochlea is reduced with ageing. The distribution of feedback gain for different ages is reduced proportionally with the age of the simulated cochlea. The hearing impaired feedback gain distributions are more decreased at positions close to the base, and increase smoothly as moving towards the apex. When the response of a hearing impaired cochlea is calculated for a given input velocity in the stapes, the velocity distribution along the cochlea has a decreased peak value, due to a reduced action of the CA. The consequent reduction of the CA action also causes a broadening of the response, leading to a lower frequency selectivity. Although the variation of the CA gain only allows to model hearing loss due to OHC malfunction, the degeneration of IHC transduction can be modelled by reducing the response of the whole cochlea. This allows to replicate a noise induced hearing loss in where only the OHC are damaged, or to model the effect of presbyacusis, in where the transduction properties of both OHC and IHC are changed due to a decrease in EP.

The cochlear model has also been used to calculate upward masking threshold for different frequencies, as those reproduced by Fletcher [227]. These curves have been obtained by determining the SPL at the ears, at each frequency, needed to produce a BM velocity equal to that which a masker tone of a certain frequency is producing. Using the feedback gains calculated for the hearing impaired it is possible to predict how the masker spreads in a hearing impaired cochlea. The obtained hearing impaired masking patterns show that the masking is increased for off-frequency value, as the cochlear filters are broader. The on-frequency val-

ues are, however, the same for both healthy and hearing impaired cochlea.

An initial study is then performed of how masking is affected by the equalisation used by the array. The effect of different equalisations has been studied in healthy and hearing impaired cochleas, and it has been found that by decreasing the level of the reproduced program in the frequency range between 1 kHz and 4 kHz is beneficial for hearing impaired listeners. This is because the excitation between 1 kHz and 4 kHz largely masks the information present around the 6-8 kHz band. By decreasing the level in the 1-4 kHz it is possible to resolve the high frequency tones, that otherwise are masked in a hearing impaired cochlea. The effectiveness of such equalisation has been checked by using the STI, and it has been shown that this is improved for positive SNRs. This equalisation might be further tuned by optimising the masking along the whole frequency band, as performed by noise reduction algorithms [242]. The effectiveness has yet to be studied with subjective tests, but the STI results are encouraging.

The masking model has also been utilised to study how masking can be applied in personal audio applications to mask such frequencies at which the array is less directional and can have a worse programme separation. A detailed analysis is contained in Appendix G, in where the masking model is employed to analyse a situation in which two listeners are watching a movie in two different languages.

Chapter 8

Conclusions and suggestions for further work

8.1 Conclusions

This thesis has introduced the design of a miniature loudspeaker array to improve speech intelligibility for hard of hearing people when watching TV. As to be effective, the array provides an enhanced version of the TV audio in a narrow spatial zone. The sound pressure produced outside of this zone is kept low as not to annoy normal hearing TV viewers.

The design process which has been carried out through this thesis can be divided into two research streams. A first stream is that of creating a superdirective control of the radiated pressure inside a room. The second stream is that of providing such equalisation to the enhanced audio which maximises speech intelligibility.

8.1.1 Personal audio in rooms

Personal audio refers to the creation of different acoustic zones in a common environment, allowing various user to listen to different audio programmes without interference. Based on the personal audio technology, this thesis has investigated the design of two small size line arrays, aimed to improve the speech intelligibility of hard of hearing TV viewers.

The design of the individual sources for the array has been first considered. The approach taken has been to use phase-shift radiators, which allow to obtain a close to hypercardioid directivity, hence minimising the energy input to the reverberant field. The phase-shift sources initially showed some limited performance. In

order to overcome such performance limitation, a new design of the enclosure has been presented. This uses a combined mass-resistance rear port, which allows to change the radiation pattern of the source with frequency. Using such rear port arrangement, a performance close to that of a theoretical hypercardioid has been obtained between 500 Hz and 8 kHz [58].

Based on the knowledge gained in phase-shift source design, two arrays have been built, which are termed the 1×8 array and the 4×8 array. The 1×8 array is built up with 8 small phase-shift sources, with each phase-shift source having an oval construction. Using such array, the performance has been measured in an anechoic chamber, with the array obtaining an acoustic contrast which grows linearly with increasing logarithmic frequency from about 10 dB at 500 Hz to about 25 dB at 8 kHz, when using an array effort lower than 6 dB [53]. The performance of the array, however, has shown to be quite low when the array is placed in a normal room.

The use of phase-shift radiators for an array may seem, at first glance, limited in terms of the steering capabilities of the array. It has been shown, however, that an array of phase-shift sources is able of obtaining the same performance of a dual-layer array between 0° (broadisde) and 75°, requiring half the amount of control systems.

Although the performance a beamformer grows with the number of sources, the robustness of the radiator with respect to mismatches in the transfer impedances decreases as more sources are used. Furthermore, the digital processing gets more demanding as more independent channels are needed. This has motivated the design of the 4×8 array. The 4×8 array is built of a pile of four 1×8 arrays, stacked one above the other, hence its name. The individual sources are wired so that each column of the array is driven in phase. Such arrangement leads to a radiator which needs of the same number of control filters as the 1×8 array, but that is able of obtaining a much larger directionality in a 3D sense. If measured in a normal room, the 4×8 array has shown to give about 5 dB more of acoustic contrast than the 1×8 array at about 5 kHz [95].

Various superdirective control approaches have been considered in this thesis, these are: the acoustic contrast maximisation (ACM), the least squares inverse filters (LSIF) and the mixed control (MC) approach. Between all these approaches, the LSIF and mixed control (MC) algorithm are preferred rather than the ACM for the application of the array as a hearing aid. The ACM maximises the ratio

of mean squared pressure between two control zones, however, does not allow control of the characteristics of the reproduced signal on the listening zone [53]. The other two methods, LSIF and MC, allow the selection of a pressure target to, so that a certain pressure field may be reproduced on the listening zone. In terms of directional performance, the three control methods obtain a very similar performance in the control geometries considered in this thesis.

The creation of superdirective control filters has been studied with the filters optimised both in frequency and time domains. In order to optimise the filters in the time domain, the formulation for the ACM, LSIF and MC control approaches has been adapted, using existing time domain theory [71]. Whilst that in the frequency domain design method the filters are optimised one frequency at a time, in the time domain design methods all the frequencies are optimised at the same time. Due to this particularity, the filters created using the time domain implementation of the ACM do not show a smooth frequency responses with frequency. For filter lengths below $I = 256$ coefficients, the filters created using the time domain implementation of the LSIF have shown to obtain a very similar directional performance to the filters optimised in the frequency domain. When the filters are optimised in the time domain, this allows to use very short filter lengths, i.e., 16 coefficients, whilst obtaining very similar directional performance to that obtained when filters of larger lengths are optimised in the frequency domain [243] as also been object of study. Furthermore, it has been shown that the time domain optimisation allows to largely reduce the modelling relay required by the superdirective filters, which leads to an enhanced audio quality.

8.1.2 Maximisation of speech intelligibility

It has been shown that by the use of the arrays, an effective personal audio device which provides an enhanced version of the TV audio in a small spatial region can be created. Apart from the acoustic performance which the array obtain, it is of importance to assess how the arrays can enhance speech intelligibility.

The assessment of speech intelligibility has been carried, first with an objective measurement method, and then by performing a set of subjective tests. As objective assessment technique, the speech transmission index (STI) has been used. The STI allows to calculate the performance, in terms of speech intelligibility, of a certain transmission channel. The STI allows to perform such prediction based on the frequency response of the channel, which then allows its off-line prediction.

In this thesis, it has been used to compare the speech intelligibility obtained with a loudspeaker simulating the built-in loudspeaker of the TV, and the speech intelligibility which is obtained when the array is working in conjunction with the loudspeaker. The predictions have shown to obtain an improvement in the STI about 0.2 in the STI, which corresponds to a change in intelligibility from fair to good or from good to excellent.

In order to test the intelligibility improvement that the array introduces in hard of hearing listeners, a subjective test has been carried out. The test has been performed by presenting processed audio cues via headphones to 30 individuals. The processed audio cues have been constructed by convolving the measured responses of the array inside a listening room with BKB sentences processed with a hearing loss simulator. At the ears of the healthy hearing subjects, the cues processed with the hearing loss simulator should have the same cognition as that when a hearing impaired subjects listen to the unprocessed cues.

The outcome of the experiment showed that by the use of the array in conjunction with the TV loudspeaker, for a given SNR, the intelligibility can be improved by 25 % without annoying the listeners with healthy hearing. This result has shown that it is possible to find a trade-off between the speech intelligibility enhancement for a hearing impaired person, and the perceived sound quality for a normal hearing person which is close to the hearing impaired listener.

This thesis has also considered the equalisation of the array signal and its effect on speech intelligibility. A cochlear model has been used to estimate hearing impaired upward masking thresholds. This has allowed to estimate the masking produced with a certain array equalisation. Predictions using the STI have shown that by reducing the spectral content of the array at certain frequency bands, masking can be reduced, and the speech intelligibility can be maximised.
The masking model can also be applied in personal audio applications. In this case the masking model has allowed to determine the equalisation needed for an array so that the audio separation between two control zone is effective.

8.2 Further work

The subjective experiments have shown that it has been possible to increase the speech intelligibility of the hard of hearing listeners with the 4×8 array, without disturbing the perceived quality of nearby normal hearing listeners. The

performance of the array, however, can still be improved, which suggests a number of areas for potential future work.

8.2.1 Personal audio in rooms

The arrays presented in this section use 8 sources. If another array is built in the future, this should use a larger number of independent channels, in order to increase the directional performance of the device.

In order to further increase the performance of an array inside a reverberant environment, it is important to know how the soundfield is changed by such environment. This could be predicted if the exact position of the array inside the environment, and with respect to its main reflectant surfaces, is known. Further research could be performed in the direction of making the device able to characterise the room in which it is working, and hence apply that knowledge to obtain a more accurate sound reproduction which provides a better speech intelligibility.

At the moment, all the research has considered a single beamforming direction, aimed towards a fixed position of the hearing impaired listener. This is an ideal situation, and in practice, the position of the listener in need of amplification will change over time. A real time tracking of the listener which adapts to the reproduction environment will certainly benefit the personal audio performance of the system. Research on this area is being performed at the moment, and may be, it could be applied in the future to enhance the performance of the array as an aid for the hard of hearing.

8.2.2 Maximisation of speech intelligibility

The presented cochlear model could be further tuned to represent more accurately the amplitude gain and frequency selectivity found in behavioural experiments. This will provide a better prediction of masking thresholds.

The hearing impaired cochlear model could be also implemented in the time domain, allowing to represent the non-linear processes found in the cochlea. This could potentially be used to study temporal coding and otoacustic emissions in the hearing impaired.

Optimisation of the equalisation of a signal could then be undertaken so that masking is minimised. The masking model could be used to find the equalisation

curve that minimises masking and hence provides a higher speech intelligibility.

Bibliography

[1] K. S. Helfer and R. L. Freyman, "Aging and speech-on-speech masking," *Ear & Hearing*, vol. 29, pp. 87–98, 2008.

[2] B. C. J. Moore, *Cochlear Hearing Loss.* Wurr Publishers Ltd, 1998.

[3] M. R. Leek, M. F. Dorman, and Q. Summerfield, "Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners," *The Journal of the Acoustical Society of America*, vol. 81, no. 1, pp. 148–154, 1987.

[4] B. A. Schneider and S. J. Hamstra, "Gap detection thresholds as a function of tonal duration for younger and older listeners," *The Journal of the Acoustical Society of America*, vol. 106, no. 1, pp. 371–380, 1999.

[5] B. C. J. Moore, *An introduction to the psychology of hearing.* Academic press, 1997.

[6] B. Schneider, "Psychoacustics and aging: Implications for everyday listening," *Revue d'orthophonie et d'audiologie*, vol. 21, pp. 111–124, 1997.

[7] G. A. Gates and J. H. Mills, "Presbycusis," *The Lancet*, vol. 366, no. 9491, pp. 1111 – 1120, 2005.

[8] J. F. Willott, *Aging & The Auditory System.* Whurr Publishers LTD, 1991.

[9] P. Borough, "35 million brits cant properly hear whats being said on TV says Widex," *Response Source*, 2011.

[10] S. Rosen, "Presbycusis Study of a Relatively Noise-Free Population in the Sudan," *Transactions of the American Otological Society*, vol. 50, pp. 133–152, 1962.

[11] NIH, "Presbycusis," National Institute of Deafness and other Communication Disorders, Tech. Rep. 4333, October 1997.

[12] *ISO 7029-Statistical distribution of hearing thresholds as a function of age*, International Standards Organization Std., 2000.

[13] S. Coren, "Most comfortable listening level as a function of age," *Ergonomics*, vol. 37, no. 7, pp. 1269–1274, 1994, pMID: 8050411.

[14] V. Yule, "Golden oldies," *New Scientist*, p. 32, March 2011, opinion letter.

[15] M. E. Lutman, "Personal communication," 2011.

[16] H. Müsch, "Aging and sound perception: Desirable characteristics of entertainment audio for the elderly," *Audio Engineering Society 125 convention*, 2008.

[17] C. D. Mathers, "A study of sound balances for the heard of hearing," BBC, Research Department, Tech. Rep., 1991.

[18] D. Walker and L. Wilcox, "The audibility of speech on television. analysis of problems," Tech. Rep., 2010.

[19] Sennheiser, "Hearing amplifiers," Sennheiser, 2012. [Online]. Available: *http : //www.sennheiser.co.uk/uk/home_en.nsf/root/private_impaired*

[20] ZVOX, "Single cabinet home teather systems for hearing impaired," ZVOX, 2010. [Online]. Available: *http : //www.zvoxaudio.com/cgi − bin/category/hearingimpaired*

[21] "TV EARS, Wireless loudspeaker as an assisted listening device for a TV." 2012. [Online]. Available: http://www.tvears.com/TV-Ears-Wireless-Speaker-System

[22] "Audio fox," 2010. [Online]. Available: http://www.audiofox.com/wireless-tv-headset

[23] Widex, "TV-DEX, TV without echo," 2011. [Online]. Available: http://www.widex.co.uk/en/products/dex/tvdex/

[24] NIH, "Captions for deaf and hard-of-hearing viewers," 2002. [Online]. Available: *http : //www.nidcd.nih.gov/health/hearing/caption.asp*

[25] M. Armstrong, J. Barrett, and M. Evans, "Enabling and Enriching Broadcast Services by Combining IP and Broadcast Delivery," BBC, Tech. Rep., 2010.

[26] F. Rumsey, "Hearing enhancement," *Journal of Audio Engineering Society*, vol. 57, pp. 353–359, 2009.

[27] H. Müsch, "Enhancement of multichannel audio," United States of America Patent 8 271 276 B1, May, 2012.

[28] M. Vinton and C. Robinson, "Automated speech/other discrimination for loudness monitoring," *Audio Engineering Society Convention 118*, 2005.

[29] *ANSI S3.5-1997 Methods for Calculation of the Speech Intelligibility Index*, American National Standard Institute Std., Rev. Reafirmed by ANSI, June 18, 2012, 1997.

[30] O. H. Christian Uhle and J. Weigel, "Speech enhancement of movie sound," in *Audio Engineering Society Convention 125*, 10 2008.

[31] J. C. G. M. van Rooij and R. Plomp, "Auditive and cognitive factors in speech perception by elderly listeners. iii. additional data and final discussion," *The Journal of the Acoustical Society of America*, vol. 91, no. 2, pp. 1028–1033, 1992.

[32] S. F. Lybarger, "Simplified fitting system for hearing aids," Raio Ear Corp., Canonsburg, PA, Tech. Rep., 1963.

[33] D. Byrne and H. Dillon, ""The national acoustics laboratories" (NAL) new procedure for selecting gain and frequency response of a hearing aid," *Ear and Hearing*, vol. 7, pp. 257–265, 1986.

[34] H. Dillon, "NAL-Nl1: A new procedure for firtting Non-linear hearing aids," *Hearing Journal*, vol. 52, pp. 10–16, 1999.

[35] G. Kedser, H. Dillon, M. Flax, T. Ching, and S. Brewer, "The NAL-NL2 prescription procedure," *Audiology Reserch*, vol. 1(1), pp. 133–14 288–90, 2011.

[36] T. Y. Ching, H. Dillon, R. Katsch, and D. Byrne, "Maximizing effective audibility in hearing aid fitting," *Ear & Hearing*, vol. 22, pp. 212–224, 2001.

[37] E. Y. Choueiri, "Optimal crosstalk cancellation for binaural audio with two loudspeakers," *Princeton University.*

[38] T. Takeuchi and P. A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers," *The Journal of the Acoustical Society of America*, vol. 6, no. 112, pp. 2786–2797, 2002.

[39] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, 1993.

[40] W. F. Druyvesteyn and J. Garas, "Personal sound," *Journal of Audio Engineering Society*, vol. 45, pp. 685–701, 1997.

[41] A. P. W.F. Druvesteyn, R.M.Aarts and A. Ruxton, "Personal sound," in *in Proceedings of the Institute of Acoustics, vol 16, p. 571*, 1994.

[42] *Sound system equipment-Part 16: Objective rating of speech intelligibility by speech transmission index (IEC 60268-16:2011)*, International Electrotechnical Commision Std., 2011.

[43] J. M. Kates, "Superdirective arrays for hearing aids," *The Journal of the Acoustical Society of America*, vol. 94, no. 4, pp. 1930–1933, 1993.

[44] H. Cox, R. M. Zeskind, and T. Kooij, "Practical supergain," *IEEE Trans. Acoust. Speech. Signal Processing*, no. 34, pp. 393–398, 1986.

[45] B. D. V. Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, no. 5, pp. 4–24, 1988.

[46] J. W. Choi and Y. H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *The Journal of the Acoustical Society of America*, vol. 111, no. 111, pp. 1695–1700, 2002.

[47] M. Shin, S. Q. Lee, F. M. Fazi, P. A. Nelson, D. Kim, S. Wang, K. H. Park, and J. Seo, "Maximization of acoustic energy difference between two spaces," *The Journal of the Acoustical Society of America*, vol. 128, no. 1, pp. 121–131, 2010.

[48] S. J. Elliott and M. Jones, "An active headrest for personal audio," *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 2702–2709, 2006.

[49] M. Shin, F. M. Fazi, P. A. Nelson, and J. Seo, "Control of velocity for sound field reproduction," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception*, Sep 2013.

[50] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduña Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Transactions on Audio Speech and Language Processing*, vol. 6, no. 2, 1998.

[51] J.-H. Chang and F. Jacobsen, "Experimental validation of sound field control with a circular double-layer array of loudspeakers," *The Journal of the Acoustical Society of America*, vol. 133, no. 4, pp. 2046–2054, 2013.

[52] S. J. Elliott, J. Cheer, H. Murfet, and K. R. Holland, "Minimally radiating sources for personal audio," *The Journal of Acoustical Society of America*, vol. 128 (4), pp. 1721–1728, 2010.

[53] M. F. Simón Gálvez, S. J. Elliott, and J. Cheer, "A superdirective array of phase shift sources," *The Journal of the Acoustical Society of America*, vol. 132, no. 2, pp. 746–756, 2012.

[54] M. Jones and S. J. Elliott, "Personal audio with multiple dark zones," *The Journal of the Acoustical Society of America*, vol. 124, no. 6, pp. 3497–3506, 2008.

[55] J. Cheer, "Designing loudspeaker directivity for mobile devices," University of Southampton, 2009.

[56] Y. K. Jordan Cheer, Stephen J. Elliott and J.-W. Choi, "The effect of finite sized baffles on mobile device personal audio," in *Audio Engineering Society Convention 130*, 5 2011.

[57] J. Cheer, S. J. Elliott, Y. Kim, and J.-W. Choi, "Practical implementation of personal audio in a mobile device," *Journal of Audio Engineering Society*, vol. 61, no. 5, pp. 290–300, 2013.

[58] J. Cheer, S. J. Elliott, and M. F. Simón Gálvez, "Design and implementation of a car cabin personal audio system," *Journal of Audio Engineering Society*, vol. 61, no. 6, pp. 412–424, 2013.

[59] J.-H. Chang, C.-H. Lee, J.-Y. Park, and Y.-H. Kim, "A realization of sound focused personal audio system using acoustic contrast control," *The Journal of the Acoustical Society of America*, vol. 125, no. 4, pp. 2091–2097, 2009.

[60] Y.-H. K. Ji-Ho Chang, Jin-Young Park, "Scattering effect on the sound focused personal audio system," *The Journal of the Acoustical Society of America*, vol. 125, no. 5, pp. 3060–3066, 2009.

[61] J.-Y. Park, J.-H. Chang, and Y.-H. Kim, "Generation of independent bright zones for a two-channel private audio system," *J. Audio journal = Journal of Audio Engineering Society, Eng. Soc*, vol. 58, no. 5, pp. 382–393, 2010.

[62] S. Q. L. K.-H. P. Hye J. Kim, Kunmo Koo and J. Kim, "High performance piezoelectric microspeakers and thin speaker array system," *ETRI Journal*, vol. 31(6), pp. 680–687, 2009.

[63] S. J. E. Naomichi Yanagidate, Jordan Cheer and T. Toi, "Car cabin personal audio: Acoustic contrast with limited sound differences," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Helsinki, Finland, 2014.

[64] M. M. Boone, W.-H. Cho, and J.-G. Ih, "Design of a highly directional endfire loudspeaker array," *Journal of Audio Engineering Society*, vol. 57, no. 5, pp. 309–325, 2009.

[65] J. Luzarraga Iturrioz, "Inverse filter design for sound field control with loudspeaker line array systems," Master's thesis, University of Southampton, ISVR, 2011.

[66] M. Shin, F. M. Fazi, F. C. Hirono, and P. A. Nelson, "Control of a dual-layer loudspeaker array for the generation of private sound," in *Internoise*, 2012.

[67] J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *The Journal of the Acoustical Society of America*, vol. 131, no. 6, pp. 4518–4525, 2012.

[68] Y. Cai, M. Wu, and J. Yang, "Sound reproduction in personal audio systems using the least-squares approach with acoustic contrast control constraint," *The Journal of the Acoustical Society of America*, vol. 135, no. 2, pp. 734–741, 2014.

[69] M. Shin, F. M. Fazi, P. A. Nelson, and F. C. Hirono, "Controlled sound field with a dual layer loudspeaker array," *Journal of Sound and Vibration*, vol. 333, no. 16, pp. 3794 – 3817, 2014.

[70] P. Coleman, P. J. B. Jackson, M. Olik, M. Mller, M. Olsen, and J. Abildgaard Pedersen, "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker arraya)," *The Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 1929–1940, 2014.

[71] O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," *Journal of Audio Engineering Society*, vol. 47, no. 7/8, pp. 583–595, 1999.

[72] P. Nelson, F. Orduna-Bustamante, and H. Hamada, "Inverse filter design and equalization zones in multichannel sound reproduction," *Speech and Audio Processing, IEEE Transactions on*, vol. 3, no. 3, pp. 185–192, May 1995.

[73] S. G. Norcross, G. A. Soulodre, and M. C. Lavoie, "Subjective investigations of inverse filtering," *Journal of Audio Engineering Society*, vol. 52, no. 10, pp. 1003–1028, 2004.

[74] Y. Cai, M. Wu, and J. Yang, "Design of a time-domain acoustic contrast control for broadband input signals in personal audio systems," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 341–345.

[75] S. J. Elliott and J. Cheer, "Regularization and robustness of personal audio systems," ISVR, University of Southampton, Technical Memorandum 995, 2012.

[76] Y. Cai, M. Wu, L. Liu, and J. Yang, "Time-domain acoustic contrast control design with response differential constraint in personal audio systems," *The Journal of the Acoustical Society of America*, vol. 135, no. 6, pp. EL252–EL257, 2014.

[77] H. Kutruff, *Room Acoustics*. London: Applied Science Publishers LTD, 1979.

[78] J. Mourjopoulos, "On the variation and invertibility of room impulse response functions," *Journal of Sound and Vibration*, vol. 102, no. 2, pp. 217 – 228, 1985.

[79] ——, "Digital equalization of room acoustics," in *Audio Engineering Society Convention 92*, Mar 1992.

[80] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *The Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100–2111, 2005.

[81] B. E. Anderson, B. Moser, and K. L. Gee, "Loudspeaker line array educational demonstration," *The Journal of the Acoustical Society of America*, vol. 131, no. 3, pp. 2394–2400, 2012.

[82] L. L. Beranek, *Acoustics*. New York: American Institute of Physics, 1987.

[83] F. Olivieri, M. Shin, F. M. Fazi, P. A. Nelson, and P. Otto, "Loudspeaker array processing for multi-zone audio reproduction based on analytical and measured electroacoustical transfer functions," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception*, Sep 2013.

[84] J. Y. Yauan Wen and W. seng Gan, "Strategies for an acoustical-hotspot generation," *IEICE Trans. Fundamentals.*, vol. E88A (7), pp. 1739–1746, 2005.

[85] S. Yon, M. Tanter, and M. Fink, "Sound focusing in rooms: The time-reversal approach," *The Journal of the Acoustical Society of America*, no. 113, pp. 1533–1543, 2003.

[86] ——, "Sound focusing in rooms: The spatio-temporal inverse filter," *The Journal of the Acoustical Society of America*, vol. 114, pp. 3044–3052, 2003.

[87] F. A. A. S. A. Canclini, D. Marković and S.Tubaro, "A room-compensated virtual surround system exploiting early reflections in a reverberant room," in *20th European Signal Processing Conference (EUPSICO 2012)*, Bucharest, Romania, 2012.

[88] T. Betlehem and M. A. Poletti, "Two dimensional sound field reproduction using higher order sources to exploit room reflections," *The Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 1820–1833, 2014.

[89] E. N. Gilbert and S. P. Morgan, "Optimum design of directive antenna arrays subject to random variations," *Bell System Technical Journal*, vol. 34, no. 3, pp. 637–663, 1955.

[90] S. Doclo and M. Moonen, "Superdirective beamforming robust against microphone mismatch," *IEEE Transactions on Audio Speech and Language Processing*, vol. 15, no. 2, pp. 617–631, 2007.

[91] S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Transactions on Audio Speech and Language Processing*, vol. 20, no. 7, pp. 2123–2133, 2012.

[92] J.-Y. Park, J.-W. Choi, and Y.-H. Kim, "Acoustic contrast sensitivity to transfer function errors in the design of a personal audio system," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. EL112–EL118, 2013.

[93] E. A. Jorswieck, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Transactions on Signal Processing*, vol. 52, no. 12, December 2004.

[94] M. F. Simón Gálvez, "Louspeaker Arrays for Family TV," Master's thesis, ISVR, University of Southampton, 2011.

[95] M. F. Simón Gálvez, S. J. Elliott, and J. Cheer, "Personal audio loudspeaker array as a complementary tv sound system for the hard of hearing," *IEICE Trans. Fundamentals.*, vol. E97(9), 2014.

[96] ——, "The effect of reverberation on personal audio devices," *The Journal of the Acoustical Society of America*, vol. 135, no. 5, pp. 2654–2663, 2014.

[97] J. Pefretzschner and M. Romera, "Directivity improvement in acoustical arrays," *Applied Acoustics*, no. 9, pp. 215–224, 1976.

[98] H. F. Olson, "Gradient loudspeakers," *Journal of Audio Engineering Society*, no. 21, pp. 86–93, 1973.

[99] JBL, "Forward Steered Arrays in Precision Directivity Speaker^{TM} Systems," JBL, Tech. Rep. 28, 2001.

[100] Meyer Sound, "PSW-6, Self Powered Cardioid Woofer," Meyer Sound, Tech. Rep., 1998.

[101] T. J. Holmes, "The "acoustic resistance box"- a fresh look at an old principle," *Journal of Audio Engineering Society*, vol. 34, no. 12, pp. 981–989, 1985.

[102] W. H. Iding, "Unidirectionally radiating loudspeakers," in *Audio Engineering Society Convention 1ce*, Mar 1971.

[103] J. Backman, "Theory of acoustical resistance enclosures," May 1999.

[104] H. F. Olson, *Acoustical Engineering*. Princeton, NJ: Van Nostrand, 1957.

[105] P. A. Nelson and S. J. Elliott, *Active Control of Sound*. London: Academic Press, 1992.

[106] A. Marker, "Using a hearing loss simulator to test a TV loudspeaker array," Master's thesis, ISVR, University of Southampton, 2013.

[107] P. Newell and K. R. Holland, *Louspeakers for music recording and reproduction*. Focal Press, 2007.

[108] I. Wolf and L. Malter, "Directional radiation of sound," *The Journal of the Acoustical Society of America*, vol. 2, pp. 201–241, 1930.

[109] F. M. Fazi, F. Hirono, and P. A. Nelson, "Dual-layer loudspeaker array for multiple listening zones," *The Journal of the Acoustical Society of America*, vol. 131, no. 4, pp. 3216–3216, 2012.

[110] B. B. Bauer, "US Patent, Conversion of wave motion into electrical energy," 1941.

[111] C. J. Powles, A. McAlpine, and R. H. Self, *Introductory Mathematics for Sound and Vibration Studies*. ISVR, 2010.

[112] S. J. Elliott and K. R. Holland, *Microphone Directivity*, ISVR, 2011.

[113] M. F. Simón Gálvez, S. J. Elliott, and J. Cheer, "Loudspeaker arrays for family TV," in *Proceedings of the 19th International Congress on Sound and Vibration*. International Institute of Acoustics and Vibration, July 2012.

[114] D. T. Blackstock, *Physical Acoustics*, D. T. Blackstock, Ed. John Wiley & Sons, 2000.

[115] *IEC International Standard 61094-3, Measurement microphone Part 3: Primary method for calibration of laboratory microphones by the reciprocity technique*, IEC Std., 1995.

[116] J. Vanderkooy, "The acoustic center: A new concept for loudspeakers at low frequencies," in *Audio Engineering Society Convention 121*, 10 2006.

[117] D. Henwood and J. Vanderkooy, "Polar plots for low frequencies: The acoustic centre," in *Audio Engineering Society Convention 120*, 5 2006.

[118] J. Vanderkooy, "Applications of the acoustic centre," in *Audio Engineering Society Convention 122*, 5 2007.

[119] M. Zampolli, N. Malm, and A. Tesei, "Improved perfectly matched layers for acoustic radiation and scattering problems," *Proceedings of the Comsol Comference, Hannover*, 2008.

[120] K. Shin and J. K. Hammond, *Fundamentals of Signal Processing for sound and vibration engineers*, Wiley, Ed., 2008.

[121] Visaton, "Visaton k20.40, Miniature moving coil loudspeaker," 04 2013. [Online]. Available: http://www.visaton.de/en/industrie/miniatur/k20_40_8.html

[122] F. Fahy and J. Walker, *Fundamentals of Noise and Vibration*, F. Fahy and J. Walker, Eds. Spon Press, 1998.

[123] J.-L. Sánchez Bote and E. Álvarez Fernández, *Transductores Electoacústicos*. Escuela Universitaria de Ingeniería Técnica de Telecomunicaciónes, UPM, 2000.

[124] N. Kino, "A comparison of two acoustical methods for estimating parameters of glass fibre and melamime foam materials," *Applied Acoustics*, vol. 73, pp. 590–603, 2011.

[125] M. Poletti, F. M. Fazi, and P. A. Nelson, "Sound-field reproduction systems using fixed-directivity loudspeakers," *The Journal of the Acoustical Society of America*, vol. 127, no. 6, pp. 3590–3601, 2010.

[126] A. Gupta and T. Abhayapala, "Three-dimensional sound field reproduction using multiple circular loudspeaker arrays," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 5, pp. 1149–1159, July 2011.

[127] Sheerwood, "S-9 3D Sound Bar System," 2013. [Online]. Available: http://www.sherwoodusa.com/

[128] Yamaha Electronics Corporation, "YSP-1100 surround soundbar," 2012. [Online]. Available: http://www.yamaha.com/yec/ysp1/idx_ysp1100.htm

[129] BOSE, "Bose Cinemate 1 SR home theater speaker system." [Online]. Available: http://worldwide.bose.com/productsupport/en˙us/web/cinemate˙1sr/page.html

[130] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, Eds. John Wiley & Sons, 2000.

[131] F. Fahy and J. Walker, *Advanced Applications in Acoustics, Noise & Vibrations*, F. Fahy and J. Walker, Eds. London: Spon Press, 2004.

[132] S. Spors and R. Rabenstein, "Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis," *Audio Engineering Society 120 convention*, May 2006.

[133] F. M. Fazi, "Sound field reproduction," Ph.D. dissertation, University of Southampton, Faculty of engineering, science and mathematics, ISVR, 2010.

[134] J. M. Carcione and F. Cavallini, "On the acoustic-electromagnetic analogy," *Wave Motion*, no. 21, pp. 149–162, 1995.

[135] P. A. Nelson, H. Hamada, and S. J. Elliott, "Inverse filters for multi-channel sound reproduction," *IEICE Trans. Fundamentals.*, vol. E75-A(11), pp. 1468–1473, 1992.

[136] G. V. Borgiotti, "The power radiated by a vibrating body in an acoustic fluid and its determination from boundary measurements," *The Journal of the Acoustical Society of America*, vol. 88, no. 4, pp. 1884–1893, 1990.

[137] F. M. Fazi and P. A. Nelson, "The ill-conditioning problem in sound field reconstruction," Oct 2007.

[138] A. Tychonoff, "On the stability of inverse problems," *Doklady Akademii Nauk SSSR*, vol. 39, pp. 195–198, 1943.

[139] E. Saff and A. Kuijlaars, "Distributing many points on a sphere," *The Mathematical Intelligencer*, vol. 19, no. 1, pp. 5–11, Dec. 1997.

[140] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *The Journal of the Acoustical Society of America*, vol. 66, no. 1, pp. 165–169, 1979.

[141] S. J. Elliott, *Signal Processing for Active Noise Control*, S. J. Elliott, Ed. Academic Press, 2001.

[142] M. R. Bai, G.-Y. Shih, and C.-C. Lee, "Comparative study of audio spatializers for dual-loudspeaker mobile phones," *The Journal of the Acoustical Society of America*, vol. 121, no. 1, pp. 298–309, 2007.

[143] E. Mabande, A. Schad, and W. Kellermann, "A time-domain implementation of data-independent robust broadband beamformers with lowfilter order," in *Hands-free Speech Communication and Microphone Arrays (HSCMA), 2011 Joint Workshop on*, May 2011, pp. 81–85.

[144] J. Fung, "Lit. Review. Effects of steering delay quantization in beamforming," Master's thesis, The University of Texas at Austin.

[145] J.-L. Sanchez-Bote and J.-J. Gomez-Alfageme, "Fractional delay in the frequency domain for beamforming applications," *Journal of Audio Engineering Society*, vol. 58, no. 3, pp. 160–172, 2010.

[146] A. Marker, S. J. Elliott, and M. F. Simón Gálvez, "Using listening tests to demonstrate the subjective performance of a superdirective TV loudspeaker array," in *In Forum Acusticum 2014*, 2014.

[147] M. Stone, B. C. Moore, K. Meisenbacher, and R. Derleth, "Tolerable hearing aid delays v. estimation of limits for open canal fittings," *Ear & Hearing*, vol. 29, pp. 601–617, 2008.

[148] O. Kirkeby, P. A. Nelson, F. Orduña Bustamante, and H. Hamada, "Local sound field reproduction using digital signal processing," *The Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1584–1593, 1996.

[149] "ISVR listening room," 2000, date last viewed 16/12/2013. [Online]. Available: http://www.isvr.co.uk/faciliti/listroom.htm

[150] *BS EN ISO 3382:1-2009, Acoustics-Measurement of room acoustic parameters*, ISO Std., 2009.

[151] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, Feb 2000.

[152] *BS EN ISO 354:2003 ,Acoustics – Measurement of sound absorption in a reverberation room*, ISO Std., 2003, acoustics-Measurement of sound absorption in a reverberation room.

[153] C. F. Eyring, "Reverberation time in "dead" rooms," *The Journal of the Acoustical Society of America*, vol. 1, no. 2A, pp. 217–241, 1930.

[154] R. H. Lyon, "Statistical analysis of power injection and response in structures and rooms," *The Journal of the Acoustical Society of America*, vol. 45, no. 3, pp. 545–565, 1969.

[155] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, "Measurement of correlation coefficients in reverberant sound fields," *The Journal of the Acoustical Society of America*, vol. 27, no. 6, pp. 1072–1077, 1955.

[156] W. M. Hartmann and Y. J. Cho, "Generating partially correlated noisea comparison of methods," *The Journal of the Acoustical Society of America*, vol. 130, no. 1, pp. 292–301, 2011.

[157] N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," *The Journal of the Acoustical Society of America*, vol. 19, no. 1, pp. 90–119, 1947.

[158] K. D. Kryter, "Methods for the calculation and use of the articulation index," *The Journal of the Acoustical Society of America*, vol. 34, no. 11, pp. 1689–1697, 1962.

[159] T. Houtgast and H. J. Steeneken, "Evaluation of speech transmission channels by using artificial signals," *Acustica*, vol. 25, pp. 355–367, 1971.

[160] ——, "The modulation transfer function in room acoustics as a predictor of speech intelligibility," *Acustica*, vol. 28, pp. 66–73, 1973.

[161] ——, *Past, present and future of the Speech Transmission Index*, S. J. van Wijngaarden, Ed. TNO Human Factors, Soesterberg, The Netherlands, 2002.

[162] H. J. M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *The Journal of the Acoustical Society of America*, vol. 67, no. 1, pp. 318–326, 1980.

[163] T. Houtgast and H. J. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *The Journal of the Acoustical Society of America*, vol. 77, no. 3, pp. 1069–1077, 1985.

[164] Y. Peled and B. Rafaely, "Objective performance analysis of spherical microphone arrays for speech enhancement in rooms," *The Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1473–1481, 2012.

[165] Y. Nejime and B. C. J. Moore, "Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise," *The Journal of the Acoustical Society of America*, vol. 102, no. 1, pp. 603–615, 1997.

[166] D. Howard and J. Angus, *Acoustics and Psychoacoustics*. Taylor & Francis, 2009.

[167] B. C. J. Moore and B. R. Glasberg, "Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech," *The Journal of the Acoustical Society of America*, vol. 94, no. 4, pp. 2050–2062, 1993.

[168] T. Baer and B. C. J. Moore, "Effects of spectral smearing on the intelligibility of sentences in noise," *The Journal of the Acoustical Society of America*, vol. 94, no. 3, pp. 1229–1241, 1993.

[169] N. H. van Schijndel, T. Houtgast, and J. M. Festen, "Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal-hearing and hearing-impaired listeners," *The Journal of the Acoustical Society of America*, vol. 110, no. 1, pp. 529–542, 2001.

[170] H. Hu, J. Sang, M. Lutman, and S. Bleeck, "Simulation of hearing loss using compressive gammachirp auditory filters," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 5428–5431.

[171] "Brian C.J. Moore, Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, United Kingdom."

[172] R. D. Patterson, I. Nimmo-Smith, D. L. Weber, and R. Milroy, "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," *The Journal of the Acoustical Society of America*, vol. 72, no. 6, pp. 1788–1803, 1982.

[173] J. Bench, s. Kowal, and J. Bamford, "The BKB (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children," *British Journal of Audiology*, vol. 13, no. 3, pp. 108–112, 1979.

[174] G. E. Box, J. S. Hunter, and W. Hunter, *Statistics for experimenters: design, innovation and discovery.* New Jersey: John Wiley & Sons, 2009.

[175] Mathworks, "Matlab R2011b," 2011.

[176] National Institute for Occupational Safety and Health, "NIOSH: Criteria for a Recommended Standard: Occupational Noise Exposure," Cincinaty, Tech. Rep., 1998.

[177] D. J. Finney, *Probit Analysis.* Cambridge, UK: Cambridge University Press,, 1971.

[178] *ITU-R BS1534-1Method for the subjective assessment of intermediate quality level of coding Systems*, International Telecommunications Union Std., 2003.

[179] G. A. Soulodre and M. C. Lavoie, "Subjective evaluation of large and small impairments in audio codecs," in *Audio Engineering Society Conference: 17th International Conference: High-Quality Audio Coding*, Aug 1999.

[180] U. Wüstenhagen, B. Feiten, T. Buholtz, R. Schwalve and J. Kroll, "Method for assessment of audio systems with intermediate audio quality," in *Proceedings of the 21st Tonmeistertagung*, 2003.

[181] A. F. Ryan, "Protection of auditory receptors and neurons: Evidence for interactive damage," *Proceedings of the National Academy of Sciences*, vol. 97, no. 13, pp. 6939–6940, 2000.

[182] M. A. Ruggero and N. C. Rich, "Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane," *Journal of Neuroscience*, vol. 11, pp. 1057–1097, 1991.

[183] B. C. J. Moore and B. R. Glasberg, "Comparisons of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairments," *The Journal of the Acoustical Society of America*, vol. 80, no. 1, pp. 93–107, 1986.

[184] G. von Békésy, *Experiments in Hearing*. McGraw-Hill, 1960.

[185] T. Irino and R. D. Patterson, "A compressive gammachirp auditory filter for both physiological and psychophysical data," *The Journal of the Acoustical Society of America*, vol. 109, no. 5, pp. 2008–2022, 2001.

[186] R. Meddis and A. L. Poveda, "Peripherial auditory system: From pinna to auditory nerve," in *Computer Models of the Auditory syte, Springer Handbook of Auditory Research*. Springer, New York, 2010.

[187] L. Robles and M. A. Ruggero, "Mechanics of the mammalian cochlea," *Physiological Reviews*, vol. 81, no. 3, pp. 1305–1352, 2001.

[188] E. Boer, "Mechanics of the cochlea: Modeling efforts," in *The Cochlea*, ser. Springer Handbook of Auditory Research, P. Dallos, A. N. Popper, and R. R. Fay, Eds. Springer New York, 1996, vol. 8, pp. 258–317.

[189] S. J. Elliott and C. A. Shera, "The cochlea as a smart structure," *Smart Materials and Structures*, vol. 21, no. 6, p. 064001, 2012.

[190] S. Ramamoorthy, N. V. Deo, and K. Grosh, "A mechano-electro-acoustical model for the cochlea: Response to acoustic stimuli," *The Journal of the Acoustical Society of America*, vol. 121, no. 5, pp. 2758–2773, 2007.

[191] M. F. Simón Gálvez and S. J. Elliott, "ISVR TM 1000, A lumped parameter model of the organ of Corti," Institute of Sound and Vibration research, Tech. Rep., 2012.

[192] C. A. Shera, "Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics," *The Journal of the Acoustical Society of America*, vol. 110, no. 1, pp. 332–348, 2001.

[193] A. Recio and W. S. Rhode, "Basilar membrane responses to broadband stimuli," *The Journal of the Acoustical Society of America*, vol. 108, no. 5, pp. 2281–2298, 2000.

[194] E. de Boer and A. L. Nuttall, "The mechanical waveform of the basilar membrane. III. intensity effects," *The Journal of the Acoustical Society of America*, vol. 107, no. 3, pp. 1497–1507, 2000.

[195] H. Fletcher, "Auditory patterns." *Review of Modern Physics*, vol. 12, pp. 47–65, 1940.

[196] B. R. Glasberg and B. C. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, no. 12, pp. 103 – 138, 1990.

[197] T. Bentsen, J. M. Harte, and T. Dau, "Human cochlear tuning estimates from stimulus-frequency otoacoustic emissions," *The Journal of the Acoustical Society of America*, vol. 129, no. 6, pp. 3797–3807, 2011.

[198] C. A. Shera, J. John J. Guinan, and A. J. Oxenham, "Revised estimates of human cochlear tuning from otoacustic and behavioral measurements," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 3318–3223, 2002.

[199] J. Meaud and K. Grosh, "The effect of tectorial membrane and basilar membrane longitudinal coupling in cochlear mechanics," *The Journal of the Acoustical Society of America*, vol. 127, no. 3, pp. 1411–1421, 2010.

[200] S. Cantrel, "A model of the three-dimensional active cochlea," Master's thesis, Institute of Sound and Vibration Research, University of Southampton, Southampton, UK, 2012.

[201] M. F. Simón Gálvez and S. J. Elliott, "Hearing impaired cochlear response simulation," *Proceedings of Meetings on Acoustics*, vol. 19, no. 1, pp. –, 2013.

[202] S. Neely and D. Kim, "A model for actrive elements in cochlear biomechanics," *The Journal of the Acoustical Society of America*, vol. 79, no. 5, pp. 1472–1480, 1986.

[203] F. Baumgarte, "A physiological ear model for the emulation of masking," *ORL*, vol. 61, pp. 294–304, 1999.

[204] S. Verhulst, T. Dau, and C. A. Shera, "Nonlinear time-domain cochlear model for transient stimulation and human otoacoustic emission," *The Journal of the Acoustical Society of America*, vol. 132, no. 6, pp. 3842–3848, 2012.

[205] Y.-W. Liu and S. T. Neely, "Suppression tuning of distortion-product otoacoustic emissions: Results from cochlear mechanics simulation," *The Journal of the Acoustical Society of America*, vol. 133, no. 2, pp. 951–961, 2013.

[206] M. Heinz and K. S. Henry, "Modeling disrupted tonotopicity of temporal coding following sensorineural hearing loss," *Proceedings of Meetings on Acoustics*, vol. 19, no. 1, pp. –, 2013.

[207] S. D. Ewert, S. Kortlang, and V. Hohmann, "A model-based hearing aid: Psychoacoustics, models and algorithms," *Proceedings of Meetings on Acoustics*, vol. 19, no. 1, pp. –, 2013.

[208] M. S. A. Zilany and I. C. Bruce, "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," *The Journal of the Acoustical Society of America*, vol. 120, no. 3, pp. 1446–1466, 2006.

[209] A. Saremi and S. Stenfelt, "Effect of metabolic presbyacusis on cochlear responses: A simulation approach using a physiologically-based model," *The Journal of the Acoustical Society of America*, vol. 134, no. 4, pp. 2833–2851, 2013.

[210] J. Saunders, B. Canlon, and A. Flock, "Changes in stereocilia micromechanics following overstimulation in metabolically blocked hair cells," *Hearing Research*, vol. 24, pp. 217–225, 1986.

[211] B. A. Bohne, G. Harding, and S. Lee, "Death pathways in noise-damaged outer hair cells," *Hearing Research*, vol. 223, pp. 61–70, 2007.

[212] H.-W. Lim, S. H. Choi, H. H. Kang, J. H. Ahn, and J. W. Chung, "Apoptotic pattern of cochlear outer hair cells and frequency-specific hearing threshold shift in noise-exposed balb/c mice," *Clinical and Experimental Otorhinolaryngology*, vol. 1, pp. 80–85, 2008.

[213] S. Stenfelt, "Towards understanding the specifics of cochlear hearing loss: A modelling approach," *International Journal of Audiology*, vol. 47, pp. S10–S15, 2008.

[214] N. B. Sleepcky, "Structure of the mammalian cochlea," in *The cochlea*, A. P.Dallos and R.R.Fay, Eds. Springer-v, 1996.

[215] R. Shcmiedt, "Effects of aging on potassium homeostasis and the endocochlear potential in the gerbil cochlea," *Hearing Research*, vol. 1-2, pp. 125–132, 1996.

[216] J. R. Dubno, F.-S. Lee, L. J. Matthews, J. B. Ahlstrom, A. R. Horwitz, and J. H. Mills, "Longitudinal changes in speech recognition in older persons," *The Journal of the Acoustical Society of America*, vol. 123, no. 1, pp. 462–475, 2008.

[217] N. M. S. J. R. D. B. A. S. H Lang, V Jyothi and R. A. Shcmiedt, "Chronic reduction of endocochlear ppotential reduces auditory nerve activity:further confirmation of an animal model of metabolic presbyacusis," *Journal for the Association for Research in Otolaryngology*, vol. 11, pp. 419–434, 2010.

[218] C. J. Sumner, E. A. Lopez-Poveda, L. P. OMard, and R. Meddis, "A revised model of the inner-hair cell and auditory-nerve complex," *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 2178–2188, 2002.

[219] S. Kurt, M.Sausbier, L. Ruttiger, N. Brandt, C.K.Moeller, U. J.Kindler, U.Zimmerman, H. van Straaten, W.Neuhuber, J.Engel, P. M.Knipper, and H.Schulze, "Critical role for cochlear hair cell bk cannels for coding the tempral strucutre and dynamic range of auditory information for central auditory processing," *FASEB Journal*, vol. 26, pp. 3834–3843, 2012.

[220] G. A. Gates, D. Mills, B. ho Nam, R. DAgostino, and E. W. Rubel, "Effects of age on the distortion product otoacoustic emission growth functions," *Hearing Research*, vol. 163, no. 12, pp. 53 – 60, 2002.

[221] S. R.A., H.Lang, H. Okamura, and B. Schulte, "Effects of furosemide applied chronically to the round window: a model of metabolic presbyacusis," *J Neurosci*, vol. 21, pp. 9643–9650, 2002.

[222] A. Saremi, S. Stenfelt, and E. Mäki-Torkko, "Changes in temporal and spectral functions of the auditory periphery due to aging and noise-induced cochlear pathologies: A comparative clinical study," in *PhD thesis, Saremi, Amin G.. Effects of Specific Cochlear Pathologies on the Auditory Functions: Modelling, Simulations and Clinical Implications. 2014.* Linköpings Universitet, 2014.

[223] M. L. Jepsen and T. Dau, "Characterizing auditory processing and perception in individual listeners with sensorineural hearing loss," *The Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 262–281, 2011.

[224] S. J. Elliott, E. M. Ku, and B. Lineton, "A state space model for cochlear mechanics," *The Journal of the Acoustical Society of America*, vol. 122, no. 5, pp. 2759–2771, 2007.

[225] K. H. Fu-Quan Chen, Hong-Wei Zheng and S.-H. Sha, "Traumatic noise activates rho-family gtpases through transient cellular energy depletion," *The Journal of Neuroscience*, vol. 32(36), p. 1242112430.

[226] M. C. Liberman, L. Dodds, and D. Learson, "Structure-function correlation in noise-damaged ears: A light and electron-microscopic study," in *Basic and Applied Aspects of Noise-Induced Hearing Loss*, ser. NATO ASI Series, R. Salvi, D. Henderson, R. Hamernik, and V. Colletti, Eds. Springer US, 1986, vol. 111, pp. 163–177.

[227] H. Fletcher, *The ASA Edition of Speech and Hearing in Communication*. The Acoustical Society of America, 1995.

[228] E. Zwicker and H. Fastl, Eds., *Psychoacoustics-Facts and models*. New York: Springer, 1999.

[229] P. Noll, "MPEG digital audio coding," *Signal Processing Magazine, IEEE*, vol. 14, no. 5, pp. 59–81, Sep 1997.

[230] E. Terhardt, "Calculating virtual pitch," *Hearing Research*, vol. 1, pp. 155–182, 1979.

[231] A. J. Klein, J. H. Mills, and W. Y. Adkins, "Upward spread of masking, hearing loss, and speech recognition in young and elderly listeners," *The Journal of the Acoustical Society of America*, vol. 87, no. 3, pp. 1266–1271, 1990.

[232] O. Murnane and C. W. Turner, "Growth of masking in sinsorineural hearing loss," *Audiology*, vol. 30, pp. 275–285, 1991.

[233] G. Evans and D. Johnson, "Stress and open office noise," *Journal of Applied Psychology*, vol. 85(5), pp. 779–783, 2000.

[234] R. L. B. Karl S. Pearsons and S. Fidell, "Speech levels in various noise environments," Office of Health and Ecological Effects, Office of Research and Development, U.S. Environmental Protection Agency, Tech. Rep., 1977.

[235] J. M. Pickett, "Effects of vocal force on the intelligibility of speech sounds," *The Journal of the Acoustical Society of America*, vol. 28, no. 5, pp. 902–905, 1956.

[236] D. Wang, U. Kjems, M. S. Pedersen, J. B. Boldt, and T. Lunner, "Speech intelligibility in background noise with ideal binary time-frequency masking," *The Journal of the Acoustical Society of America*, vol. 125, no. 4, pp. 2336–2347, 2009.

[237] R. J. Niederjohn and J. H. Grotelueschen, "The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression," *IEEE Trans. Acoust. Speech. Signal Processing*, vol. 24(4), pp. 277–282, 1976.

[238] H. Dillon, *Hearing Aids*, H. Dillon, Ed. Thieme, 2001.

[239] T. Baer, B. C. J. Moore, and S. Gatehouse, "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times," *Journal of Rehabilitation Research and Development*, vol. 1, pp. 49–72, 1993.

[240] R. Pedlow, "Spectral shaping and the intelligibility of speech in noise," *The Journal of the Acoustical Society of America*, vol. 83, no. S1, pp. S69–S69, 1988.

[241] I. B. Thomas, "The influence of first and second formants on the intelligibility of clipped speech," *Journal of Audio Engineering Society*, vol. 16, no. 2, pp. 182–185, 1968.

[242] S. Gustafsson and P. Jax, "A new approach to noise reduction based on auditory masking effects," Institute of Communication Systems and Data Processing, RWTH, Aachen, Germany.

[243] M. F. Simón Gálvez, S. J. Elliott, and J. Cheer, "Superdirective time domain beamforming for personal audio applications," in *EAA Forum Acusticum 2014 Conference, Krakow, Poland*, 2014.

[244] Visaton, "Visaton k20.40 moving-coil driver." [Online]. Available: http://www.visaton.com/en/industrie/klein/k20_40_8.html

[245] M. Bouchard, S. G. Norcross, and G. A. Soulodre, "Inverse filtering design using a minimal-phase target function from regularization," in *Audio Engineering Society Convention 121*, Oct 2006.

[246] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[247] D. Fernandez Comesaña, S. Steltenpool, G. Carrillo Pousa, H.-E. de Bree, and K. Holland, "Scan and paint: theory and practice of a sound field visualization method," *ISRN Mechanical Engineering*, vol. 2013, no. 241958, pp. 1–11, 2013.

[248] J. B. Allen, "Cochlear micromechanics—a physical model of transduction," *The Journal of the Acoustical Society of America*, vol. 68, no. 6, pp. 1660–1670, 1980.

[249] S. T. Neely and D. Kim, "An active cochlear model showing sharp tuning and high sensitivity," *Hearing Research*, vol. 9, no. 2, pp. 123 – 130, 1983.

[250] Y.-W. Liu and S. T. Neely, "Suppression tuning of distortion-product otoacoustic emissions: Results from cochlear mechanics simulation," *The Journal of the Acoustical Society of America*, vol. 133, no. 2, pp. 951–961, 2013.

[251] E. de Boer and A. L. Nuttall, "The mechanical waveform of the basilar membrane. I. frequency modulations ("glides") in impulse responses and cross-correlation functions," *The Journal of the Acoustical Society of America*, vol. 101, no. 6, pp. 3583–3592, 1997.

[252] S. Ellliott, R. Pierzycki, and B.Lineton, "ISVR TM 967 Lumped-parameter Models for Cochlear Micromechanics," Institute of Sound and Vibration Research, University of Southampton, Tech. Rep., 2006.

[253] S. J. Elliott, "Active control of vibration in aircraft and inside the ear." Active 09 Conference (Ontario), 2009.

[254] C. Geisler and C. Sang, "A cochlear model using feed-forward outer-hair-cell forces," *Hearing Research*, vol. 86, no. 12, pp. 132 – 146, 1995.

[255] P. Dallos, "Overview: Cochlear neurobiology," in *The Cochlea*, ser. Springer Handbook of Auditory Research, P. Dallos, A. N. Popper, and R. R. Fay, Eds. Springer New York, 1996, vol. 8, pp. 258–317.

[256] J. O. Pickles, *An Introduction to the Physiology of Hearing*. Academic Press, 1988.

[257] S. J. Elliott, B. Lineton, and G. Ni, "ISVR Tehnical Memorandum 990, Fluid Coupling between the Elements in a Discrete Model of Cochlear Mechanics," Institute of Sound and Vibration Research, University of Southampton, Tech. Rep., 2010.

[258] S. J. Elliott and G. Ni, "Near field fluid coupling between internal motion of the organ of Corti and the basilar membrane," in *in Mechanics of Hearing*, 2014.

[259] Y. Li, "Corrections regarding to JASA paper," 2012.

[260] S. J. Elliott, B. Lineton, and G. Ni, "Fluid coupling in a discrete model of cochlear mechanics," *The Journal of the Acoustical Society of America*, vol. 130, no. 3, pp. 1441–1451, 2011.

[261] S. T. Neely, "Mathematical modeling of cochlear mechanics," *The Journal of the Acoustical Society of America*, vol. 78, no. 1, pp. 345–352, 1985.

[262] S. J. E. Guangjian Ni and J. Baumgart, "Modelling motions within the organ of Corti," in *Mechanics of Hearing*, 2014.

[263] C. R. Steele and L. A. Taber, "Comparison of WKB calculations and experimental results for three-dimensional cochlear models," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 1007–1018, 1979.

[264] S. Ramamoorthy and A. L. Nuttall, "Half-octave shift in mammalian hearing is an epiphenomenon of the cochlear amplifier," *PLOS One*, vol. 7, p. e45640, 2012.

[265] J. Zakis and M. Witte, "Modelling of the cochlea using Java 3D," *IEEE Engineering in Medicine and Biology Society*, pp. 1–5, 2001.

[266] D. P. Egolf, W. A. Kennedy, and V. D. Larson, "Occluded-ear simulator with variable acoustic properties," *The Journal of the Acoustical Society of America*, vol. 91, no. 5, pp. 2873–2823, 1991.

[267] M. Kringlebotn, "Network model for the human middle ear," *Scandinavian Audiology.*, vol. 17, pp. 75–85, 1988.

[268] E. M. Ku, "Modelling the human cochlea," Ph.D. dissertation, ISVR, University of Southampton, 2008.

[269] J. Young, "Modelling the cochlear origins of distortion product otoacustic emissions," Ph.D. dissertation, ISVR, University of Southampton, 2010.

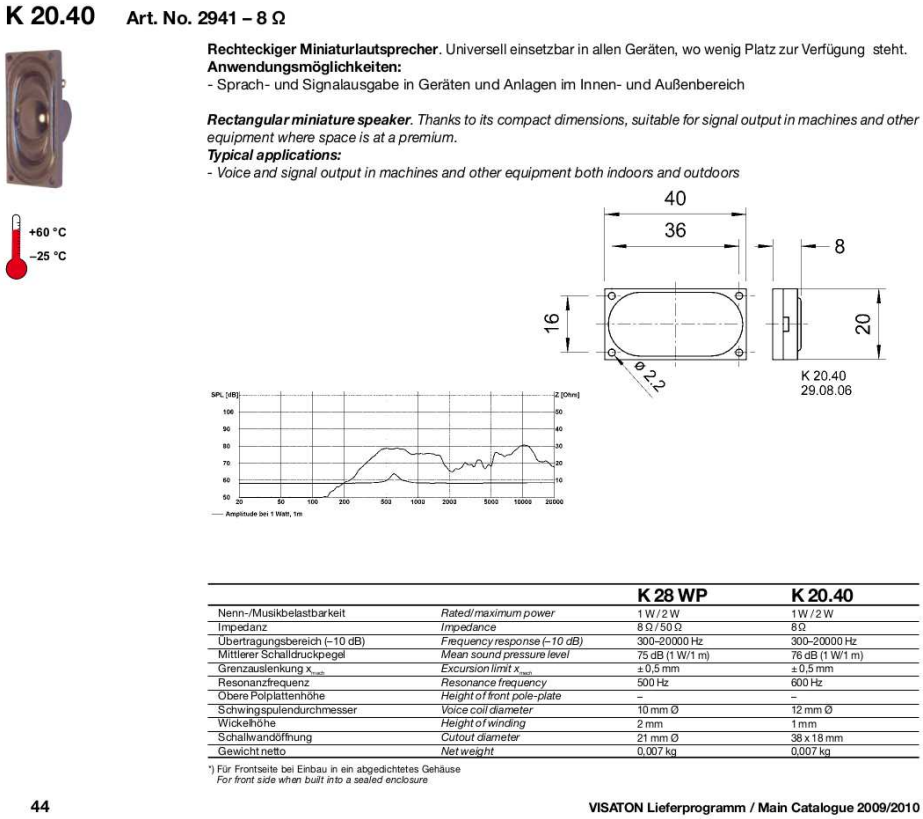
[270] S. Puria, "Measurements of human middle ear forward and reverse acoustics: Implications for otoacustic emissions," *Journal of Acoustical Society of America*, vol. 113, p. 2773, 2003.

Appendices

Appendix A

Moving-coil loudspeaker

A.1 Visaton k20.40 moving-coil driver



K 20.40 **Art. No. 2941 – 8 Ω**

Rechteckiger Miniaturlautsprecher. Universell einsetzbar in allen Geräten, wo wenig Platz zur Verfügung steht.
Anwendungsmöglichkeiten:
- Sprach- und Signalausgabe in Geräten und Anlagen im Innen- und Außenbereich

Rectangular miniature speaker. *Thanks to its compact dimensions, suitable for signal output in machines and other equipment where space is at a premium.*
Typical applications:
- *Voice and signal output in machines and other equipment both indoors and outdoors*

+60 °C
–25 °C

		K 28 WP	K 20.40
Nenn-/Musikbelastbarkeit	*Rated/maximum power*	1 W / 2 W	1 W / 2 W
Impedanz	*Impedance*	8 Ω / 50 Ω	8 Ω
Übertragungsbereich (–10 dB)	*Frequency response (–10 dB)*	300–20000 Hz	300–20000 Hz
Mittlerer Schalldruckpegel	*Mean sound pressure level*	75 dB (1 W/1 m)	76 dB (1 W/1 m)
Grenzauslenkung x_{max}	*Excursion limit* x_{max}	± 0,5 mm	± 0,5 mm
Resonanzfrequenz	*Resonance frequency*	500 Hz	600 Hz
Obere Polplattenhöhe	*Height of front pole-plate*	–	–
Schwingspulendurchmesser	*Voice coil diameter*	10 mm Ø	12 mm Ø
Wickelhöhe	*Height of winding*	2 mm	1 mm
Schallwandöffnung	*Cutout diameter*	21 mm Ø	38 x 18 mm
Gewicht netto	*Net weight*	0,007 kg	0,007 kg

*) Für Frontseite bei Einbau in ein abgedichtetes Gehäuse
For front side when built into a sealed enclosure

Figure A.1: Datasheet provided by Visaton about the k20.40 moving-coil source. Reproduced from [244].

Appendix B

Superdirectivity and audio performance

B.1 Time artefacts created by excessive control effort

The increase in array effort needed for obtaining superdirective control at mid-low frequencies often carries a decrease in audio quality, due to the extra control effort, which results in longer time series for the filters. So, it is quite important, to stablish a balance between the amount of superdirectivity used to increase the directional characteristics and the quality of the reproduced program, which should be of excellent attributes in order to contribute to the improvement of speech intelligibility. In this section, an analysis is presented which studies the effect of the amount of superdirectivity used in the quality of the reproduced audio.

Considering the 4×8 array defined in Section 3.4 in a control zone as that of Fig. 3.4, the LSIF formulation defined in Section 3.1.2 have been use to create two sets of filters with $I = 1024$ coefficients, with the array effort limited to be lower than 6 dB and 20 dB. The transfer functions from the 4-th source of the array to the bright control points have been used as a target pressure, without any modelling delay[1]. The results for the directional and reproduced response obtained with such sets of filters are shown in Fig. B.1.

The results of Fig. B.1 show how a larger acoustic contrast with a lower limita-

[1]See block diagram of Fig. 3.8. A modelling delay is usually employed when calculating the inverse filter to assure that the result is causal. As in this case the optimisation is performed one frequency at a time, no modelling delay is needed.

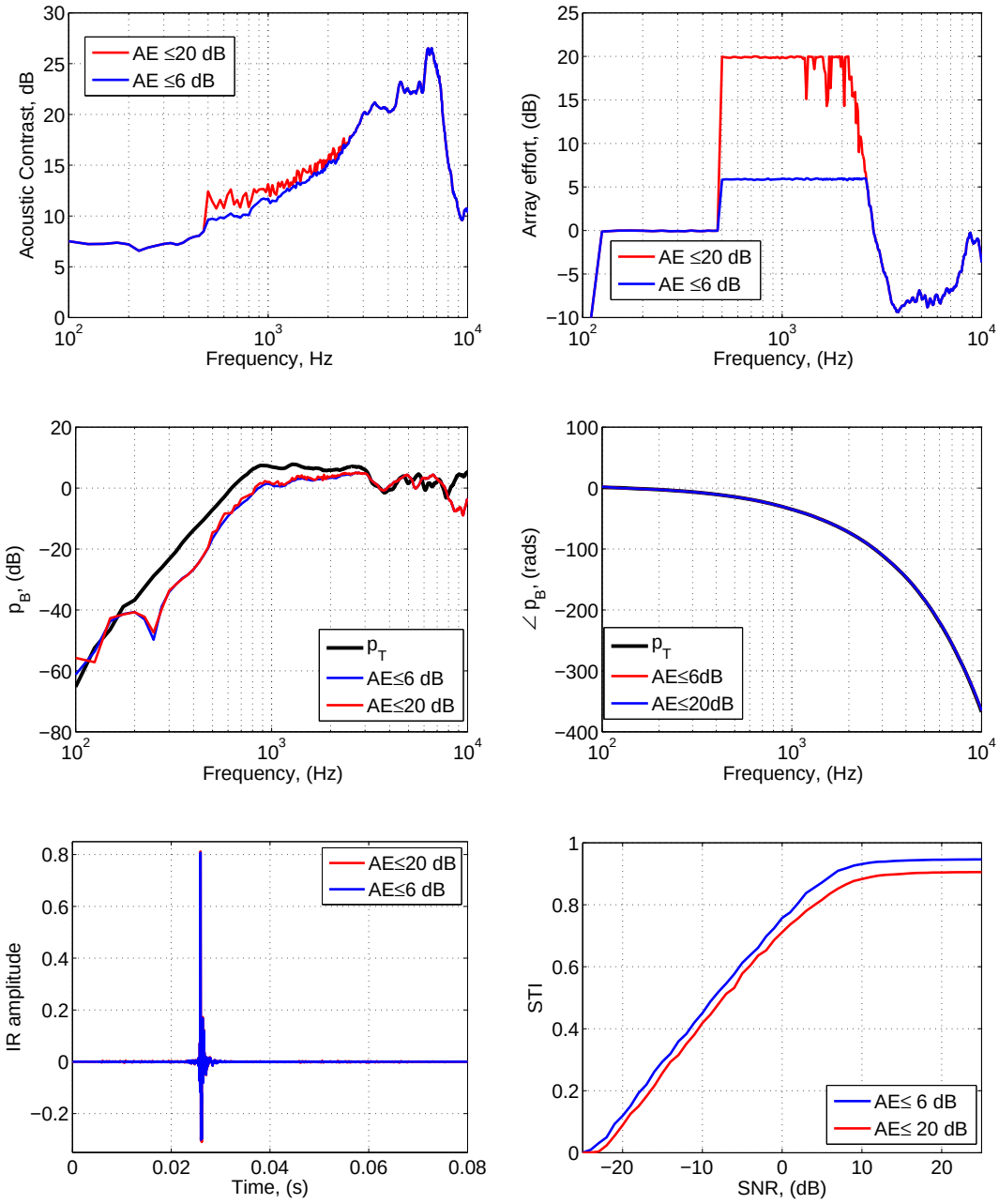


Figure B.1: Acoustic contrast performance and array effort for the 4×8 array in a control geometry as that of Fig. 3.4, with filters created with the array effort limited to 6 dB and to 20 dB. The middle graphs show the bode plots of the reproduced signal when compared to the target pressure. In the lower graphs it is shown the IR of the reproduced signal and the STI obtained with both sets of filters.

tion in array effort. The array effort is controlled by the regularisation parameter, β, which establishes the ratio between smaller and the greater singular values of the matrix $\mathbf{Z}$. By applying regularisation, the ratio between the singular values is reduced. This can be observed in the upper left hand side plot of Fig. B.2, where it can be seen how the condition number is much larger when the array effort is limited to 20 dB. The magnitude of the regularisation parameter needed

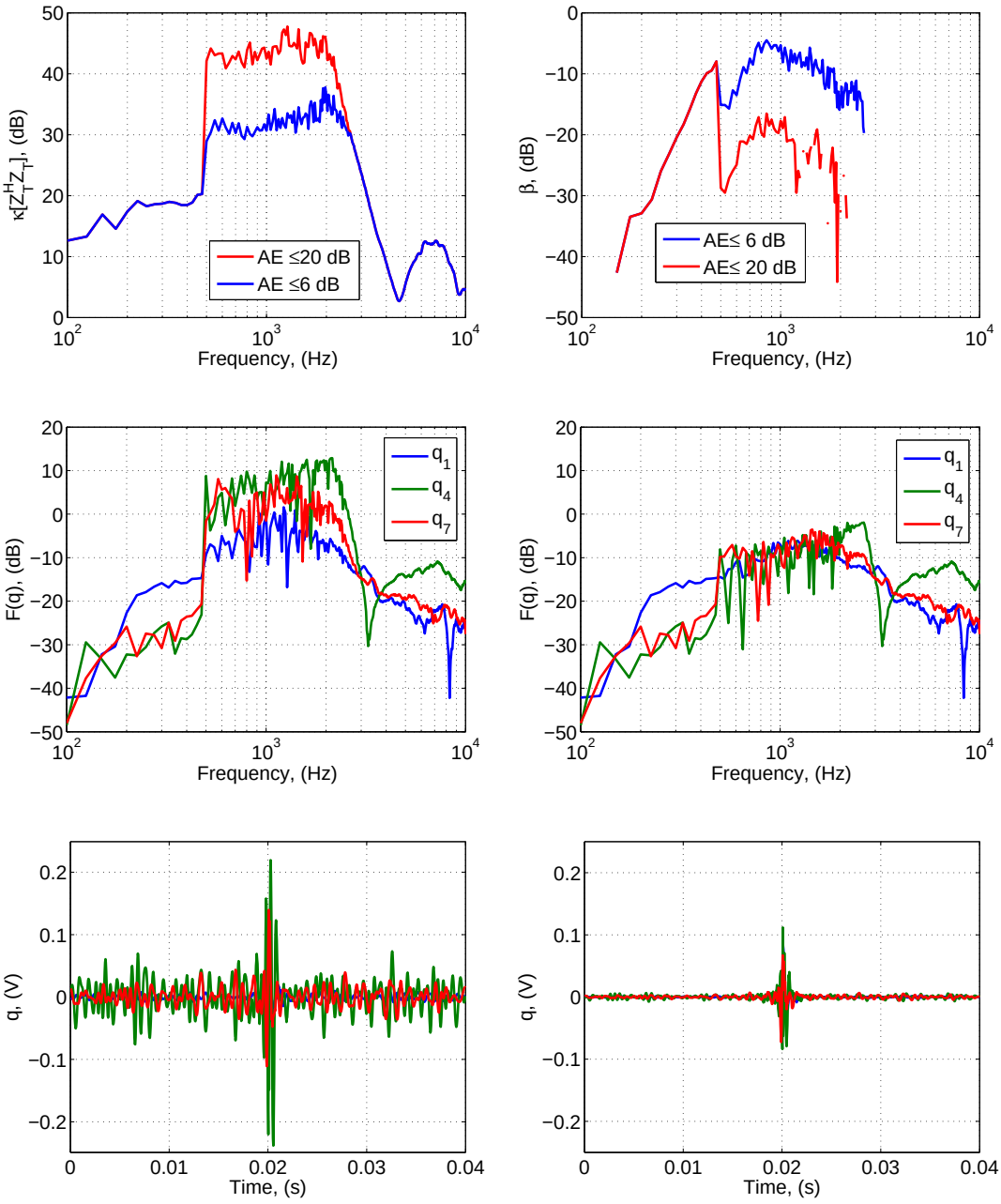


Figure B.2: Condition number with respect to the inversion of the matrix $\mathbf{Z}^H\mathbf{Z}$ and the value of the regularisation coefficient, β, over frequency to obtain the results of Fig. B.1. The middle and lower plots show the magnitude of the frequency responses and the IRs of the filters correspondent to the 1st, 4th and 7th source of the array, with the results for the AE limited to 20 dB in the left hand side plot and the results with the array limited to 6 dB in the right hand side plot.

to limit the array effort in both cases is shown in the upper right hand side plot of Fig. B.2. It can be observed how above 500 Hz the regularisation parameter is large in order to limit the AE to 6 dB. The larger condition number is reflected in filters with a greater magnitude response, as can be observed in the middle plots of Fig. B.2. The time series of the filters with a less severe AE limitation

lead to impulse responses with a higher amount of pre- and post-ringing, as can be compared in the lower plots of Fig.B.2.

The effect of a lower AE limitation at the expense of obtaining a larger superdirectivity has direct effects on the quality of the perceived audio. This can be observed in the middle plots Fig. B.1, where it is shown the bode diagram of the FFT of array response to an unitary impulse. The responses are compared with that of the target pressure, and it can be observed how above 3 kHz, when the condition number is low, the reproduced response is quite close to that of the target. Below 3 kHz, the response of both set of filters is a few dB lower than that of the target pressure. The response of the set of filters with the array effort limited to 20 dB presents a higher amount of ripple. The higher control effort means that the array puts more energy in obtaining control at certain frequencies, which is appreciable in the fact that peaks are present in the frequency response magnitude and acoustic contrast. Such energy excesses lead to an IR which presents pre- and post-ringing, which decrease the quality of the reproduced signal.

The results obtained are in agreement with those obtained by Norcross *et al.* [73], in where the authors assessed the subjective performance of frequency domain and time domain LSIF. The authors showed that with the correct amount of frequency dependent regularisation a better subjective grade was obtained, which corresponded to a lower amount of energy outside of the peak of the IR. The effect of pre- and post-ringing has been previously studied by Bouchard *et al.* [245] in the equalisation of the response of a single source inside a room. The authors used a minimum phase target function which allowed to control the amount of pre- and post-ringing. The effect of selecting a minimum phase target for the array has been tested, however, the obtained results have shown to be very similar to that when using the complete form of the transfer functions of the source[2]. The analysis has centred in the use of least squares filters, however, the findings which are here presented are common to the rest of superdirective approaches, and are found in every control approach which involves the inversion of a matrix, as the acoustic contrast maximisation algorithm or the Chang and Jacobsen mixed control approach.

[2]It is possible to decompose every signal $H(j\omega)$ into a minimum phase part which accounts for the magnitude response, $H_{MP}(j\omega)$, and an all-pass part which accounts for the phase response $H_{AP}(j\omega)$, so that $H(j\omega) = H_{MP}(j\omega)H_{AP}(j\omega)$. This is of importance as the exact inverse only exists for the minimum phase part of the response.

B.2 Conclusions

This appendix introduces an analysis of the effects of superdirectivity in the obtained audio quality. Apart from the fact that the individual sources of the array can not be driven with filters using a high array effort, as otherwise they will be overdriven and burned, a large amount of superdirectivity leads to filters which provide a low audio quality. An analysis has been presented with two sets of filters in where the array effort has been limited to be below 6 dB or below 20 dB. Whilst that the first set of filters obtained a good audio quality, assessed by informal listening, the former obtained a response with pre echoes, which reduced the quality of the response.

If the application requires a good audio quality, as in the case of using the array as a hearing aid, it is important then that the reproduced response is perceptually accurate. It has been shown, however, that by the use of 6 dB of array effort, an effective superdirectivity with a good audio quality is obtained.

Appendix C

Time domain filters creation via frequency domain responses

C.1 Time series obtention

The formulation contained in Section 3.1.4 calculates the optimal source strength vector at a single frequency. The optimisation is calculated between $-f_S/2$ and $f_S/2$ with a spam of Δ_f, where f_S is the sampling frequency. Δ_f depends on the number N of frequency bins used in the fast Fourier transform (FFT) for the calculation of the transfer responses $\mathbf{Z}_B$ and $\mathbf{Z}_D$.

After the optimisation is calculated for the whole number of frequency bins, the optimal frequency responses of each filter are windowed to reduce the energy at frequencies out of the bandwidth where an effective control is to be obtained. The time series of the filters are obtained via inverse fast Fourier transformation (IFFT), to give impulse responses of length

$$I = \left(\frac{f_S}{\Delta_F + 1} \right). \tag{C.1}$$

After the obtention of the impulse responses, these are shifted $(I-1)/2$ samples to assure causality.

For the proposed application, in where the array has to work complementary with the audio system of the TV, the processing of the filters to create the superdirective beamforming has to be quite fast, as otherwise it will lead to a complementary audio signal which is out of phase with the TV audio. To this end, the length of the filters has to be relatively short. Considering a sampling frequency, f_S=25.6

kHz, as the one used for processing the array responses[1], the modelling delay required for filters of different lengths is shown in Table C.1.

Table C.1: Modelling delays needed for the processing of filters of different length with $f_S = 25.6$ kHz.

Filter length, I	32	256	512	1024	2048
Delay, s	0.00065	0.005	0.010	0.020	0.040

As observed in Table C.1, a short filter length is ideal in terms of minimising the processing delay. In order to obtain a good superdirective performance, a certain number of frequency bins are, however, needed for the characterisation of the transfer responses of the drivers. Fig. C.1 shows an analysis of the acoustic contrast, array effort and accuracy of the reproduced signal which is obtained with filters of different lengths. As can be observed, in order to obtain a good directional and frequency response, the filters length has to be at least of 512 coefficients. If the length of the filters is shorter than such figure, dips in the directional and frequency responses are obtained, which correspond to an impulse response with a large amount of pre- and post- ripple. This constraints the signal which comes out of the array to be delayed a certain quantity, which may not be ideal for working complementary with the TV audio system.

[1]See appendix J in where a list of the equipment used in the measurements is contained.

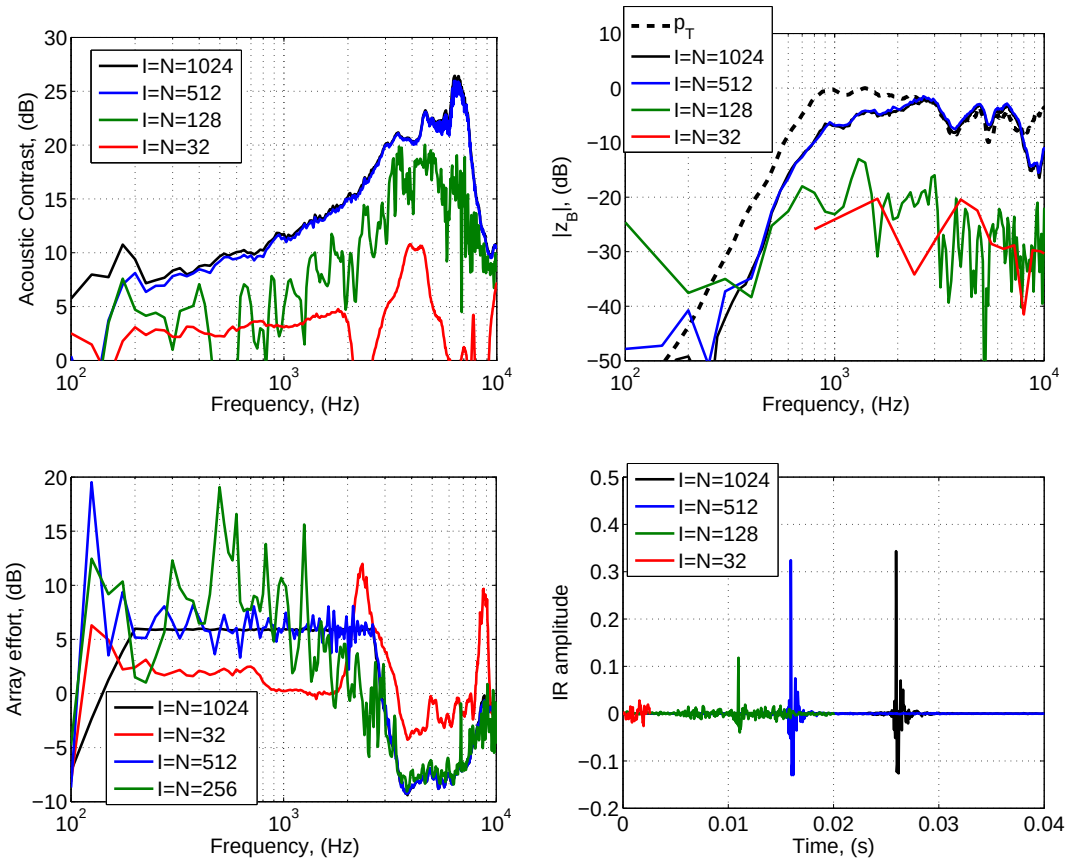


Figure C.1: Acoustic contrast and array effort for the 4×8 array efforts for filters with different lengths (left hand side plots). Also shown are the frequency and impulse responses at the centre of the listening zone (right hand side plots). The array is optimised with LSIF filters in a control zone as that of Fig. 3.4, with the array effort limited not to exceed 6 dB.

C.2 Truncating the impulse responses

In order to further reduce the array filters, another possibility consists in calculating the response of the filters with a larger number of frequency bins, e.g., N=2048 or $N = 4096$, and then truncating so that the delay introduced by the filter is smaller. In order to perform such truncation, a Hanning window with a desired length, I_T, is multiplied at the centre of the original time responses. After the multiplication, the coefficients of the response which are equal to zero are neglected, so that a new impulse response with a length equal to I_T is obtained. An example of the truncation is shown in Fig. C.2, where it is possible to observe the original impulse response, the Hanning window used in the truncation and the truncated impulse response.

Fig. C.3 shows the directional and frequency response calculated as in Fig. C.1, but with filters which their impulse responses have been truncated. The filters

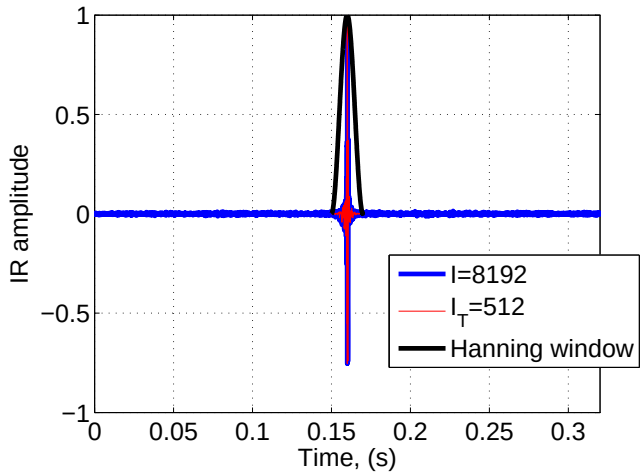


Figure C.2: Impulse response truncation process. The original impulse response with a length $I = 8192$ is truncated by a Hanning window to obtain a shorter impulse response of length $I_T = 512$.

have been calculated for the 4×8 array using the LSIF method in a control geometry as that of Fig. 3.4. In this case the original impulse responses have been calculated with I=1024 samples with the array effort limited to be below 6 dB.

When the impulse responses are truncated, the directional response is larger than when the same number of frequency bins are used to calculate the optimal filters, with the directional responses being consistent if the number of coefficients is greater than 64. As the number of coefficients is decreased, the energy contained in the tails of the impulse responses is erased. This causes the filter not to control in the same way the frequencies that need a large effort, what causes the array effort to decrease as the number of coefficients of the filters is also decreased. The effect on the directional response is, however, minimal. The frequency responses show a very flat distribution with frequency. The corresponding impulse responses show a similar peak height, which slightly decreases with lowering the number of coefficients. As the original impulse responses have been calculated with the array effort limited below 6 dB, the responses are free of pre- and post-ripple. Furthermore, the ripple is further reduced as it is truncated by the window.

Whilst it is not possible to create short filters by calculating the response of the filters with a low number of frequency bins, short filters which obtain a good frequency and directional response can be obtained by obtaining an optimal set of filters with enough frequency bins and then truncating the impulse responses.

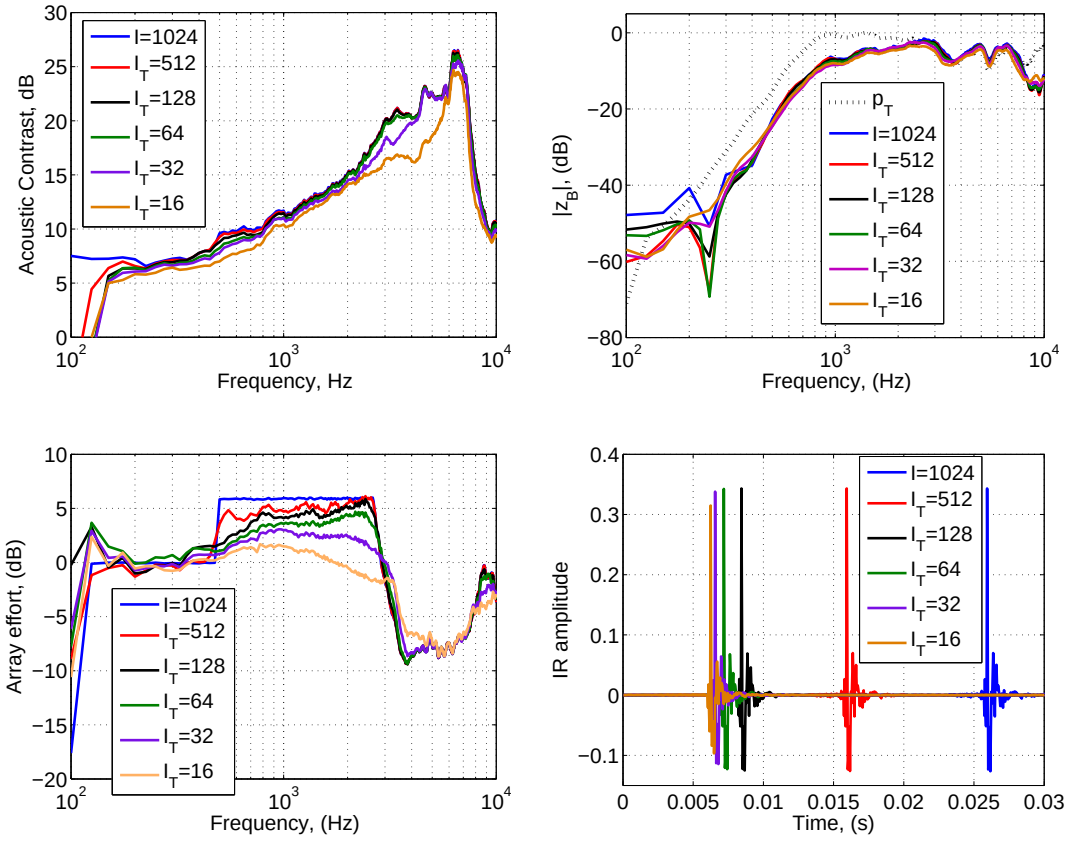


Figure C.3: Acoustic contrast and array effort for the 4×8 array efforts for filters with different values of $I = N$ (left hand side plots). Also shown are the frequency and impulse responses at the centre of the listening zone (right hand side plots). The array is optimised with LSIF filters in a control zone as that of Fig. 3.4, with the array effort limited no to exceed 6 dB.

C.3 Conclusions

This appendix analyses the way that the time series of the superdirective filters are created via inverse fast Fourier transformation from the optimal set of transfer functions. It has been observed that the most effective way of creating filters which are short in length and that at the same time provide good performance is by truncating the vectors of optimal transfer functions.

Filters of 128 coefficients at $f_S = 25.6$ kHz created in this way are able of obtaining an effective contrast.

Appendix D

Simplifying the Array construction process

Superdirective methods require the knowledge of the transfer functions from each source of the array to each control microphone. To this end, the transfer functions of the array must be measured in an anechoic chamber or in the environment where the array is placed. This appendix presents two methods of simplifying the measurement process of the array source transfer functions. The first method is based on the use of analytical transfer functions based in point source. The second method consists in measuring the transfer function of a single source and applying a delay and amplitude correction based on the placement of the different sources of the array. The performance of these two approaches is presented here by means of off-line simulations and real-time measurements. All the responses are calculated based on filters created with the acoustic contrast maximisation algorithm (ACM).

D.1 Modelling the source transfer function with a point source Green function

The transfer function from each source of the array to each control point is modelled as a point hypercardioid source[1], with a mathematical expression depending on the angle θ given by

$$Z(\omega, \Psi, \theta) = j\omega\rho_0 \frac{e^{-jkr}}{4\pi r}(1 - \Psi + \Psi\cos\theta), \tag{D.1}$$

with Ψ being a factor which controls the directivity, which is equal to 0.75 for the case of a hypercardioid source.

[1]The hypercardioid pattern is close to that offered by a phase-shift source.

Each radius r from each control point to each control microphone is calculated according to the separation of the sources of the array and the distance at where the *virtual* control points are placed. The transfer responses based on point source Green functions are used to form the elements of the matrices of impedances $\mathbf{Z}_B$ and $\mathbf{Z}_D$. The formulation described in Section 3.1.4 is afterwards used to obtain the optimal set of source strengths based on point source Green functions, $\mathbf{q}_{GF}$.

The optimal source strengths can be used to perform an off-line simulation with measured transfer impedances. The pressure at the elements of the bright zone is obtained as

$$\mathbf{p}_B = \mathbf{Z}_{B,M}\mathbf{q}_{GF}, \tag{D.2}$$

where the subscript $_M$ stands for measured. The pressure at each element of the dark zone is similarly obtained by multiplying the optimal set of source strengths by $\mathbf{Z}_{D,M}$. The acoustic contrast is afterwards calculated by performing

$$C = \frac{N_D}{N_B}\frac{\mathbf{q}_{GF}^H\mathbf{Z}_{B,M}^H\mathbf{Z}_{B,M}\mathbf{q}_{GF}}{\mathbf{q}_{GF}^H\mathbf{Z}_{D,M}^H\mathbf{Z}_{D,M}\mathbf{q}_{GF}}. \tag{D.3}$$

Fig. D.1 shows an off-line simulation of performance of an 8 phase-shift source line array, as this presented in Simón Gálvez *et al.* [113]. The filters created using Green functions need a strong regularisation at low frequency to overcome the differences in response between the Green function transfer function and that of the source. However, the array shows a reasonable figure of acoustic contrast if a strong regularisation is applied. The simulations using an array effort lower than 0 dB predict an acoustic contrast bigger than 10 dB for almost all the frequency band, except for some little nulls. The performance using Green function transfer impedances is not much lower than that obtained using measured transfer impedances for calculating the optimal set of source strengths, "(Offline, I=1000)", with the AE also limited to be below 0 dB.

The real-time performance of filters created in this fashion has been measured inside an anechoic chamber. The results are presented in Fig. D.2, where it can be observed how the measured results are quite close to the off-line simulations. The measured performance of the filters based on Green functions is very similar to that of filters created with free-field measured transfer functions. The measured results are also compared with those of a delay and sum. The response of the superdirective filters is larger than that of the delay and sum for the whole bandwidth.

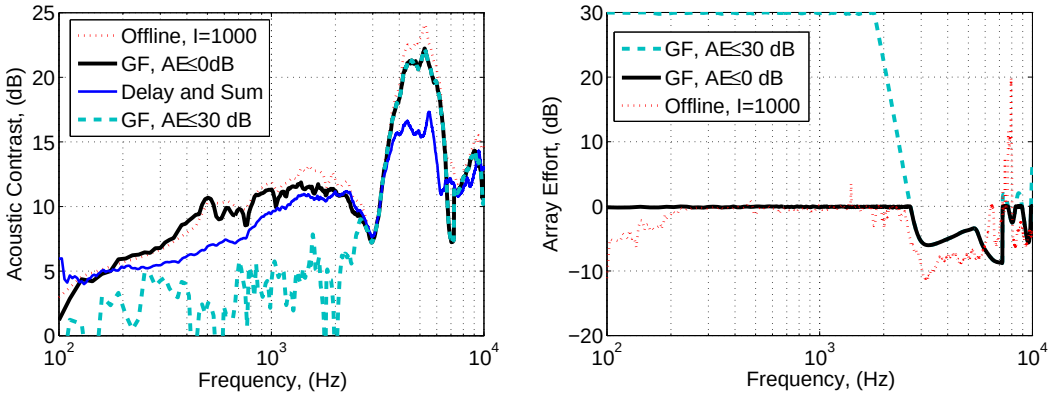


Figure D.1: Results of an off-line simulation of acoustic contrast and array effort for filters based in free-field transfer functions, "Off-line,I=1000", filters based in point hypercardioid Green functions, "GF, AE≤0 dB" and "GF, AE≤0 dB" and for a delay and sum.

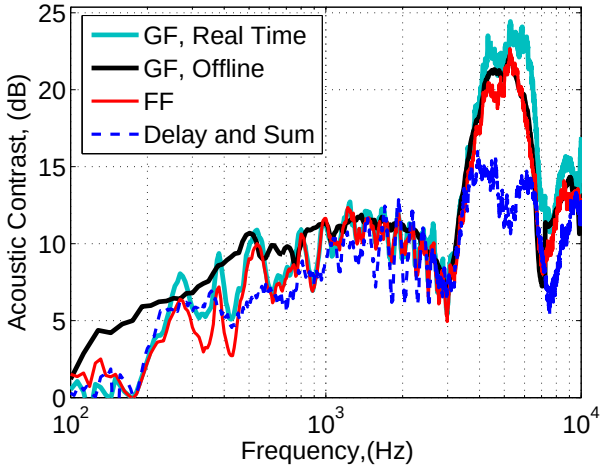


Figure D.2: Results of real-time measurements with the filters calculated using point Green functions. "GF, Real Time" refers to the measured results, "GF, Offline" is an off-line simulation included for comparison and "FF" corresponds to the results obtained based on free-field measured responses. All the filters are calculated with the array effort limited to be below 0 dB.

D.2 Calculating the source strengths assuming that every source behaves equally

Another approach to simplify the filters design process is to measure only the transfer impedance of a single source, and then assume that the rest of the sources have an identical response. A correction in amplitude and phase is then applied

depending on the position of the sources with respect to the array position and the separation between the sources.

Having measured the directivity of a source at different angles, this is used to build the columns of the matrices of impedances to each control zone. In this case $\mathbf{Z}_B$ is built as

$$\mathbf{Z}_{B_R} = \begin{bmatrix} Z_{B_{11}} & Z_{B_{11}} & \cdots & Z_{B_{11}} \\ Z_{B_{21}} & Z_{B_{21}} & \cdots & Z_{B_{21}} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{B_{N1}} & Z_{B_{N1}} & \cdots & Z_{B_{N1}} \end{bmatrix}, \tag{D.4}$$

where $_R$ stands for replicated. In this case the *original* source is the source number 1, i.e., m_0=1. The amplitude and phase of each element of the matrix is now adjusted according to the relative difference that the source number m_0 has with a certain control point, and that the rest of the sources have with the same control point, .i.e.,

$$\Xi = \frac{r_{(m_0)}}{r_{(m)}} e^{-jk\Delta_r}, \tag{D.5}$$

where $\Delta_r = r_{(m_0)} - r_{(m)}$. By calculating the relative differences between each source and each control point, a matrix to adjust the amplitude and phase of each source is built, $\mathbf{\Xi}$.

The matrix of transfer impedances is then built by

$$\mathbf{Z} = \mathbf{Z}_{B_R} \cdot \mathbf{\Xi}, \tag{D.6}$$

which can be afterwards use to create superdirective filters.

The effectiveness of this approach has been tested using off-line simulations. Fig. D.3 shows the performance using filters in which the optimal set of source strengths has been obtained by measuring the free-field impedance of just one source, "$\mathbf{Zq}_R$", in comparison with the acoustic contrast obtained when the optimal set of source strengths is calculated with the measured transfer impedance of each source, "$\mathbf{Zq}$". Also shown for comparison is the predicted contrast assuming that each source of the array would have the same transfer impedance, ""$\mathbf{Z}_R\mathbf{q}_R$". The predicted results show a performance just a few dB lower than that obtained by measuring transfer function of each of the array source. It is also possible to observe how if the transfer impedance of each source would be identical, the acoustic contrast performance would be much larger.

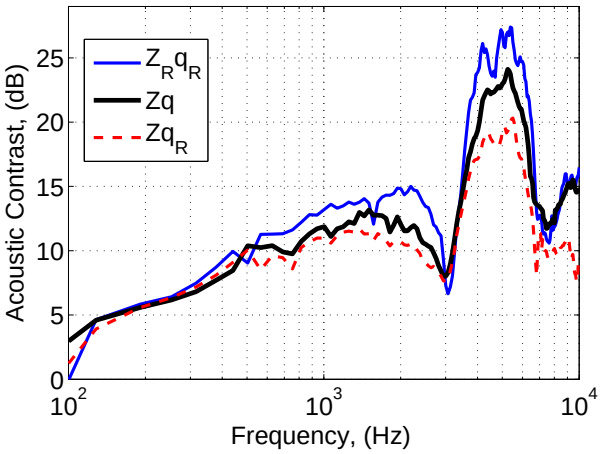


Figure D.3: Prediction of acoustic contrast for an array effort limited to be lower than 0 dB. $\mathbf{Z}_R\mathbf{q}_R$, stands for the case when the optimal strengths are calculated for $\mathbf{Z}_R$ transfer impedances and these impedances are used in the calculation of acoustic contrast (all the sources of the array are identical in response). $\mathbf{Zq}$ stands for the case where the optimal strengths are calculated for the measured transfer impedances (the filters match the measured transfer responses). $\mathbf{Zq}_R$ stands for the case when the optimal strengths are calculated using $\mathbf{Z}_R$ transfer impedances but the estimation of acoustic contrast is calculated using the real, measured, transfer impedances (it is assumed that the directivity of each source is identical when calculating the optimal strengths).

Fig. D.4 shows the measured acoustic contrast using the formulation presented in this section, with the array effort limited to 0 dB. The results based on filters created using the free-field impedance of a single source, "Single source", show a very similar performance to those calculated using the transfer impedances of the whole set of sources, "FF". The obtained results are also very close to that predicted off-line, "Single source, Offline".

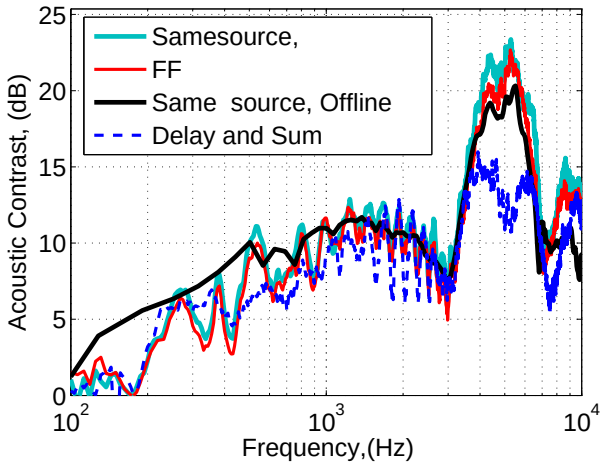


Figure D.4: Results of real-time measurements with filters calculated assuming that all the sources are delayed copies of a single source, "Same source". "Same source, Offline" refers to an off-line simulation with measured transfer impedances and"FF" refers to filters created with transfer impedances measured in the big anechoic chamber. The performance of a delay and sum is shown for comparison.

D.3 Comments to these approaches

The results presented here show that the filters created using model transfer impedances work very close to that obtained when measuring the transfer functions of each source. However, this is not always the case. The previous results are measured in the small anechoic chamber of the ISVR, which is not a totally anechoic environment. The obtained results, are hence biased, and show a strong comb effect due to a strong reflection. Another set of measurements is presented on Fig. D.5, in where the performance is assessed for the 1×8 array.

Although at some frequencies the performance for the filters based on Green functions or in a single source is higher, the low frequency performance is improved significantly by the filters created with the measured transfer impedances. The improvement of performance at low frequency is due to the fact that the condition number at low frequency is greater, being the system less robust to errors and mismatches at these frequencies. As the real transfer impedances differ from the ones used for calculating the filters based on Green functions or on a single source, the obtained volume velocities are influenced by these little differences, giving a lower performance.

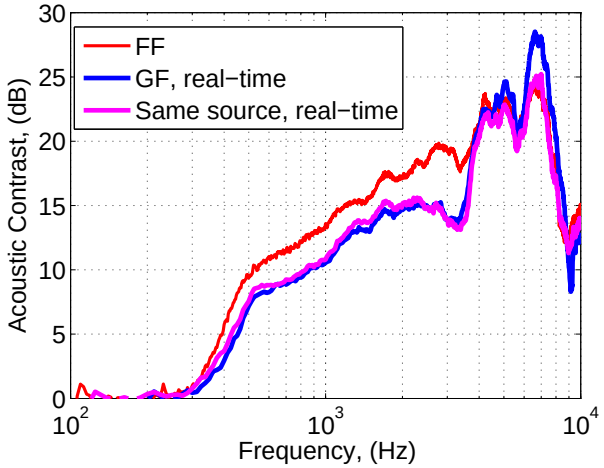


Figure D.5: Results of real-time measurements with the filters calculated assuming that all the sources behave equally, "Same source", and also with point hypercardioid Green function transfer functions, 'GF, real-time". "FF" refers to filters created using the measured transfer impedances of each source.

D.4 Conclusions

The approaches presented here show how the filters of the array can be created in a more simple way by modelling the transfer impedances from each source to each control zone. Two approaches have been presented, one based on point source transfer impedances, and the other based on the transfer impedances of a single source. The obtained results are just slightly lower than that obtained by measuring the transfer functions of the whole array in the free-field. In order to use these approaches, the filters need to be created using strong regularisation, to overcome the differences in response between the model and the real transfer function.

An example of the reverberant performance obtained with filters created using transfer functions based in point source Green functions is contained in Section 5, where it is seen how the performance is identical to that obtained using measured transfer functions.

Appendix E

A study of the steering capabilities of a phase-shift source array with respect to a dual layer array

The use of passive directional speakers for reducing the radiation towards the back of an array is advantageous respect to the use of a dual layer array, as half the number of digital filters is needed, which makes a more economic and robust array. However, a dual layer array is more flexible than an array of passive directional radiators in terms of steering, as the directivity of those last ones is fixed, and normally pointing to the front of the array. In order to assess to which extent it is desirable to use each of these two kind of radiators, a set of simulations employing pairs of monopoles and point hypercardioid Green functions initial simulations is carried out to test the capacity of beamforming at different angles. As to control the sound radiation, least squares inverse filters have been used.

E.1 Simulations based on arrays of point sources

A double layer array has been considered, which is formed of a row of eight double point monopoles, with a total of 16 independently controlled sources. This array is compared with an array which uses 8 point hypercardioid sources, each of them perpendicular to the main axes of the array.

The beamforming that both arrays obtain is benchmarked using a control geometry as that shown in Fig. E.1. This control geometry has a bright control point, which is steered in increments of 12° between broadside, 0°, and 84°, and a total

of 258 dark points, which are all equally spaced.

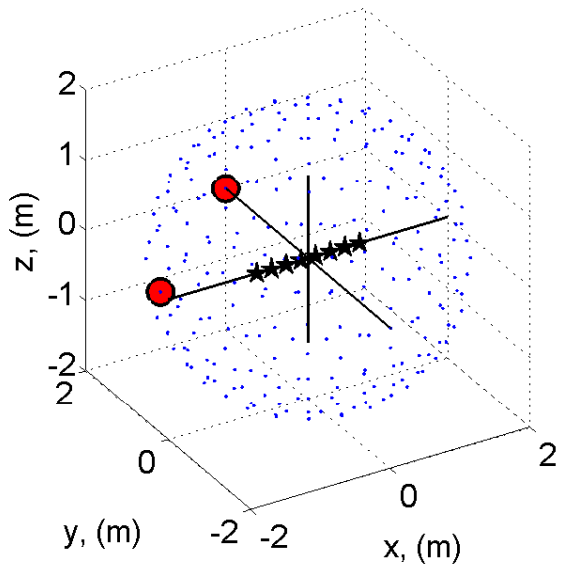


Figure E.1: Control geometry used for beamforming at different angles. The red points symbolise the beamforming direction, when the bright zone is selected to be broadside and when this is selected to be 84°.

The acoustic contrast performance of both arrays for different steering angles in such control zone is shown in Fig. E.2. These results have been obtained limiting the normalised array effort to be lower than 20 dB. Although in principle it may be thought that the performance of the hypercardioid sources is only greater at steering positions close to broadside configurations, the results show that this is good for a broad range of steering angles. The acoustic contrast for the phase-shift sources array is greater at every frequency from 0° to 48°. When the array is steered at 60° the performance of the dual layer array is greater at high frequency, and at 84° the performance of the dual layer array is greater throughout the whole of the frequency range.

Fig. E.3 shows the x, y and y, z and vertical projections of the directivity balloons for both arrays, for $f = 2500$ Hz. This form of plotting a directivity is convenient, as it can reveal details which are not seen from directivities plotted from an horizontal or vertical measurement slide. The directivity balloons corresponding to the steering angles between 0° and 48° complement the results shown in Fig. E.2, showing, in the x, y projections, how the dual layer array creates a larger number of secondary lobes. At 60° the array of hypercardioid sources starts however to create larger secondary lobes than the dual layer array, which is more noticeable as the steering angle increases. At these steering angles the pressure radiated by

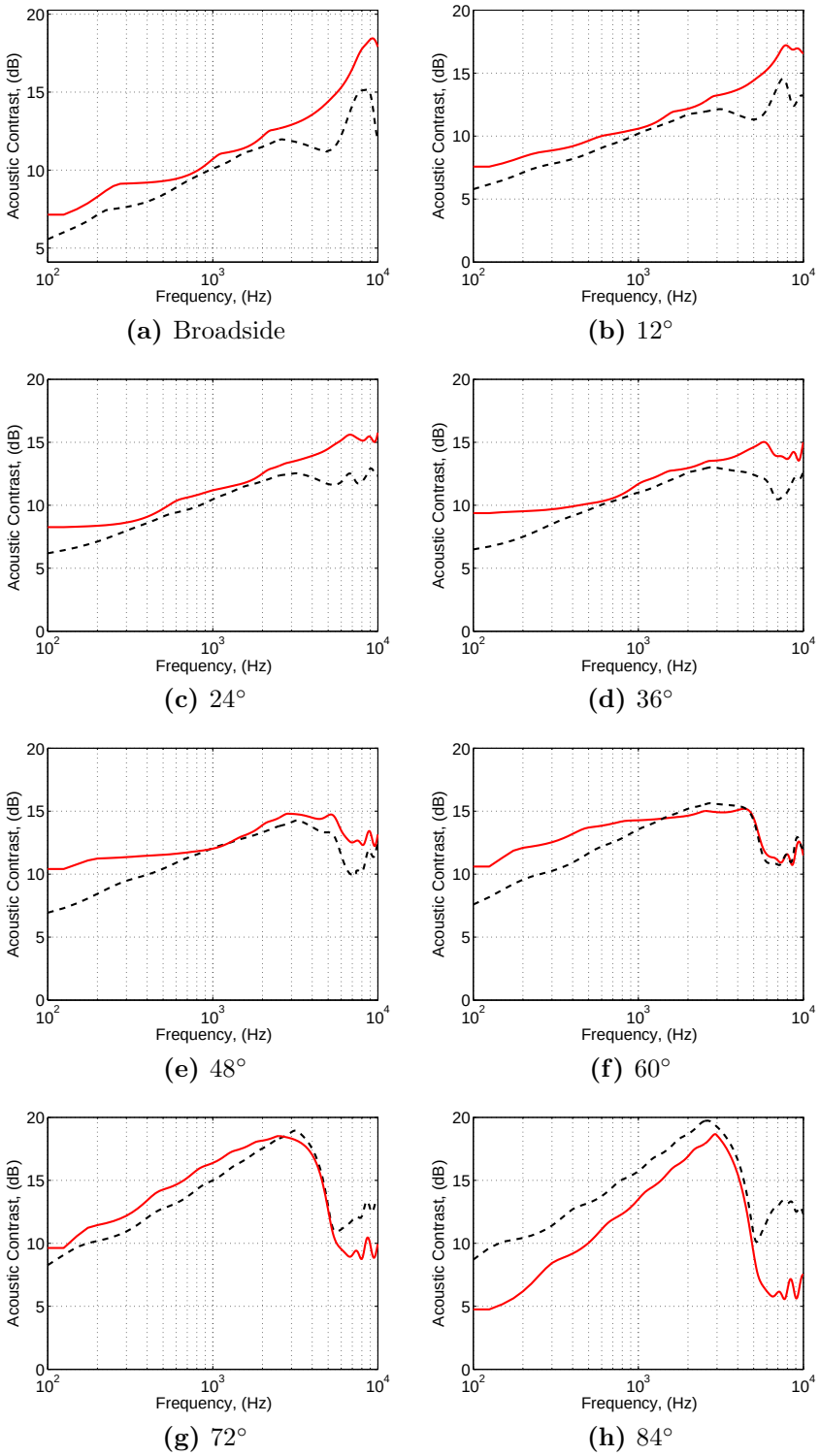
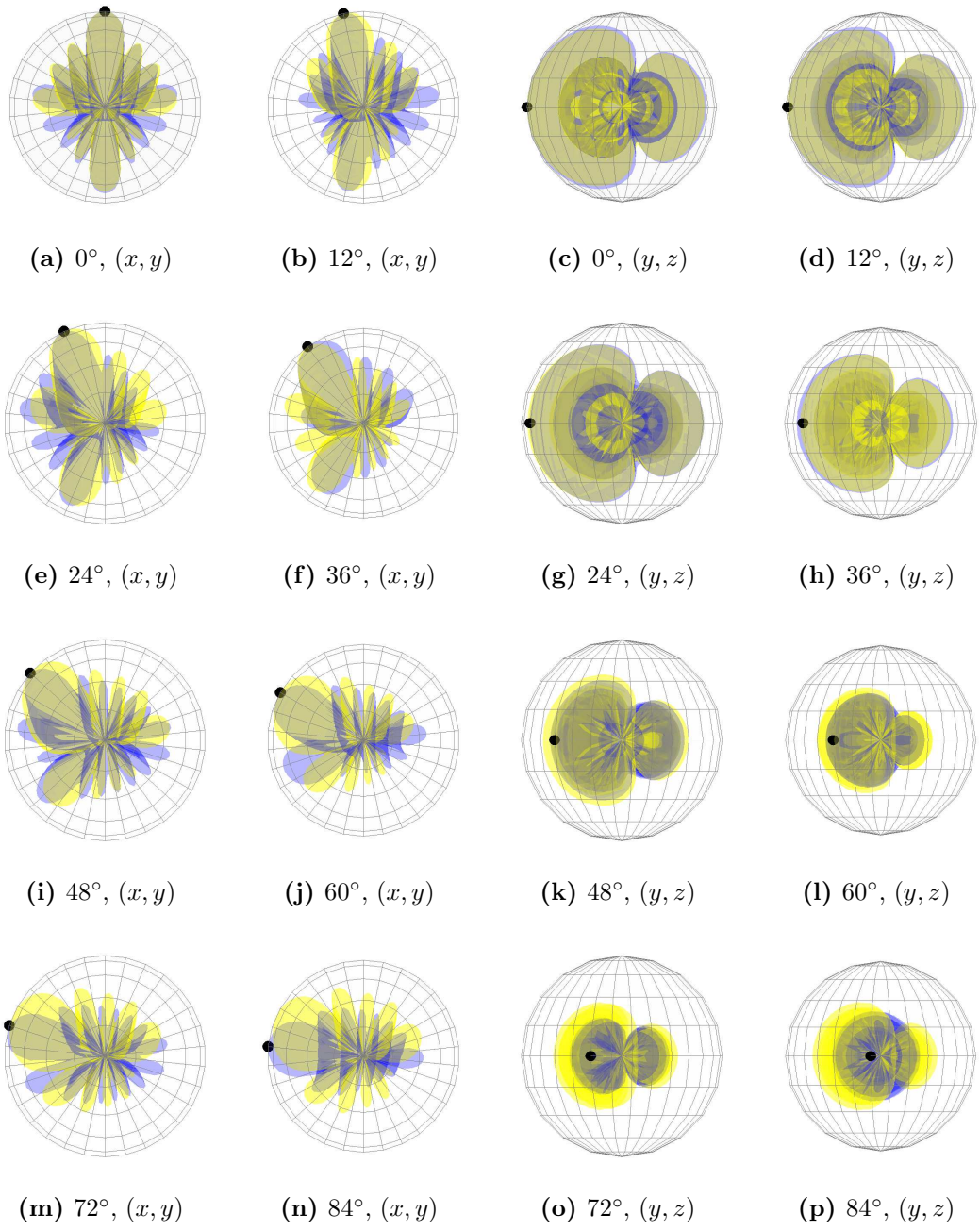


Figure E.2: Acoustic contrast for a 8 point sources hypercardioid array (red solid line) and for a 8 point double monopoles array (black dashed line) for different bright zone steering angles. The normalised array effort has been limited to be lower than 20 dB for both arrays.

the hypercardioid sources is much lower than that radiated on axis, which makes the array unable to steer effectively. This is particularly observed at 72° and 84°, where it is seen how the radiation pattern which maximises the contrast does



(a) 0°, (x, y) **(b)** 12°, (x, y) **(c)** 0°, (y, z) **(d)** 12°, (y, z)

(e) 24°, (x, y) **(f)** 36°, (x, y) **(g)** 24°, (y, z) **(h)** 36°, (y, z)

(i) 48°, (x, y) **(j)** 60°, (x, y) **(k)** 48°, (y, z) **(l)** 60°, (y, z)

(m) 72°, (x, y) **(n)** 84°, (x, y) **(o)** 72°, (y, z) **(p)** 84°, (y, z)

Figure E.3: x, y and y, z projections of the 3D directivities at $f = 2500$ Hz for the 8 point sources hypercardioid array (yellow) and for the 8 point double monopoles array (blue), given at different steering angles. The plots are normalized in a dB scale to the point of maximum radiation. The closed point shows the beamforming direction.

not place the main lobe in the beamforming direction. By looking at the (y, z) projections it can be seen how between 0° and 72° both arrays present a hypercardioid radiation pattern, which is slightly smaller in the case of the phase-shift sources array. It can be seen how the amplitude of the hypercardioid decreases as the arrays are steered towards an endfire position. For 84° it is observed that the dual layer array radiates almost omnidirectionally in the (y, z) plane, as an endfire array, whilst that the hypercardioid sources array still radiates following

a hypercardioid radiation pattern.

The use of point hypercardioid sources to simulate a phase-shift hypercardioid sources array assumes that the sources present a constant directivity along all the frequency range. This is not the case however in a practical array, where the directivity of the sources falls apart at a given frequency. The performance given by the dual layer array is lower at many steering points from that of the hypercardioid sources array, as the array effort increases due to the use of double the sources. In this simulation, it is not taken in account, however, that when a driver is put inside a phase-shift enclosure its frequency response changes, making it less sensitive at low frequency. If this would be taken into account, it could be that both arrays would present the same low frequency performance. These is confirmed in the results shown in Section 3.3, where it is seen that both arrays, the phase-shift array and the dual layer array obtain the same directional performance.

E.2 Conclusions

The simulations presented here have shown that an array of point hypercardioids presents a directional response which is slightly better from that of a dual layer array, for steering orientations between broadside and 60°, using the same amount of superdirective filters.

Appendix F

Effect of a hard reflecting surface at the back of an array

This appendix presents an investigation of the performance of the soundfield created by an 8 source line array when a reflective hard wall is placed at the back of the array. This is important, as the array is assumed to be placed below a TV and TVs are normally placed very close to walls. As the distance at which the TV is placed modifies the radiated soundfield, it is interesting to know how and how much the soundfield is modified by the reflections from a close surface.

A previous study was performed in [94], in where a plane surface was placed at the back of the array, showing little modification of the created soundfield. In this work, simulations are first carried out using the method of acoustic images, where the array and its reflection from the wall are represented by point sources. This allows to estimate the frequency response in the listening zone and the directive performance, both under the influence of the wall.

A set of field measurements are also introduced, in where the array is placed pointing upwards in a hemianechoic chamber, hence using the floor of the chamber as a reflecting surface. Both the point source simulations and the measured results show a performance which decreases proportionally to the separation between the array and the back reflecting source.

F.1 Free field environment simulations based on the method of acoustic images

The reflection of a source with a rigid boundary can be amended with the *acoustic imaging* method [246, 130]. Let a source of spherical propagation be placed in a medium which extends throughout the space. If this source is located on the y axis at a distance d from the origin, a spherical wave exists that propagates according to

$$p_i = \frac{A}{r_-} e^{-jkr_-}, \tag{F.1}$$

where

$$r_- = \sqrt{x^2 + (y-d)^2 + z^2}, \tag{F.2}$$

being r_- the distance to each field point from the point $(0, d, 0)$. If a second source, the *image*, of equal strength, frequency, and initial phase is placed at $(0, -d, 0)$

$$p_r = \frac{A}{r_+} e^{-jkr_+}, \tag{F.3}$$

where

$$r_+ = \sqrt{x^2 + (y+d)^2 + z^2}, \tag{F.4}$$

the normal component of the velocity vanishes on the x, z plane. The fluid at the negative point of the x, z plane can be replaced by a rigid boundary at $y = 0$. The pressure in the region $y > 0$ is given by the summation of Eqs. F.1 and F.3.

$$p = p_i + p_r = A(\frac{1}{r_-} e{-}jkr_- + (1-\alpha)\frac{1}{r_+} e{-}jkr_+). \tag{F.5}$$

However, this approximation is only valid in the case of a reflection with a rigid wall [130]. If the boundary is not rigid, the method of images can be used as a reasonably good approximation only if the source is many wavelengths away from the boundary. For the results presented here, it was desirable to estimate small separation distances between the array and the wall, so that only a rigid boundary is considered.

The performance of the array is assessed when a hard wall is placed at the back of the array at a certain distance. The distances used for the study are 1 cm, 10 cm, 20 cm and 50 cm. The directive performance is measured in a control zone as that shown in Fig. F.1, which posses a total of 97 control points and a bright zone extending from 0° to 15°.

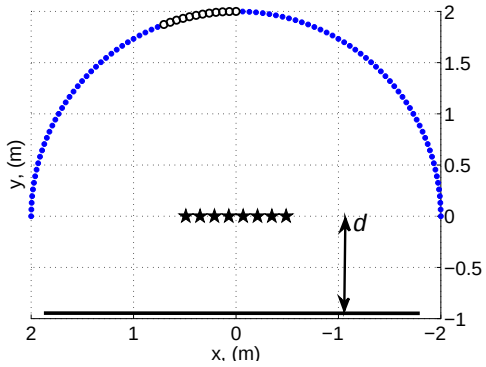


Figure F.1: Control geometry used to measure the performance under the influence of a wall. Blue dots represent dark zone microphones and open circles represent bright zone microphones. The stars represent the array sources and d is the distance between the array and the wall.

In the following results three different graphs corresponding to different configurations are plotted, which are all based on the control geometry of Fig. F.1. The three different configurations are:

- "Free field": The performance is estimated in the free field with filters calculated based on free field transfer responses.
- "Wall": Being the array driven by filters created via free field transfer responses, a wall is placed at the back of the array.
- "TFs with wall": In this case, the array is driven with filters which take into account the influence of the wall.

The arrays are driven with least squares inverse filters (LSIF) [50] with the array effort limited to be lower than 6 dB.

The only reflection due to the wall modifies the pressure soundfield radiated around the array. The summation between the direct pressure from the array and the reflected wave from the hard surface causes a comb filtering effect in the frequency domain. The frequency response of a comb filter is given by [77]

$$|Z(j\omega)| = \sqrt{(1+\zeta^2) + 2\zeta\cos(\omega\tau)}, \tag{F.6}$$

where $\zeta = 1-\alpha$ is the amplitude of the reflected wave, α is the absorbent coefficient of the wall and τ the time delay between direct and reflected signals, being the notchs in frequency spaced each $1/\tau$ Hz. For the distances between the array and the wall considered here, the theoretical frequency spacing of the comb filter

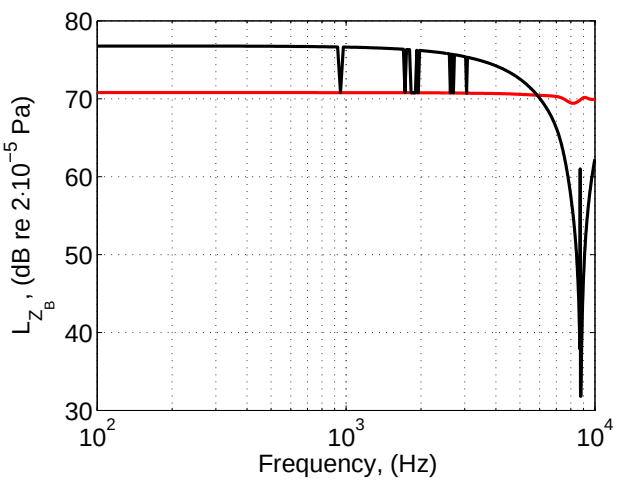
nulls can be observed in Table. F.1.

Table F.1: Frequency spacing of the cancelling effect due to the first reflection introduced by the wall.

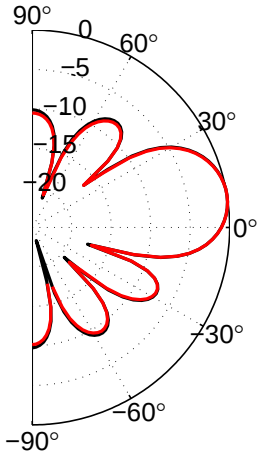
d, (cm)	1	10	20	50
$\mathbf{\Delta_f}$**, (Hz)**	17150	1715	860	344

By observing the top and middle left hand side plots of Figures F.2 to F.4 it can be seen how the nulls in the frequency response, $\mathbf{Z}_B$, and also in the acoustic contrast responses are spaced about the amount given in Table F.1 with the first null in the directivity given at a frequency $f = c/(2d)$. The results show how the on-axis response is largely reduced at the notch frequencies, which leads to a correspondent reduction in acoustic contrast. As the array is moved apart from the wall, the time delay of the first reflection increases and hence the frequency spacing diminishes. The level of the first reflection is smaller and hence the quality factor of the notches is also reduced. When the distance with the wall is large, i.e. 50 cm, the reduction in performance is more noticeable at high frequency.

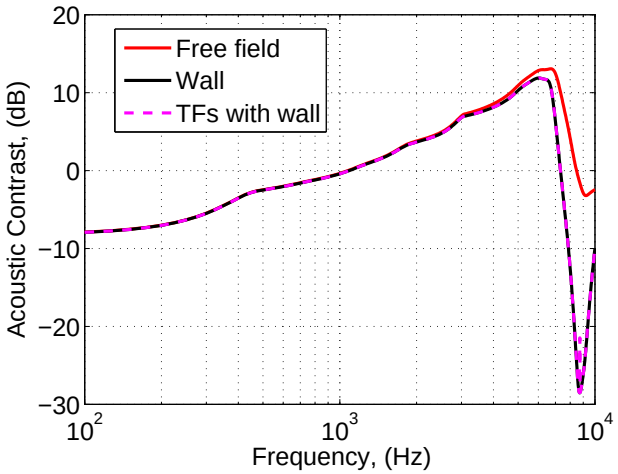
Similar results are obtained by placing the wall behind the array, as in the "Wall" configuration, and those in which also the reflection from the wall is taken into account when calculating the optimal set of array filters, as in the "TFs with wall" configuration. The later set of filters is able of obtaining a slightly higher amount of acoustic contrast at frequencies just above each performance notch. By observing the array effort plots it can be seen how at frequencies close to the notches in response, a small reduction is obtained. However, a small increase in the array effort above the limiting figure, 6 dB, is present at certain frequencies when the wall is not considered for the calculation of the filters. An example can be observed in Fig. F.3, where about 500 Hz there is a small peak in the acoustic contrast.



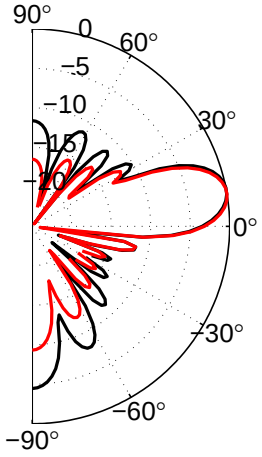
(a) $\mathbf{Z}_B$ averaged response



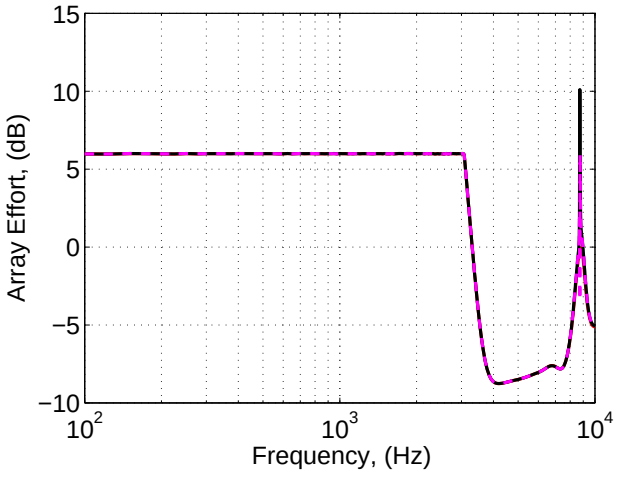
(b) 2300 Hz



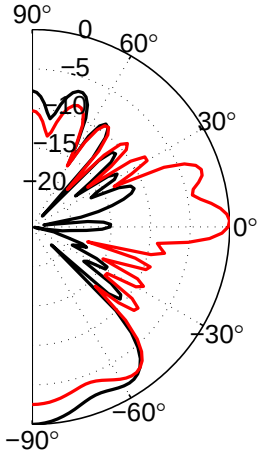
(c) Acoustic Contrast



(d) 6900 Hz

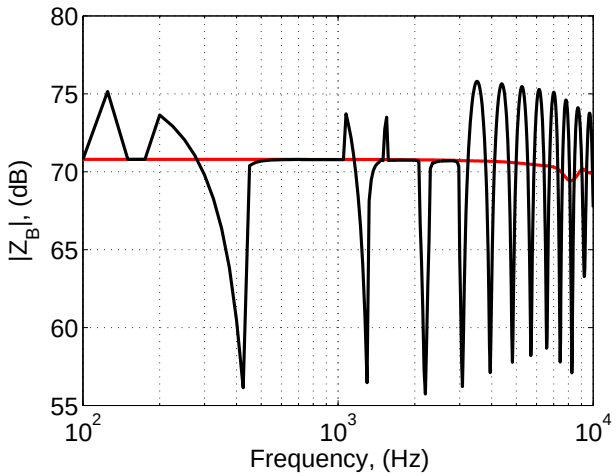


(e) Array Effort

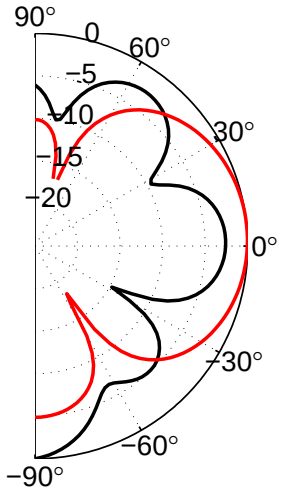


(f) 8700 Hz

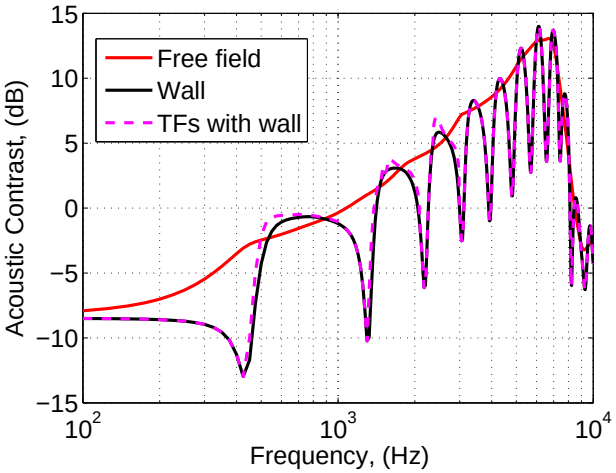
Figure F.2: Responses for a 8 point monopole sources array with a wall 1 cm apart from its back. The right hand side plots show the directivities at various frequencies.



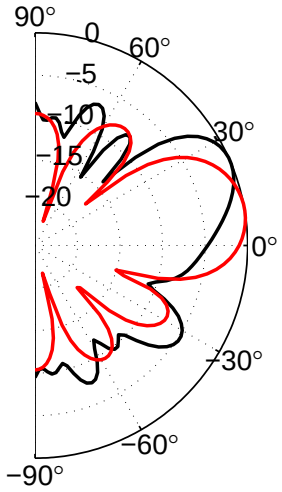
(a) $\mathbf{Z}_B$ averaged response



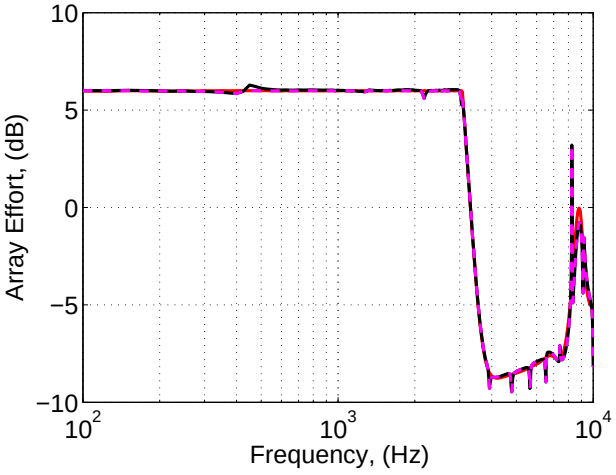
(b) 475 Hz



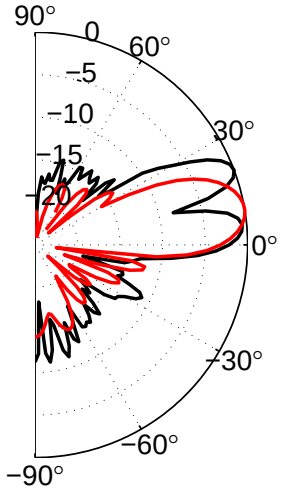
(c) Acoustic Contrast



(d) 2600 Hz

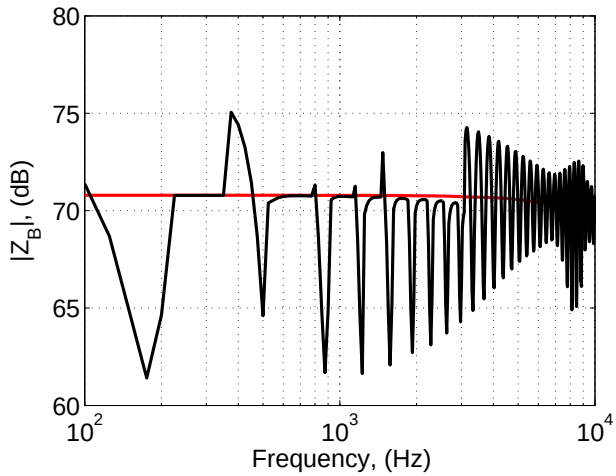


(e) Array Effort

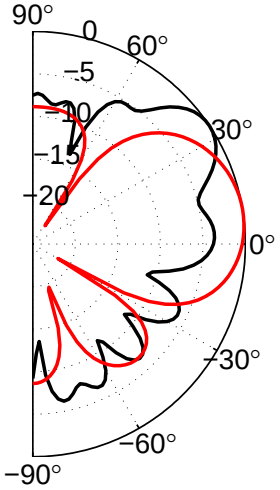


(f) 6600 Hz

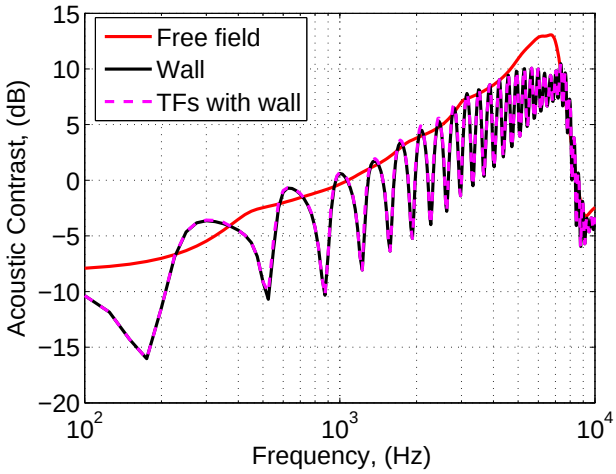
Figure F.3: Responses for a 8 point monopole sources array with a wall 20 cm apart from its back. The right hand side plots show the directivities at various frequencies.



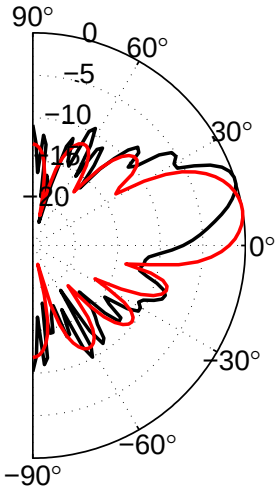
(a) $\mathbf{Z}_B$ averaged response



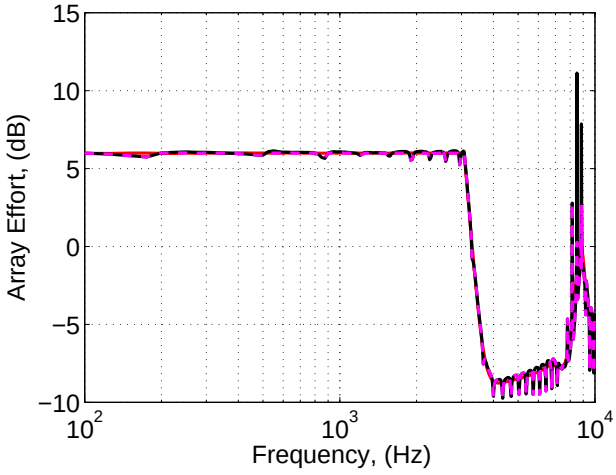
(b) 1225 Hz



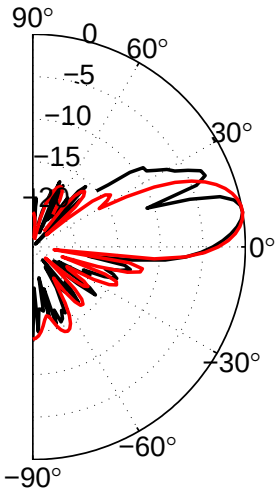
(c) Acoustic Contrast



(d) 3600 Hz



(e) Array Effort



(f) 6600 Hz

Figure F.4: Responses for a 8 point monopole sources array with a wall 50 cm apart from its back. The right hand side plots show the directivities at various frequencies.

F.2 Influence of source directivity

The directivity of the individual sources of an array plays and important role on how big the reflections coming from the wall are. In Fig. F.5 it is shown the average reduction in acoustic contrast that occurs when an array is moved away from a wall. The results are calculated for a control geometry as that shown in Fig. F.1, for two 8 point source arrays using either monopole sources or hypercardioid sources. The reduction that the monopole array oversees when this is moved away is much greater than that observed by the hypercardioid array, due to the lower level of reflections produced by the later one.

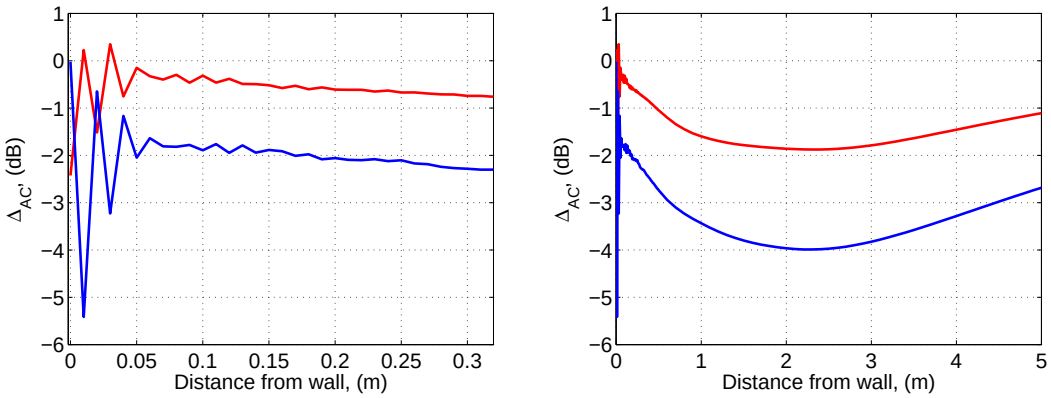


Figure F.5: Variation of the frequency averaged acoustic contrast as a function of distance, calculated in a control geometry as shown in Fig. F.1 when a rigid wall is placed at the back of the array. The red line results refer to an array of 8 point hypercardioid sources and the blue line refers to an array of 8 point monopole sources.

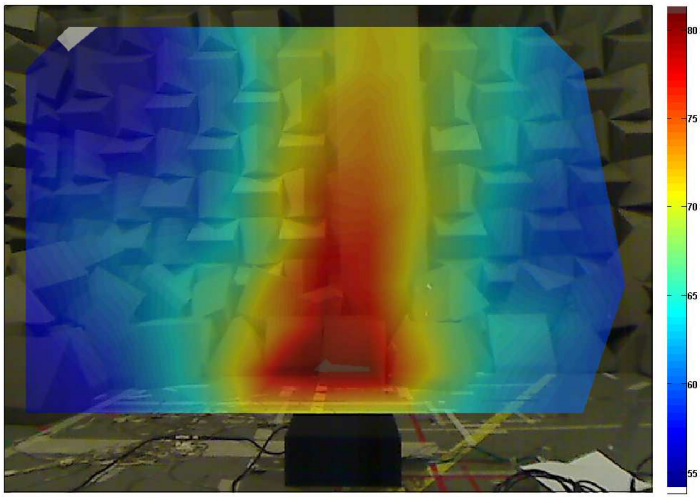
It can be observed in Fig. F.5 how the reduction in contrast is lower for the array of hypercardioids, as the level of the reflection with the wall is reduced. If the array of hypercardioid sources is placed at 0 cm from the wall, the performance is however decreased about 2.5 dB. This is due to the fact that the summation of two hypercardioids that point to opposite directions leads to a omnidirectional radiation pattern. Moving slightly away from the wall the results of both arrays present a large ripple, which is stabilised when the array is about 5 cm apart from the wall. This strong ripple is caused by the cancellation that occurs when the array is very close to the wall and the reflection is large, as can be observed in Fig. F.2, where the reduction in contrast is about -25 dB. When the wall is very far from the array the reflections are much lower and the acoustic contrast increases again.

F.3 Real time measurements of performance

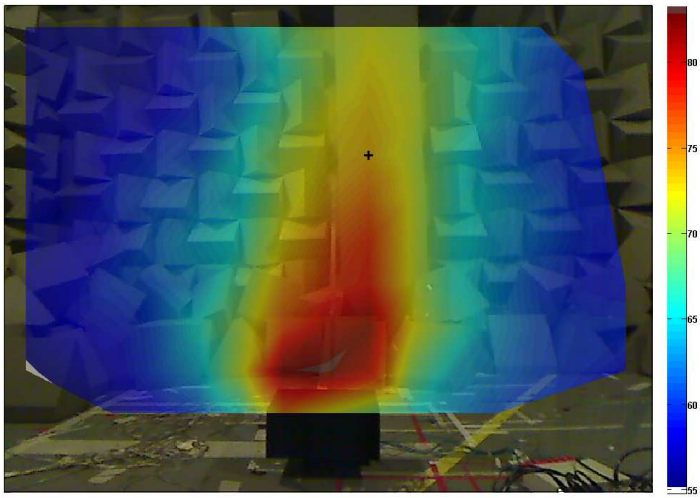
A set of measurements is presented which have been performed in a hemianechoic environment, in order to create an environment similar to the simulations performed using the method of acoustic images. The measurements have been performed in an anechoic chamber with the floor removed. The array has been placed pointing towards the ceiling, so that the floor is acting as a wall. As these measurements needed a large number of control points, the Scan & Paint [247] method has been used to make a faster and less complex estimation of the acoustic field. By the use of this technique, the sound field produced by the array has been mapped when the array is placed on the floor, and at 20 and 50 cm from it. The sound field has been sampled afterwards to lead to a control geometry similar to that shown in Fig. F.1, allowing to estimate the acoustic contrast.

Fig. F.6 shows the acoustic pressure maps averaged in the frequency bandwidth from 500 Hz to 8 kHz of the array at various distances from the floor. It is possible to observe how the soundfield maps are very similar for the cases when the array is on the floor or it is at 20 cm from the floor. When the array is at 50 cm from the floor, however, the main lobe of radiation has broadened.

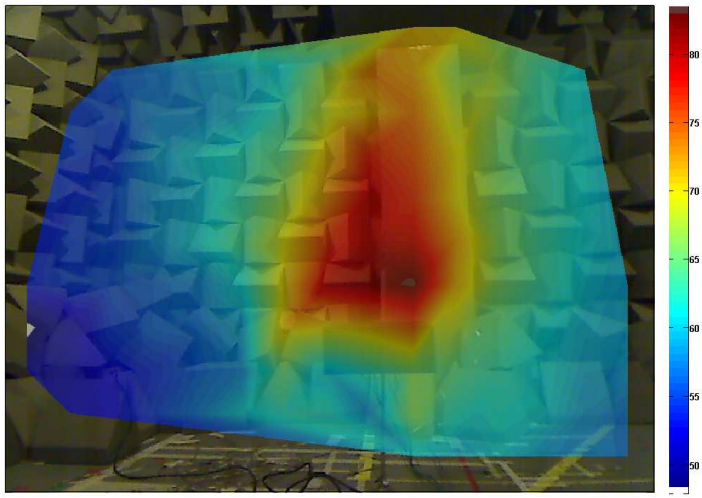
The acoustic contrast has been estimated by sampling the pressure maps at the points given by Fig. F.1, which leads to the results of Fig.F.7. As in the case of the point source simulations, the measured results show how the performance when the array is on the floor or at 20 cm is very similar. When the array is at 50 cm from the floor, the performance is about 2 dB to 5 dB lower along all the frequency band, as in the results presented in Fig. F.4.



(a) Floor



(b) 20 cm



.

(c) 50 cm

Figure F.6: Acoustic pressure maps captured with the Scan & Paint method [247] when a reflective surface is placed at the back of the array.

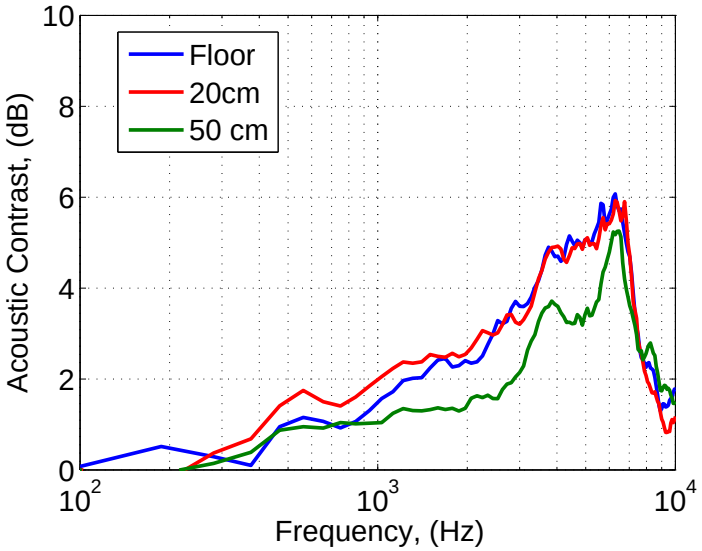


Figure F.7: Acoustic contrast results for the array at different distances from the floor of a hemianechoic chamber, which simulates a wall placed behind the array.

F.4 Conclusions

Reflections from surfaces close to the array will decrease the directional performance of an array. In this appendix, the effect that a wall placed at the back of the array has been studied, as TVs are typically placed close to a wall.

It has been shown that an array of hypercardioid sources is in clear advantage with respect to an array of monopole sources, as the on axis response is less affected by reflections from the wall. The closer that an array is placed to a wall, the lower the reduction in performance due to reflections, with the exception of very close distances for an array of hypercardioid sources. This fact is common in the simulated results and in the measured results. The measured results, however, are not so critical with the distance at which the array is placed from the wall when this distance is small, obtaining similar results for the cases when the array is at 20 cm from the floor or when this is on the floor.

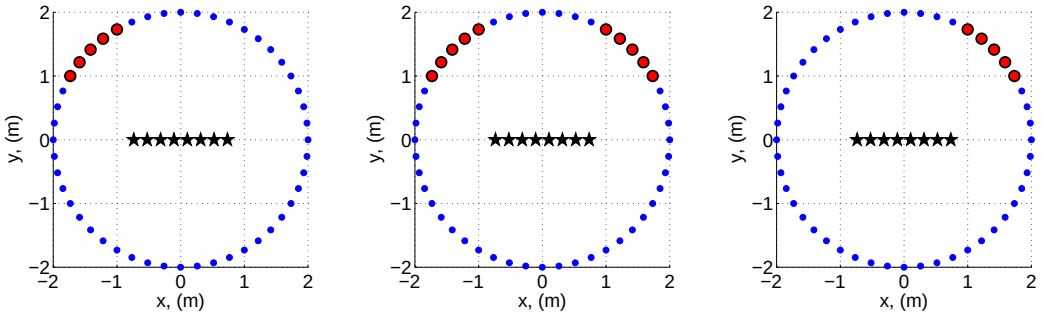
Appendix G

Case study: masking as an aid to personal audio

Masking could be beneficial for some sound zoning, or personal audio, applications. An example may be a situation where a personal audio device is used for creating a pair of beams, each one radiating a different speech aimed for different users. As the array may not work ideally at some frequencies, this can be detrimental to the acoustic contrast between both zones. At this point, a lower frequency masking sound can be of help in order to cover the disturbance generated by the unwanted speech.

G.1 Reproducing various speech programmes inside same room

An example of a personal audio application is that in which two listeners watch a film in different languages in a common environment. In this case the speech program of each listener is different, but they share the same audio and sound effects background soundtrack. To this end, the control geometries in Fig. G.1 may be used. Although this program requires a very good audio quality and an adequate mix of the material being reproduced, a first mixing approach may consist of reproducing speech in one language in a bright zone steered to the left, and reproducing speech in the other language in a bright zone steered to the right. As the background sound track is identical for both languages programme, this is reproduced in both bright control zones with same level.
The estimated SPL produced in each control zone by the array has been calculated off-line using measured transfer functions, and it can be observed in Fig. G.2. The background noise level is on average 20 dB SPL lower than that of the speech levels. The figure shows three different quantities in broadband and in



(a) Left control geometry "L" **(b)** Background control geometry, "BKG" **(c)** Right control geometry, "R"

Figure G.1: Control geometries for creating three different sound programmes in a 2D horizontal plane.

$1/3-rd$ octave bands: the spectrum in the left bright zone due to the program intended to be reproduced in the left bright zone, "L in L", the spectrum in the left bright zone of the program intended to be reproduced in the right bright zone,"R in L", considered as the interferer, and the spectrum of the background soundtrack, "BKG". The audio program present in left and right control geometries is selected to be the LTASS of normal speech and the audio program present in the background zone is the LTASS of music.

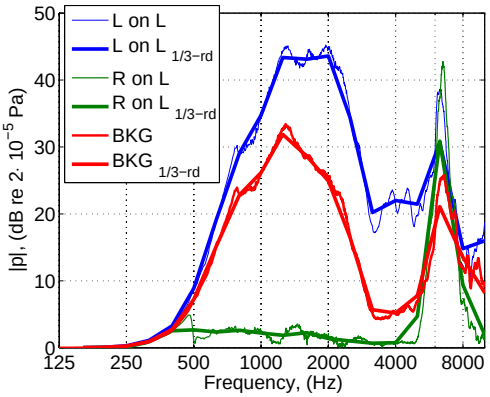


Figure G.2: LTASS of the program intended to be reproduced in zone L in zone L, from the program intended to reproduced in zone R in zone L, and from the background soundtrack, intended to be reproduced in both zones, L and R.

If the masking that each 1/3-rd octave band produces in a normal listener is calculated using the cochlear model, the result for a 20 years old listener is shown in Fig. G.3. The masking patterns for the program intended to be reproduced in the left zone, "L in L", are much greater than the interferer for almost all the frequency band, as expected from the directivity pattern of the array. The level of the interferer is, however, similar to that of the intended program at some frequencies. At low frequency, the level of the interferer, "R in L" is similar to that of the program, "L in L", as the array is quite omnidirectional and at 6.3

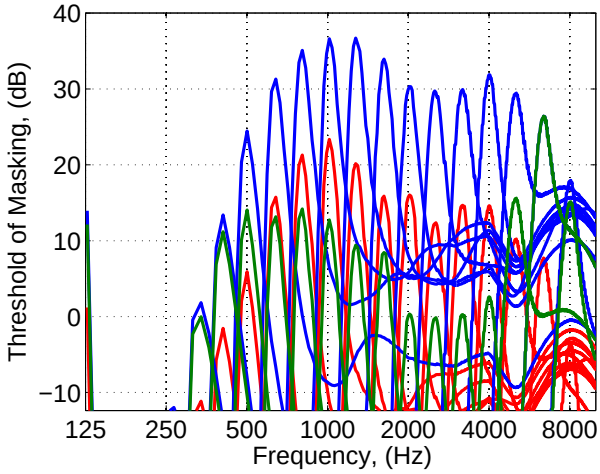


Figure G.3: Thresholds of masking produced by each of the LTASS shown in Fig. G.2 in $1/3-rd$ octave bands. The blue lines show that produced by the program in its own zone, i.e., "L in L", that produced by the interferer, "R in L", and that produced by the background program ,"BKG".

kHz both the interferer and the program are identical due to the directivity break up of the phase-shift sources . The effect of separation between program and the interferer is greater increased by the background soundtrack, "BKG", as the predicted masking thresholds are well above those of the interferer, except from low frequencies and for the tones corresponding to 5 kHz and 6.3 kHz. Although the masking curves predict that the program produced by the array is greater than the interferer, the quality of the speech separation may be compromised at low frequency and at around 6.3 kHz.

In order to palliate this problem, an alternative equalisation can be used. As to get a HI-FI reproduction of the audio program an alternative low frequency sound reproduction mechanism is needed, such system can be used to mask the interferer. Fig. G.4 show the masking predicted by such sound systems, calculated at the $1/3-rd$ bands of 250 Hz, 315, 400 and 500 Hz, which is equalised to produce the level shown in Fig. G.2 for the background program. As observed, the program produced by the low frequency sound-system masks that produced by the array at those frequency bands, and so the lack of directionality and separation between both speech programs is amended. The lack of separation between both programs at 6.3 kHz is, however, more difficult to adjust, and is only possible to be overcomed by attenuating the output of the array at such frequency, so that it is masked by lower frequency tones, which can result in a possible reduction of the speech quality and the speech intelligibility. Nevertheless, this should not

[0]See comments of Section 2.4.

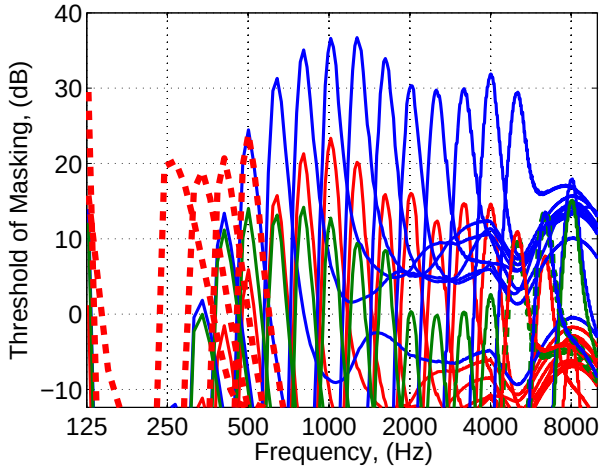


Figure G.4: Thresholds of masking using an alternative equalisation to that of Fig. G.3 in order to maximise the separation between the two speech programs. The thick red dashed lines show the masking patterns produced by the low frequency sound system, and the thick green dashed show the masking for the 6.3 kHz after the output of the array has been reduced at such frequency.

represent a large decrease of the reproduction quality, as the quality of the speech and the contribution to speech intelligibility is most attributed to the first formants, which are typically placed at lower frequencies [241].

G.2 Conclusions

The masking model used in Section 7.2.2 has also been used to estimate how masking can be beneficial for personal audio applications. In this case the response is equalised so that a disturbing sound produced by an interferer personal audio program is masked and minimises the distraction from a target program. The case study application presented has been that of using an array for reproducing the speech material of a film to two listeners sat in a sofa in two different languages. As the soundtrack material of the movie is the same for both listeners, this is equalised so that at the frequencies where the program separation is minimum this is equalised to mask the lack of program separation.

Although the effective separation provided with such equalisation has to be verified with measurements, the masking curves predicted by the model can be used as a first approach to find an equalisation which provides a suitable separation between both programs. It is also needed to take into account that the predictions given by the model are based on long term average spectra, and that the performance of the audio separation between left and right control zones depends

on the instantaneous level of each signal.

Appendix H

A physiologically reasonable lumped parameter model of the human organ of Corti

It is important to construct models of cochlear mechanics to test our comprehension of the peripheral hearing system and the physical processes involved. Modelling helps our understanding by providing predictions of system responses, based on assumptions of the underlying physical mechanisms. The action of the cochlea is often modelled as a bank of filters, such as gammatone or gammachirp filters which are able of replicating the key features of the auditory system, such as tuning, critical bands compression and frequency- and level-dependent responses [186]. These models, however, reproduce the overall psychoacoustic responses without directly linking these outcomes with the fundamental biophysical details and underlying cellular sub processes.

One of the first resonant models was introduced by Allen [248] in 1980, where the author presented a relation between shear and bending motion of the TM. The model did not include OHC motility, but instead Allen proposed that the TM formed a second resonant system with a slightly lower resonant frequency than the resonant system according to a specific BM position, which allowed sharp peaks to be obtained in the resulting frequency response curves.

A similar model was published by Neely and Kim [249], in which the coupling between TM and the OC was achieved by a negative damping. The negative damping was interpreted as an energy-providing or active element, which could represent some physical action of the OHC. Neely and Kim revised this model in 1986 [202], where they related the OHC with the cochlear amplifier (CA). This model

has been employed in various research of the cochlear dynamics [224, 250, 201].

These models are able of reproducing the frequency responses of the cochlea, however, they are not based on a mechanical representation of the physiology of the organ of Corti. This fact makes these models not able to replicate with precision some effects found in the temporal response of the cochlea, as for example the near-zero invariance of the IRs with input excitation or the frequency glides [251].

The model presented here is based in a third degree of freedom system, in where the masses can be associated with the BM, and the transverse and shear motion of the TM. The motion in both axis is translated into a single axis using the relations presented by Allen [248], and discussed by Elliott *et al.* [252, 253, 189]. Models using similar dynamics have been previously developed by Geisler *et al.* [254] and by Ramamoorthy *et al.* [190], although, these models use a feedforward approach to the active pressure due to the OHC forces, which is not used here.

The model is based in that presented by Ramamoorhty *et al.* [190], with a feedback force depending on the relative displacement of TM and RL used to model the action of the OHC. The relative displacement of TM and RL causes the displacement of the fluid present in the sub-tectorial space, which bends the cilia of the IHC and transmits information towards the auditory nerve (AN). The shearing between the RL and TM also excites the cilia of the OHC, which drives the CA [255].

The Ramamoorhty *et al.* [190] model was originally developed using finite elements, with parameters chosen to model a guinea pig cochlea. The model has been converted into a lumped-parameter model by Simón Gálvez and Elliott [191]. The parameters used in the guinea pig have been also modified so that the responses replicate those found in a human cochlea. This leads into a lumped parameter model of the organ of Corti in which the action of the CA is physiologically realisable.

H.1 BM and TM dynamics

Fig. H.1a shows a cross section of the cochlea, in which the organ of Corti (OC) is sitting on the top of the basilar membrane (BM). The main components of the OC are the tectorial membrane (TM) and the reticular lamina (RL). The space between the RL and TM is filled with fluid. In this space floats the stereocilia

of the inner hair cells (IHC), which have forward transduction properties [255]. The transduction of the IHC is proportional to its stereocilia deflection, which is dependent on the relative displacement of RL and TM. The BM and the RL are separated by the outer hair cells OHC, which account for the active enhancement of the mechanical response.

The arch of Corti is believed to pivot about its left bottom corner, which is attached to the BM and causes the OC to rotate as a more or less rigid body around this point. Similarly, the RL is believed to rotate about the top vertex of the arch of Corti [256]. In the passive cochlea, the OHC are not active, and

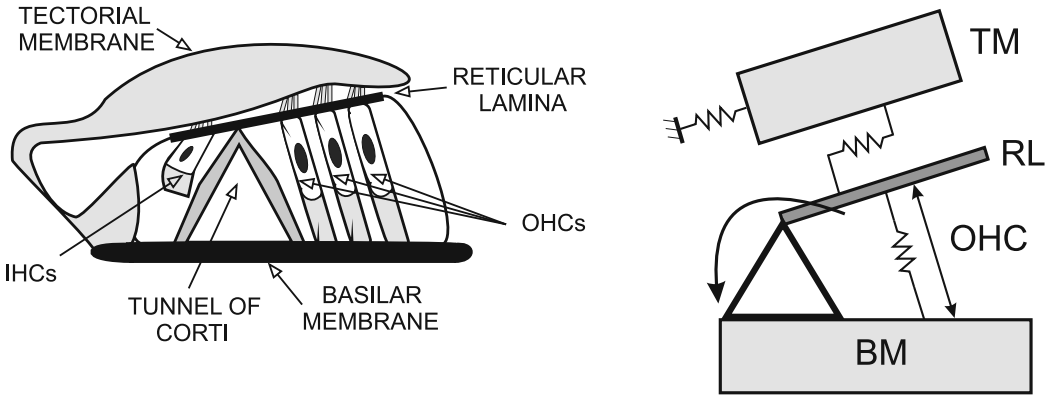


Figure H.1: Sketch cross section of the OC as it appears on the basal turn (left), and the equivalent lumped parameter representation (right).

these, together with the TM mainly act as mass loading in the BM. The passive response of the OC can thus be reasonably well represented as a single degree of freedom system. In order to recreate the active behaviour that leads to a greater frequency selectivity, it is needed to include one or two extra masses to create a higher order resonant system [248].

Fig. H.1b shows how the OC can be transformed into a lumped parameter model, where motion is assumed to be present radially (shear motion) and transversally (bending motion). This shear motion is then transformed into an equivalent bending motion, which can be performed by assuming that the RL pivots around the top corner of the tunnel of Corti. This geometry change was introduced by Allen [248] who assumed that BM and TM move parallel to one another. For the case of the model presented here, this parallel motion occurs between the TM and RL.

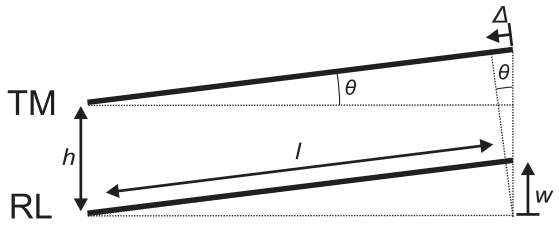


Figure H.2: Relation between shear and bending displacement on the TM.

As the two right angle triangles shown in Fig. H.2 are similar, the angle of rotation θ is also similar, which leads to the relation [252]

$$\sin\theta = \frac{w}{l} \text{ and } \sin\theta = \frac{\Delta}{h}, \tag{H.1}$$

so that

$$\frac{\Delta}{w} = g = \frac{h}{l}, \tag{H.2}$$

where g is known as the *shear gain* and depends on the geometry of the OC. Using this relation, it is possible to relate the transverse motion to the shear motion, i.e.,

$$w_{RL} = \frac{\Delta_{RL}}{g}. \tag{H.3}$$

Fig. H.3 shows the physical lumped parameter model of the cochlea and its equivalent form for entirely transverse motion. As can be observed in Fig. H.3b, the radial motion of the TM due to radial pressure p_{SR} is governed by the equation

$$p_{SR} = k_4'\Delta_{TM} + s^2 m_3'\Delta_{TM}, \tag{H.4}$$

where k_4' and m_3' are the physical shear stiffnesses and masses of the TM. This pressure is generated by the relative radial displacements of TM and RL

$$p_{SR} = k_3'(\Delta_{TM} - \Delta_{RL}). \tag{H.5}$$

We can define an equivalent transverse pressure acting on the RL due to the TM, p_{ST}. Since the work per unit area done by the RL in the shear direction, $\Delta_{RL}p_{SR}$, must be equal to that in the equivalent transverse direction, $w_{RL}p_{ST}$, then using Eq. H.3, the equivalent transverse pressure can be related to the shear pressure by $p_{ST} = gp_{SR}$. Since h is approximately equal to 1, a value of 1 is used for g here, so that the equivalent transverse pressure is equal to the physical radial

pressure. Using this relation and the relation of Eq. H.2, p_{ST} can be defined as

$$p_{ST} = g^2(k'_4 w_{TM} + s^2 m'_3 w_{TM}), \tag{H.6}$$

which can be written as

$$p_{ST} = (k_4 + s^2 m_3) w_{TM}, \tag{H.7}$$

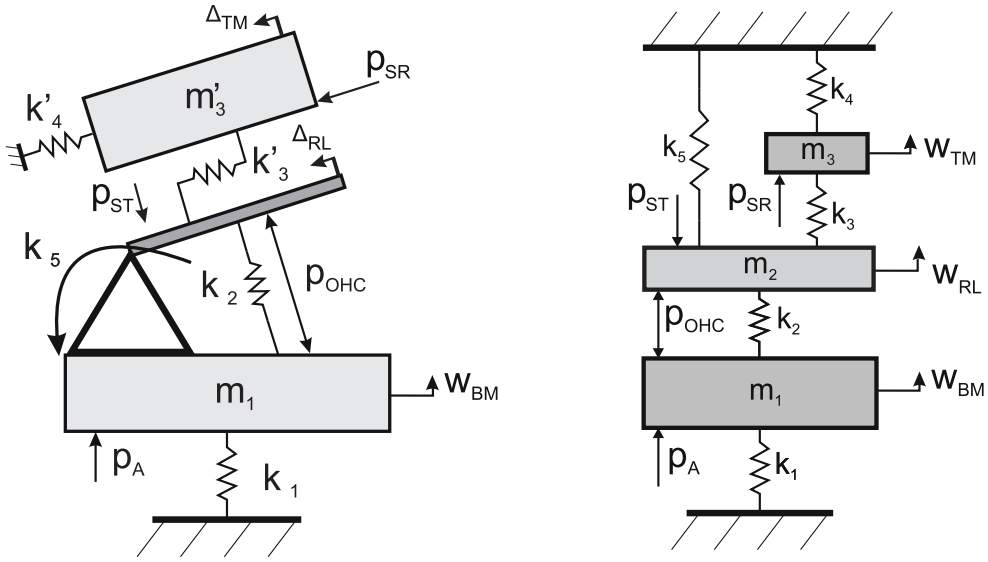
where $k_4 = g^2 k'_4$ and $m_3 = g^2 m'_3$ refer to the equivalent transverse stiffness and mass. Similarly, the equation describing the radial stiffness of the cilia, Eq. H.5 can be transformed into transverse motion to give

$$p_{ST} = g^2 k'_3 (w_{BM} - w_{TM}), \tag{H.8}$$

or

$$p_{ST} = k_3 (w_{BM} - w_{TM}), \tag{H.9}$$

where $k_3 = g^2 k'_3$. The transverse force required to drive the RL must, in addition to p_{ST}, also include that required to overcome the transverse inertia of the RL and TM masses, m_2, and the equivalent stiffnesses due to the RL rotation, k_5. By the use of these relationships, the OC model is transformed from having motion in two axes, as shown in Fig. H.3a to a model with motion in only one axis, as in Fig. H.3b, where p_A is the pressure difference due to the fluid in the cochlear chambers.



(a) Shear and transversal motion. **(b)** With motion in only one axis.

Figure H.3: Lumped parameter models of the OC. The damping associated with each stiffness is not shown for notational convenience.

In this representation m_2 represents the mass of the RL, although this is assumed to be a massless rigid body by Ramamoorhty *et al.*, plus the mass for the bending degree freedom of the TM. The mass m_3 accounts for the shear degree of freedom of the TM. Thus, the displacement between m_2 and m_3, i.e., RL and TM, is proportional to the deflection of the IHC.

H.2 Dynamics of the organ of Corti

Three equations of motion are introduced for the three masses which form the organ of Corti. A first equation describes the displacement of the BM

$$p_A - p_{OHC} = s^2 m_1 w_{BM} + k_1 w_{BM} + k_2(w_{BM} - w_{RL}), \tag{H.10}$$

where m_1 represents the mass of the BM, and w_{BM} its displacement, p_A is the force due to acoustic excitation and p_{OHC} is the force generated by the electrical excitation of the OHC. A second equation is established for the displacement of the RL

$$p_{OHC} = s^2 m_2 w_{RL} + k_2(w_{RL} - w_{BM}) + k_3(w_{RL} - w_{TM}) + k_5 w_{RL}, \tag{H.11}$$

where m_2 represents the mass of the RL, which is the mass of the TM in the bending mode, and w_{rl} its displacement. The equation which describes the equivalent shear motion of the TM is defined as

$$s^2 m_3 w_{TM} + k_3(w_{TM} - w_{RL}) + k_4 w_{TM} = 0, \tag{H.12}$$

where m_3 represents the transformed mass of the TM due to its radial motion and w_{TM} its equivalent transversal motion as discussed in Section H.1. Each stiffness is assumed to be complex and can be written as $k + sc$ where k is the real stiffness and c the associated damping. The receptance for each mass when these are driven by p_A and p_{OHC} are obtained here by performing Gaussian elimination.

The displacement of the TM, w_{TM}, is specified as a function of that of the RL, w_{RL}, by using Eq. H.12. A transfer function denoted $T_{TM/RL}$ is created

$$(s^2 m_3 + k_3 + k_4) w_{TM} - k_3 w_{RL} = 0 \rightarrow \frac{w_{TM}}{w_{RL}} = T_{TM/RL} = \frac{k_3}{s^2 m_3 + k_3 + k_4}, \tag{H.13}$$

which is referred to as the *shear transfer function*[1]. The same operation is per-

[1]This term was first introduced by Allen [248] to relate bending displacements of TM and BM.

formed with Eq. H.11, which combined with Eq. H.13 allows us to express w_{RL} as a function of the displacements of the BM and TM, w_{BM} and w_{RL} respectively. This transfer function is denoted as $T_{RL/(BM+TM)}$

$$(s^2 m_2 + k_2 + k_3 + k_5) w_{RL} - k_3 w_{TM} - k_2 w_{BM} = 0,$$

$$\left[(s^2 m_2 + k_2 + k_3 + k_5) - k_3 T_{TM/RL}\right] w_{RL} = k_2 w_{BM}. \tag{H.14}$$

Hence

$$\frac{w_{RL}}{w_{BM} + w_{TM}} = k_2 T_{RL/(BM+TM)} = k_2 \left[(s^2 m_2 + k_2 + k_3 + k_5) - k_3 T_{TM/RL}\right]^{-1}. \tag{H.15}$$

A final relationship is introduced to relate w_{BM} with w_{RL}. This is denoted as $T_{BM/RL}$ and is obtained by setting Eq. H.10 equal to zero

$$(s^2 m_1 + k_1 + k_2) w_{BM} - k_2 w_{RL} = 0,$$

$$T_{BM/RL} = \frac{w_{BM}}{w_{RL}} = \frac{k_2}{s^2 m_1 + k_1 + k_2}. \tag{H.16}$$

By using these three relationships, the receptance of the three masses of the micromechanical model with respect to p_A and p_{OHC} can be obtained.

The receptance of the BM with respect to the acoustic driving pressure, $G_{BA} = w_{BM}/p_A$, is obtained by using Eq. H.14 and expressing w_{RL} as a function of w_{BM}

$$p_A = (s^2 m_1 + k_1 + k_2) w_{BM} - k_2 w_{RL} = (s^2 m_1 + k_1 + k_2 - k_2^2 T_{RL/(BM+TM)}) w_{BM},$$

$$G_{BA} = \frac{w_{BM}}{p_A} = \frac{1}{s^2 m_1 + k_1 + k_2(1 - k_2 T_{RL/(BM+TM)})}. \tag{H.17}$$

Similarly, the receptance of the BM when driven by the active pressure, $G_{BO} = w_{BM}/p_{OHC}$, is obtained by equalling Eq. H.10 and Eq. H.11 to p_{OHC} and expressing w_{RL} and w_{RL} displacements as a function of w_{BM}. When including the active pressure, w_{RL} is related to w_{BM} as

$$\left[(s^2 m_2 + k_2 + k_3 + k_5) - k_3 T_{TM/RL}\right] w_{RL} = p_{OHC} + k_2 w_{BM},$$

$$w_{RL} T^{-1}_{RL/(BM+TM)} = p_{OHC} + k_2 w_{BM},$$

$$w_{RL} = k_2 T_{RL/(BM+TM)} w_{BM} + T_{RL/(BM+TM)} p_{OHC}.$$

So, for Eq. H.10

$$(s^2m_1+k_1+k_2)w_{BM}-k_2\left(k_2T_{RL/(BM+TM)}w_{BM}+T_{RL/(BM+TM)}p_{OHC}\right)=-p_{OHC},$$

$$\left(s^2m_1+k_1+k_2-k_2^2T_{RL/(BM+TM)}\right)w_{BM}=p_{OHC}\left(-1+k_2T_{RL/(BM+TM)}\right),$$

$$G_{BO}=\frac{w_{BM}}{p_{OHC}}=\frac{-1+k_2T_{RL/(BM+TM)}}{s^2m_1+k_1+k_2-k_2^2T_{RL/(BM+TM)}}. \tag{H.18}$$

Other displacement of interest is that given by the relative displacements of RL and TM, $w_{RL}-w_{TM}$, which represents the displacement that the OHC stereocilia overcome, w_{ST}. First we introduce the stereocilia displacement for a given acoustic pressure input at the BM, $G_{SA}=w_{ST}/p_A$. For obtaining this displacement, it is needed first to express w_{RL} as a function of p_A.

$$w_{RL}\left((s^2m_1+k_1+k_2)\frac{T^{-1}_{RL/(BM+TM)}}{k_2}-k_2\right)=p_A,$$

$$\frac{w_{RL}}{p_A}=\left((s^2m_1+k_1+k_2)\frac{T^{-1}_{RL/(BM+TM)}}{k_2}-k_2\right)^{-1}. \tag{H.19}$$

By using Eq. H.13, w_{TM} is expressed as a function of w_{RL},

$$w_{ST}=w_{RL}-w_{TM}=w_{RL}-T_{TM/RL}w_{RL}=(1-T_{TM/RL})w_{RL}. \tag{H.20}$$

So that if Eq. H.19 and Eq. H.20 are combined

$$G_{SA}=\frac{w_{ST}}{p_A}=(1-T_{TM/RL})\left((s^2m_1+k_1+k_2)\frac{T^{-1}_{RL/(BM+TM)}}{k_2}-k_2\right)^{-1}. \tag{H.21}$$

The last receptance of interest is the one that relates the displacement of the stereocilia due to the active pressure, $G_{SO}=w_{ST}/p_{OHC}$. This quantity is obtained by expressing Eq. H.10 and H.11 as a function of w_{RL} and p_{OHC}.

For Eq. H.10

$$(s^2m_1+k_1+k_2)w_{BM}-k_2w_{RL}=-p_{OHC}\rightarrow w_{BM}=(k_2w_{RL}-p_{OHC})(s^2m_1+k_1+k_2)^{-1}. \tag{H.22}$$

Eq. H.22 is now combined onto Eq. H.11

$$w_{RL}T_{RL/(BM+TM)}=p_{OHC}+k_2w_{BM},$$

$$w_{RL}T_{RL/(BM+TM)}=p_{OHC}+k_2(k_2w_{RL}-p_{OHC})(s^2m_1+k_1+k_2)^{-1},$$

$$\left(T_{RL/(BM+TM)} + \frac{k_2^2}{s^2 m_1 + k_1 + k_2}\right) w_{RL} = \left(1 - \frac{k_2}{s^2 m_1 + k_1 + k_2}\right) p_{OHC},$$

which is reduced by using, $T_{BM/RL}$, Eq. H.16

$$\left(T_{RL/(BM+TM)} + k_2 T_{BM/RL}\right) w_{RL} = \left(1 - T_{BM/RL}\right) p_{OHC},$$

so that

$$\frac{w_{RL}}{p_{OHC}} = \frac{1 - T_{BM/RL}}{T_{RL/(BM+TM)} + k_2 T_{BM/RL}}, \tag{H.23}$$

which if combined with Eq. H.20 gives G_{SO}

$$G_{SO} = \frac{w_{ST}}{p_{OHC}} = \left(1 - T_{TM/RL}\right) \frac{1 - T_{BM/RL}}{T_{RL/(BM+TM)} + k_2 T_{BM/RL}}. \tag{H.24}$$

A detailed study of the frequency response of each of these receptances is presented in Section H.4.3, using the parameters described in Section H.3.

H.3 Distribution of parameters along the cochlea

In the original model of Ramamoorhty *et al.* [190] the input quantity to the model was a force. In the lumped parameter model of a guinea pig cochlea of Simón Gálvez and Elliott [191], the parameters given by Ramamoorhty *et al.* [190] are normalised by the width of the BM so that the model is made one dimensional and the input quantity is a pressure instead of a force. Apart from this normalisation, very few changes are made to the original parameters of Ramamoorhty *et al.*.

To adapt the original guinea pig model to reproduce human responses and assure that it was still physically realisable, however, several changes were necessary. The first step was to select the desired frequency range of the passive BM. The mass of the BM is obtained from the original value of $m_{BM_{3D}} = 0.05\ kgm^{-2}$ plus the added value of $\rho_f T$. Based on the total BM mass value, a stiffness for the BM is selected, assuming that the BM resonant frequency follows a certain spatial dependence. In the case of the model, it is assumed that the resonant frequency of the element at the base is $f_{BM}(0)$=28 kHz. A spatial dependence is assumed so that

$$f_{BM}(x) = f_{BM}(0) e^{-x/l}, \tag{H.25}$$

where $l = 0.07$ is the *natural frequency length scale*, as given by Elliott *et al.* [257]. The stiffness for the BM, k_{BM} is obtained as

$$k_1(x) = k_{BM}(x) = m_{BM}(x)(2\pi f_{BM}(x))^2. \tag{H.26}$$

A coefficient for normalisation is obtained, based on the ratio between the "human" BM stiffness given by Eq. H.26 and the original BM stiffness of a guinea pig

$$\Delta_{k_{BM}}(x) = \frac{k_{BM}(x)}{k_{BM_{G.P.}}(x)}, \tag{H.27}$$

which is used to normalise the rest of stiffness. The mass of the m_{RL} and the m_{TM} are assumed to be 0.02 kgm^{-3} at the base. The resonance frequency of the TM is given by

$$f_{TM}(x) = \frac{1}{2\pi}\frac{2k_{HB}(x) + k_{TM} + k_{RL}}{m_{TM}} = \frac{1}{2\pi}\frac{2k_3(x) + k_4 + k_5}{m_{TM}} \tag{H.28}$$

The natural frequency of the TM determines the CF of the active coupled response. As it is desired that a half octave band shift occurs between passive and active responses, all the stiffness except k_{BM} are normalised by a factor $\Delta_{f_{TM}}$, which is given by

$$\Delta_{f_{TM}}(x) = m_{TM}(x)(1.2 \cdot 2\pi f_{BM})^2. \tag{H.29}$$

The use of these *human, normalised* parameters gives the distribution of natural frequencies shown in Fig. H.4, in comparison with the natural frequencies for a guinea pig cochlea. The variation of the TM and RL masses with length is also a parameter taken into account for the fine adjustment of the model. The value of the masses increases with the length of the cochlea, as in the original model of Ramamoorhty *et al.* [190]. In the human lumped parameter model these masses increase even more with the length of the cochlea in order to assure stability. This is in concordance with the results shown by Elliott and Ni [258], in where the authors suggest that the effective mass of the OC due to fluid loading grows along the cochlea.

The value of each of the normalised parameters, in comparison with those of the guinea pig model [191] are shown in Table H.1.

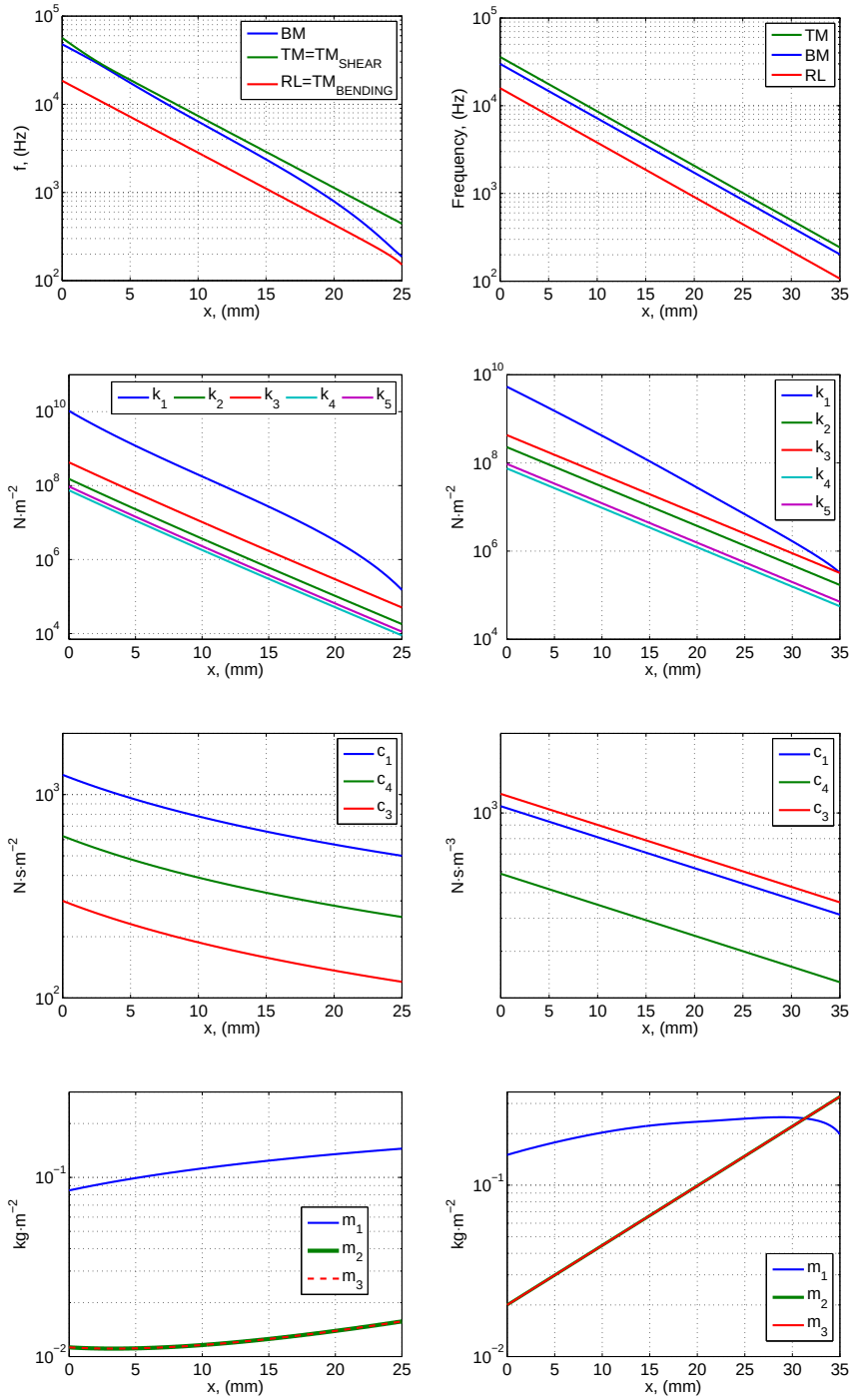


Figure H.4: From top to bottom: Distribution of resonant frequencies, stiffnesses, dampers and masses for the guinea pig cochlea (left hand side plot) and for the human cochlea (right hand side plot).

Table H.1: Micromechanical parameters for the human model of the cochlea, which have been converted from the parameters used for a guinea pig cochlea by Simón and Elliott [191]. The magnitude $b_{BM}^{(x)}$ stands for the width of the BM of the guinea pig as used by Ramamoorthy *et al.* [190].

Element	Guinea pig value [191]	Units	Conversion	Human value	Source
k_1	$8.40 \cdot 10^5 (h/h_0)^3 (b_0/b)^4$	Nm^{-3}	$m_1 w_{BM}^2$	$5.20 \cdot 10^9 e^{-253x}$	Adjusted manually
k_2	$2 \cdot 6.08 \cdot 10^3 e^{-325x}/b_{BM}^{(x)}$	Nm^{-3}	$k_2/(\Delta_{k_{BM}} \Delta_{f_{TM}})$	$2.28 \cdot 10^8 e^{-206x}$	Corrections from Li [259] ,K_{OHC}
k_3	$3.42{\cdot}10^9 e^{-325x}/b_{BM}^{(x)}$	Nm^{-3}	$k_3/(\Delta_{k_{BM}} \Delta_{f_{TM}})$	$4.27{\cdot}10^8 e^{-206x}$	Ref. [190] ,K_{HB}
k_4	$6.00{\cdot}10^9 e^{-220x}/b_{BM}^{(x)}$	Nm^{-3}	$k_4/(\Delta_{k_{BM}} \Delta_{f_{TM}})$	$7.49{\cdot}10^7 e^{-206x}$	Ref. [190], K_{TMs}
k_5	$7.60{\cdot}10^9 e^{-325x}/b_{BM}^{(x)}$	Nm^{-3}	$k_5/(\Delta_{k_{BM}} \Delta_{f_{TM}})$	$9.48{\cdot}10^7 e^{-206x}$	Ref.[190], K_{RL}
m_1	$0.08_{\text{(base) to}}\ 0.14\ _{\text{(apex)}}$	kgm^{-2}	$0.05 + m_F$	$0.15_{\text{(base) to}}\ 0.20\ _{\text{(apex)}}$	Ref. [257, 260], M_{BM}
m_2	$0.015 e^{50x}/b_{BM}^{(x)}$	kgm^{-2}		$0.02 e^{80x}$	Adjusted manually
m_3	$g m_2$	kgm^{-2}		$0.02 e^{80x}$	From [190], assumed $g = 1$
c_1	$0.10{\cdot}10^3/b_{BM}^{(x)}$	Nsm^{-3}		$1061{\cdot}10^3 e^{-27x}$	Corrections from Li [259], C_{BM}
c_2	0.00	Nsm^{-3}		0.00	
c_3	$0.024{\cdot}10^3/b_{BM}^{(x)}$	Nsm^{-3}		$1181{\cdot}10^3 e^{-27x}$	Ref. [190], C_{TMs}
c_4	$0.05{\cdot}10^3/b_{BM}^{(x)}$	Nsm^{-3}		$589{\cdot}10^3 e^{-27x}$	Ref. [190], C_{TMs}
c_5	0.00	Nsm^{-3}		0.00	

H.4 Active model of the organ of Corti

H.4.1 Formulation

The active response of the OC is introduced in the model of Fig. H.3b by including a feedback loop whereby the control pressure, p_{OHC} is related to the shear displacement of the stereocilia, w_{ST}. The passive response of the OC is indicated as the responses inside the box in Fig. H.5, in terms of the displacements of the BM, w_{BM}, and the shear displacement of the stereocilia between the RL and the TM, w_{ST}, in response to either the acoustic pressure along on the BM, p_A, and the pressure due to the OHC, p_{OHC}. Hence, p_{OHC} is then assumed to be driven by w_{ST} via the feedback response H.

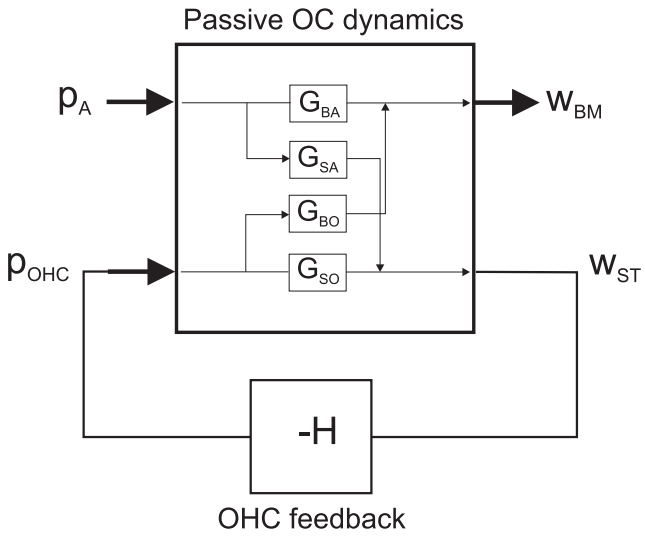


Figure H.5: Block diagram of the active amplification that takes place in the organ of Corti (OC) model of Fig. H.3b.

A set of equations for w_{BM} and w_{ST} can be obtained according to the block diagrams of Fig. H.5, for the case of the open loop

$$w_{BM} = G_{BA}p_A + G_{BO}p_{OHC}, \tag{H.30}$$

$$w_{ST} = G_{SA}p_A + G_{SO}p_{OHC}. \tag{H.31}$$

The closed loop response can be derivated from these equations and the block diagram of Fig. H.5. Given that

$$p_{OHC} = -Hw_{ST}, \tag{H.32}$$

hence

$$p_{OHC} = -H\left(G_{SA}p_A + G_{SO}p_{OHC}\right) \rightarrow p_{OHC} = \frac{-HG_{SA}}{1+G_{SO}H}p_A. \tag{H.33}$$

So that the closed loop displacement of the BM is given as

$$w_{BM} = \left(G_{BA} - G_{BO}\frac{HG_{SA}}{1+G_{SO}H}\right)p_A, \tag{H.34}$$

and hence the closed loop mobility as

$$Y_{BM} = \frac{sw_{BM}}{p_A} = s\left(G_{BA} - G_{BO}\frac{HG_{SA}}{1+G_{SO}H}\right), \tag{H.35}$$

where H represents a suitable controller, which can be written such as

$$H(j\omega) = \frac{\gamma g_H}{1+\frac{j\omega}{\omega_{OHC}}}. \tag{H.36}$$

The derivation of such controller is described below.

H.4.2 Simplified OHC model

The following assumptions are performed here with respect to the OHC model of Ramamoorhty *et al.* [190]:

- A single row of OHC is used instead of three rows as in the original model.
- No longitudinal coupling on the BM. The current flowing on a single OHC is due to the current of the OHC itself and not by potential differences on any of the scales.

By modifying Eq. 21 of [190], p_{OHC} can be defined as

$$p_{OHC} = k_2(w_{RL} - w_{BM}) + \gamma\varepsilon_3\phi_{OHC}, \tag{H.37}$$

where γ allows us to select the level of active amplification, ε_3 represents the electromechanical coupling coefficient and ϕ_{OHC} represents the voltage across the OHC. As no longitudinal coupling is modelled here, this last quantity is defined as the product of the current through the OHC, I_{OHC}, by the OHC electrical impedance, Z_m,

$$\phi_{OHC} = Z_m I_{OHC}, \tag{H.38}$$

where $Z_m = (1/R_m + j\omega C_m)^{-1}$. I_{OHC} is proportional to the steady voltage at the OHCs, V_0, to the conductance function, G_a, which varies with the position along

the cochlea, and to the stereocilia displacement, $w_{ST} = w_{TM} - w_{RL}$

$$I_{OHC} = V_0 G_a w_{ST}, \tag{H.39}$$

where G_a is defined by Ramamoorthy *et al.* [190], based on numerical experiments, as

$$G_a = G_a(0)\left(-10^5 x^3 + 5.910^3 x^2 - 109x + 1\right) e^{-150x}, \tag{H.40}$$

p_{OHC} is then defined as

$$p_{OHC} = k_2(w_{BM} - w_{RL}) + \gamma \varepsilon_3 Z_m V_0 G_a w_{ST}, \tag{H.41}$$

which can also be expressed as

$$p_{OHC} = k_2(w_{BM} - w_{RL}) + H w_{ST}, \tag{H.42}$$

where H is defined as

$$H = \frac{\gamma g' \varepsilon_3 Z_m V_0 G_a}{1 + j\omega R_m C_m} = \frac{\gamma g_H}{1 + \frac{j\omega}{\omega_{OHC}}}, \tag{H.43}$$

where $\omega_{OHC} = (R_m C_m)^{-1}$, and g_H is defined as the MET overall gain

$$g_H = g' \varepsilon_3 Z_m V_0 G_a. \tag{H.44}$$

The value of the parameters used in the formulation, obtained from Ramammorthy *et al.* [190] are presented in Table. H.2.

Table H.2: Value for the electrical parameters used in the OHC compensator.

Element	**Value**	**Units**
ε_3	$(-8 \cdot 10^{-6} \cdot 10^{-5} x)$	N/m/mV
$1/R_m$	$5100_{\text{(base) to}}\ 360_{\text{ (apex)}}$	μS/m
C_m	$1800_{\text{ (base) to}}\ 4200_{\text{ (apex)}}$	nF/m
$G_a(0)$	$5.447{\cdot}10^6$	S/m^2
V_0	$150 - 1000x$	mV
$\omega_{OHC}/2\pi$	$450_{\text{(base) to}}\ 13.7_{\text{ (apex)}}$	Hz

Ramamoorhty *et al.* [190] defined the MET sensitivity as the parameter that controls the amount of gain of the CA in their model, instead of the gain of the feedback amplifier, γ, as Neely and Kim [261]. The model presented here follows the original description of Ramamoorhty *et al.* [190], but the parameter γ has been left for convenience. The model hence depends non-linearly and strongly

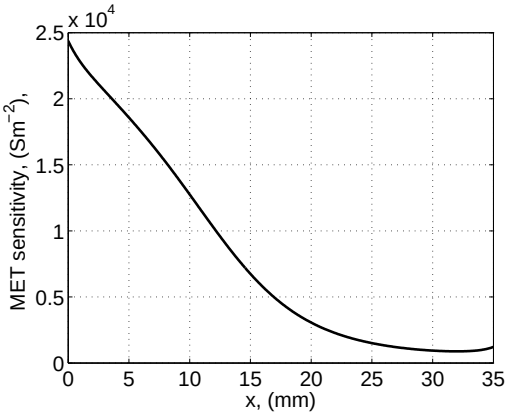


Figure H.6: Assumed variation of mechano-electrical transduction sensitivity of the OHC along the length of the cochlea.

on the MET sensitivity. Indeed, if the MET gain is increased beyond a certain limit, or γ is increased more than 1, the model becomes unstable. As previously observed by Ramamoorhty *et al.* [190], the MET sensitivity varies with the length of the cochlea so that the global response is stable. The required gain to apply to H in order to obtain an active amplification, g', is given by

$$g' = \frac{1}{\Re\{H(fp)G_{SO}(fp)\}}, \tag{H.45}$$

where fp represents the point of the response with -0.5 cycles phase, which is placed close to the (-1,0) point, and the responses are those of the coupled cochlea. In order to reproduce the responses present in this thesis g has been approximated by the polynomial

$$\begin{aligned} g'(x) = {} & 2.81035128320330 \cdot 10^{25}x^{12} - 6.36974893956170 \cdot 10^{24}x^{11} \\ & + 6.39271799459847 \cdot 10^{23}x^{10} - 3.73237386749530 \\ & \cdot 10^{22}x^{9} + 1.40060580245442 \cdot 10^{21}x^{8} - 3.52472603132536 \\ & \cdot 10^{19}x^{7} + 6.03796978510629 \cdot 10^{17}x^{6} - 7.04710877272383 \\ & \cdot 10^{15}x^{5} + 55931323311456.2x^{4} - 299441843783.816x^{3} \\ & + 989220790.214985x^{2} - 2606452.95733168x + 29178.7935867540 \end{aligned} \tag{H.46}$$

which is plotted against x in Fig. H.6. It can be observed how for lower distances from the base the gain has to be higher, whilst this is reduced greatly when x is greater than 25 mm. The shape of the function is similar to that used by Ramamoorhty *et al* [190].

H.4.3 Active response

According to the block diagram of Fig. H.5, the active response of the cochlea depends of the individual quantities defined in Section H.2, these are: The recep-

tance of the BM when this is excited by the acoustic pressure, G_{BA}, Eq. H.17; the receptance of the BM when excited by the force due to the OHC electromotility, G_{BO}, Eq. H.18; the displacement of the stereocilia of the OHC, when excited by the acoustic pressure that drives the BM, G_{SA}, Eq. H.21; and when the stereocilia is displaced by the contraction of the OHC, G_{SO}, Eq. H.24. These values are shown here for the position of 3.75 mm from the base.

Fig. H.7 shows the individual frequency responses due to the excitation of individual modes, calculated with the damping shown in Table H.1 and also with this reduced by a 90 % in order to better observe the three modes. The responses between the fluid pressure excitation and the BM response, G_{BA}, clearly show the BM mode at around 18 kHz, and thus for the OHC excitation to the stereocilia response, G_{SO}, has a peak at above 20 kHz, corresponding to the TM shear mode. The mode corresponding to the bending action of the RL and the TM can be observed to be around 9 kHz. Also for reference, the individual responses of the RL and TM are shown, w_{RL} and w_{TM}, when excited by the OHC, where G_{SO} is the difference between these responses. The results are in concordance with those obtained using a finite elements model of the organ of Corti, in where three main differentiated resonances corresponding to each one of the degrees of freedom are observed.

Fig. H.8 shows the response of an isolated element calculated using such responses for the micromechanical element of 3.75 mm. It can be seen how the response is slightly reduced at frequencies between the mode of the RL and the mode of the BM. At the resonance of the BM, a big increase is found in the response and at the shear mode of the TM a large notch is found in the modulus of the active response, as the phase becomes -0.5 cycles at around 22 kHz. This result is in concordance with results obtained with a finite element model of the organ of Corti [262]. The Nyquist diagram and open loop response, HG_{SO} are also shown. The closed loop response shows how around the BM resonance this adds positively to Y_{BM}, as both quantities have a similar phase. The notch in the mobility of the BM is found at the frequency in which the phase of the open loop is 0 cycles.

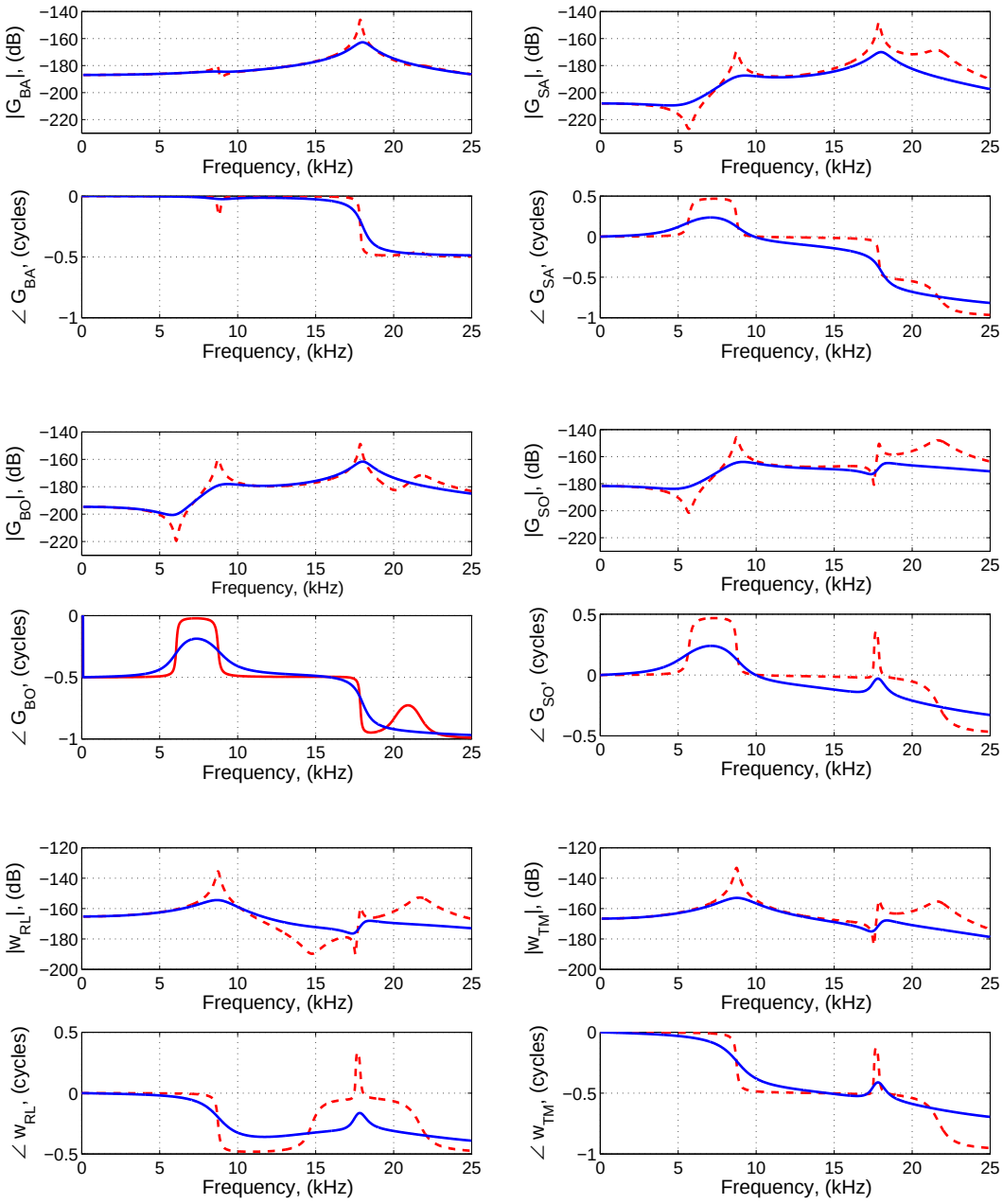
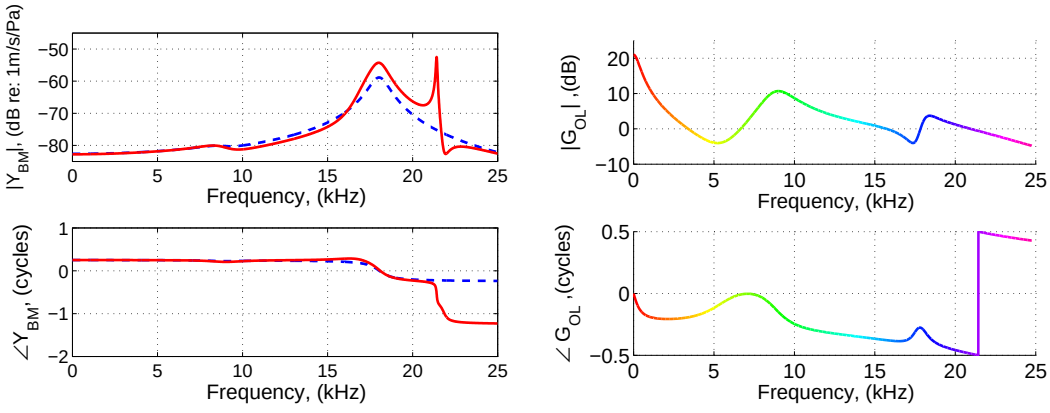
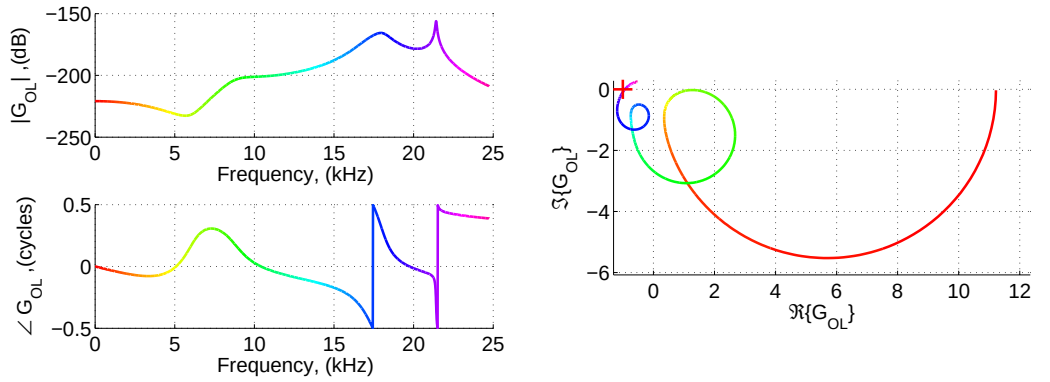


Figure H.7: Dynamics of the micromechanical model at 3.75 mm normalized to 1m·Pa^{-1}. The graphs of w_{RL} and w_{TM} represent the case when these are excited by p_{OHC}. The dashed results have the dampers reduced by a 90% to more clearly see the three modal frequencies.



(b) HG_{SO}



(c) $G_{BO}\frac{HG_{SA}}{1+G_{SO}H}$ **(d)** HG_{SO}

Figure H.8: The upper left plot shows the frequency response of the mobility for a passive (dashed blue), $\gamma = 0$, and fully active (red), $\gamma = 1$, micromechanical element. The upper right plots shows the open loop response, $G_{OL} = HG_{SO}$. The lower plots show the closed loop response, $G_{CL} = G_{BO}\frac{HG_{SA}}{1+G_{SO}H}$, and the Nyquist diagram of the open loop response, G_{OL}. All the responses correspond to the micromechanical element of 3.75 mm.

H.5 Coupled model of the cochlea

H.5.1 Elemental formulation

The cochlea can be modelled as a set of N micromechanical elements, being each of them tuned to a different resonant frequency depending on its position along the cochlea, and coupled through the inertia of the fluid in the chambers. For the case of the simulations presented, these have been calculated using $N = 512$ elements. The elements in the cochlea are assumed to be equally spaced, as shown in Fig. H.9.

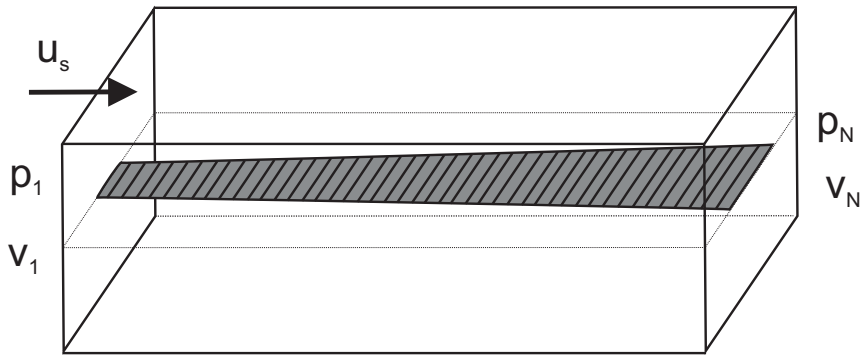


Figure H.9: *Long wave* macromechanical model of the cochlea. Each slice represents one from a total of N micromechanical elements as the one of Fig. H.3b.

We can define vectors of complex pressures and velocities at a single frequency at the position of each micromechanical element, $\mathbf{p}$ and $\mathbf{v}$

$$\mathbf{p} = [p(1), p(2), \ldots, p(N)]^T, \quad \mathbf{v} = [v(1), v(2), \ldots, v(N)]^T. \tag{H.47}$$

The BM, however, is assumed only to extend from element 2 to element N-1. Element 1 accounts for the stapes velocity, u_s in Fig. H.9, whilst the final element, N, accounts for the behaviour of the helicotrema. If the stapes velocity is set to zero, the vector of pressures due to the vector of BM velocities can be written as

$$\mathbf{p} = \mathbf{Z}_{FC}\mathbf{v}, \tag{H.48}$$

where $\mathbf{Z}_{FC}$ is the matrix due to fluid coupling. This matrix can be associated with two components, the one which accounts for the 1D propagation of the fluid and the near field component, which is associated with the 3D propagation of the fluid near the point of excitation, so that $\mathbf{Z}_{FC} = \mathbf{Z}_F + \mathbf{Z}_N$. In the same way, the vector of BM velocities can be written as

$$\mathbf{v} = -\mathbf{Y}_{BM}\mathbf{p}, \tag{H.49}$$

where $\mathbf{Y}_{BM}$ is the diagonal mobility matrix, as it is assumed to react only locally. The diagonal elements represent the admittance of each micromechanical element at a given frequency, and extend only from 2 to $N-1$. It is formed as

$$\mathbf{Y}_{BM} = j\frac{\omega\rho}{H}\begin{bmatrix} 0 & & & & \\ & Y_{cp}(2) & & & \\ & & \ddots & & \\ & & & Y_{cp}(N-1) & \\ & & & & 0 \end{bmatrix}. \tag{H.50}$$

The total vector of velocities from cam be obtained using linear superposition, so that

$$\mathbf{v} = \mathbf{v}_S - \mathbf{Y}_{BM}\mathbf{p}, \tag{H.51}$$

where $\mathbf{v}_S$ is defined as the sources vector, being its first element equal to the stapes velocity, and the rest to zero

$$\mathbf{v}_s = \begin{bmatrix} u_S T & 0 & \cdots & 0 \end{bmatrix}^T. \tag{H.52}$$

By combining Equations H.48 and H.51 two expressions that account for the coupled response of the fluid and BM dynamics are obtained.[2]

$$\mathbf{v} = \left[\mathbf{I} + \mathbf{Y}_{BM}\mathbf{Z}_{FC}\right]^{-1}\mathbf{v}_s, \tag{H.53}$$

$$\mathbf{p} = \left[\mathbf{Z}_{FC}^{-1} + \mathbf{Y}_{BM}\right]^{-1}\mathbf{v}_s. \tag{H.54}$$

H.5.2 1D fluid coupling

The far field component of the pressure difference or 1D fluid coupling, can be obtained by using a wavenumber analysis [263, 260]. Alternatively, the pressure distribution can be obtained analytically from the solution of the differential equation for the far field coupling [260], as

$$p_l(n) = \frac{j16\omega\rho\Delta^2 v_0}{\pi^2}\sqrt{B(n_0)B(n)}\sum_{n'=n_0}^{N}\frac{1}{S_F(n')}, \quad 0 < n < n_0 - 1, \tag{H.55}$$

$$p_l(n) = \frac{j16\omega\rho\Delta^2 v_0}{\pi^2}\sqrt{B(n_0)B(n)}\sum_{n'=n}^{N}\frac{1}{S_F(n')}, \quad n_0 < n < N, \tag{H.56}$$

[2] A complete description can be found in [257].

where $S_F(x)$ is the effective chamber area, given by

$$S_F(x) = \frac{S_1(x)S_2(x)}{S_1(x) + S_2(x)}, \tag{H.57}$$

being $S_1(x)$ and $S_1(x)$ the upper (scala media and scala vestibuli) and lower (scala tympani) chambers, and $n' = |n - n_o|$, where n_0 is the element being excited. For a one dimensional propagation of the fluid, the columns of $\mathbf{Z}_{FC}$ can be obtained by normalizing Eqs. H.55 and Eq. H.56 by the input velocity, v_0.

H.5.3 3D fluid coupling

The formulation of the pressure difference due to the fluid can be extended so that the effects of the near field, or 3D propagation, taking into account higher order fluid modes, are also considered. In this case, the pressure difference is due to far and near field components, so that

$$p_{FC} = p_L + p_N, \tag{H.58}$$

where the subscript L stands for long wave and the subscript N stands for near field. Consequently, the fluid coupling matrix can be formed as

$$\mathbf{Z}_{FC} = \mathbf{Z}_L + \mathbf{Z}_N, \tag{H.59}$$

where $\mathbf{Z}_L$ is given by normalizing Equations H.55 and H.56 by the input velocity.

For the case of cochlear box of uniform dimensions, the effective thickness of the BM due to fluid loading can be calculated by using [260]

$$T = \frac{8BH}{3\pi^3 W} + \sum_{n=1}^{\infty} \coth(n\pi H/W) \left[\frac{\cos(n\pi C/W) + \cos(n\pi(C+B)/W)}{1 - n^2 b^2/W^2} \right]^2, \tag{H.60}$$

where B is the width of the basilar membrane and W and H are the physical width and height of one cochlear chamber. If the dimensions of the cochlear chambers vary along the length of the cochlea, as in a real case, the near field fluid loading can be calculated by inverse Fourier transformation of the wavenumber of the fluid coupling [260]. At a given point along the cochlea, the near field pressure is given by

$$p_N = j\omega_0 \rho_0 T v_0. \tag{H.61}$$

The near field pressure distribution along the cochlea is obtained using an approximation to the averaged near field pressure due to a single element of the

BM, obtained from [260], so that

$$p_N(n) = j\omega_0\rho_0 \left(Q_1 e^{-n'\Delta/l_1} + Q_2 e^{-n'\Delta/l_2}\right) v_0, \tag{H.62}$$

where $Q_1 =$ and Q_2 are a function of T, l_1 equal to W/3.47 and l_2 to W/12.8. The columns of the 3D fluid coupling matrix, $\mathbf{Z}_N$, are obtained by normalizing both the far field and the near field pressures at each position by the input velocity.

H.5.4 Comparative of fluid coupling configurations

Three different fluid coupling configurations can be used to calculate the coupled response of the cochlea. This depends on the magnitude of the higher order modes used to account for the 3D propagation, which is used as an extra tuning parameter for the cochlear responses, as already performed by [190]. These configurations are:

1D Fluid coupling

In this case the pressure due to the 1D propagation of the fluid is obtained from Equations H.55 and H.56. At the position of excitation, a delta symbolises the added mass due to the loading of the 3D propagation of the fluid, as can be observed in Fig. H.10a, so that

$$p_N(x) = M_{FC}\delta(x - x_0). \tag{H.63}$$

The mass of the BM is obtained by the value given by Elliott et al. [257], m_{BM}=0.05 Kgm^{-2} plus the value due to the effective thickness of the BM, which symbolises the delta function $\delta(x - x_0)$. This allows the total "effective" mass of the BM, including near field fluid loading, to be taken into account, allowing the mechanics of the organ of Corti to be studied. The pressure distribution at an excitation point can be observed in Fig. H.10a, and the coupled response obtained with this fluid distribution for a shoe box model of the cochlea is shown in Fig. H.12.

3D Fluid coupling

In this case the 3D fluid propagation is given by Eq. H.60 for a cochlea of uniform fluid chambers or by inverse Fourier transformation of the fluid wavenumber. The fluid coupling matrix is formed as in Eq. H.59. The BM mass has the value given by Elliott et al. [257], m_{BM}=0.05 Kgm^{-2} and the extra mass due to the fluid near field is implicit included in the fluid coupling. The pressure distribution at an excitation point using this configuration is shown in Fig. H.10b, and the results for the velocity distribution

of the coupled cochlea for a shoe box model are shown in Fig. H.12.

Diminished 3D coupling with extra mass

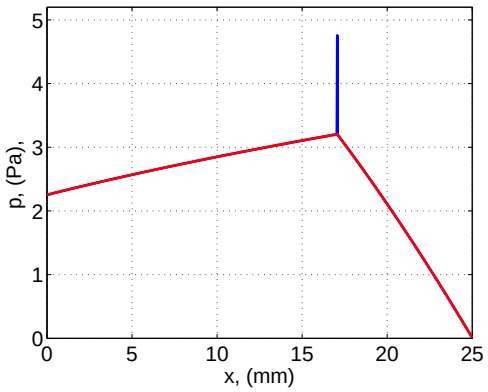
This configuration is a combination of the 1D and 3D cases. The pressure distribution of this fluid coupling configuration is similar to that of the 1D case, in where a delta function of pressure loading at the excitation position simulates the added BM, as shown in Fig. H.10c. In this case the fluid coupling matrix is formed as

$$\mathbf{Z}_{FC} = \mathbf{Z}_L + \Delta_T \mathbf{Z}_N, \tag{H.64}$$

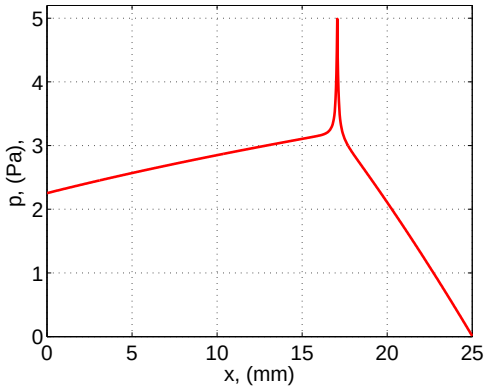
where $\mathbf{Z}_N$ is obtained from Eq. H.60 for a straight cochlear box or by inverse Fourier transformation of the fluid wavenumber for a cochlea of varying chambers [260]. The BM mass is obtained using the value given by Elliott et al. [257]

$$m_{BM} = 0.05 + \Delta_M \rho_f T. \tag{H.65}$$

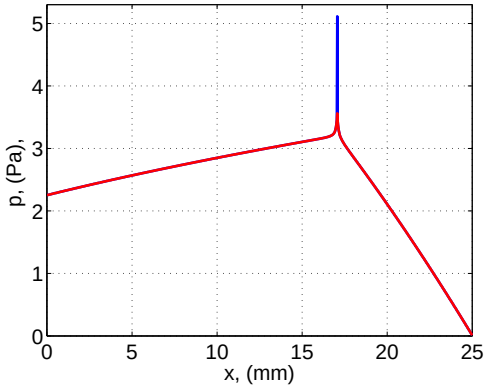
The ratio $\frac{\Delta_M}{\Delta_T}$ is used as a tuning parameter, allowing to select the amount of higher order fluid modes to be considered when solving the coupled response. These parameters can be varied from Δ_M=1, $\Delta_T = 0$, to represent a 1D fluid coupling configuration, until Δ_M=0, $\Delta_T = 1$ to give a 3D fluid coupling configuration. A detailed analysis of the effect the number of higher order modes has in the coupled response is shown between Figures H.12 and H.14.



(a) 1D fluid coupling



(b) 3D fluid coupling



(c) Diminished 3D fluid coupling

Figure H.10: Total pressure distribution for a guinea pig cochlea, including increased effective BM thickness, for the different fluid coupling configurations, with a BM excitation at 17 mm. The delta function components in blue symbolise the added mass due to the effective thickness of the BM, whilst the delta function components in red are due to fluid coupling. Results reproduced from [191].

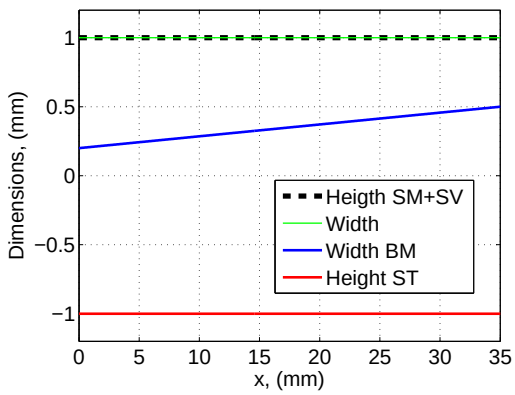
H.6 Initial responses of the model

A first set of responses have been obtained with a locally reacting model. These responses have been studied and have been used to carefully tune the model. The model uses the parameters shown in Table H.1, where it is also shown the derivation from the original values given by Ramamoorhty *et al.*[190] for a guinea pig.

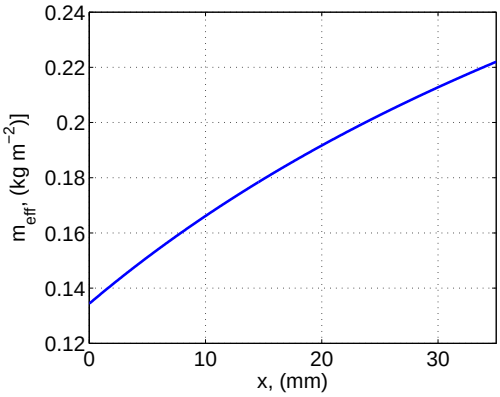
H.6.1 Responses based in a shoe box cochlear model

The most simple approach to simulate a cochlea is to use a rectangular box model of the cochlear chambers, commonly denominated "shoe box" model, as shown by de Boer [188]. In a first instance, this is the approach that has been used to model the cochlea and assess the effect that the different fluid configurations have on the coupled response.

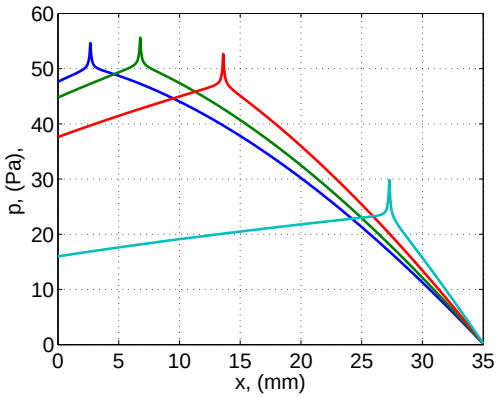
The dimensions of the cochlear chambers are shown in Fig. H.11a. The height of both cochlear chambers and the width are assumed to be equal to 1 mm, and the BM is assumed to increase from 0.2 mm at the base to 0.5 mm at the apex. in where the long field coupling is calculated using Eqs. H.55 and H.56 and the higher order modes have been calculated using Eq. H.60. The effective added mass on the BM due to the action of the higher order modes of the fluid is shown in Fig. H.11b. The distribution of pressure difference is shown at some positions along the cochlea for a straight box model in Fig. H.11c, where the combination of the far- and near-field components can be distinguished.



(a) Dimensions of cochlear chambers and BM

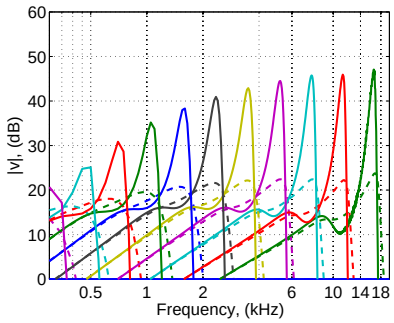


(b) $\rho_f T$

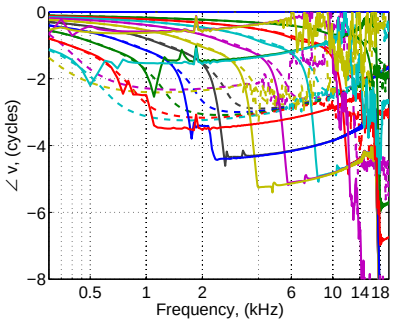


(c) Pressure difference in the chambers

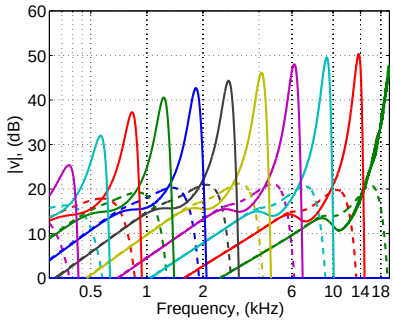
Figure H.11: Shoe box cochlear chamber and BM dimensions and corresponding effective mass due to the near field fluid. The lower image shows the calculated mean pressure due to far- and near-field components in the fluid matrix when only a single element of the discrete BM at x= 2.7. 7, 14 or 27 mm is driven sinusoidally with a velocity of 10 mms^{-}1 at 1 kHz.



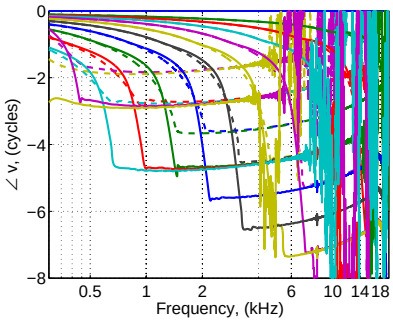
(a) 1D



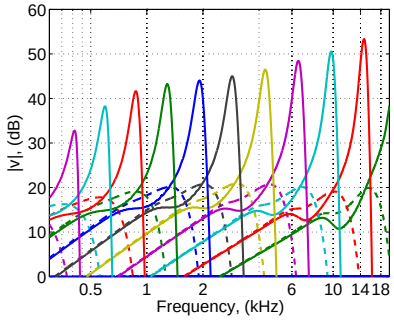
(b) 1D



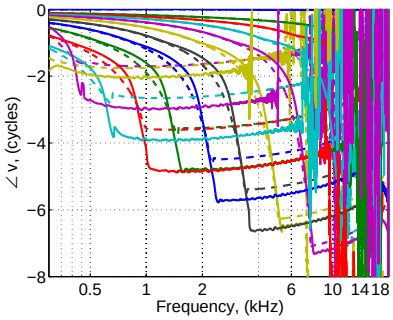
(c) $\frac{\Delta_M}{\Delta_T}$=0.5



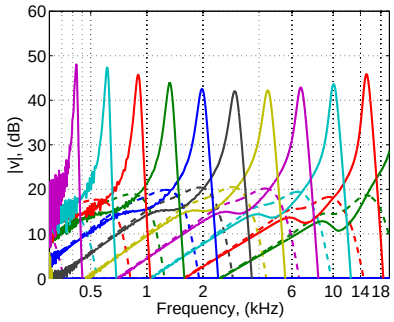
(d) $\frac{\Delta_M}{\Delta_T}$=0.5



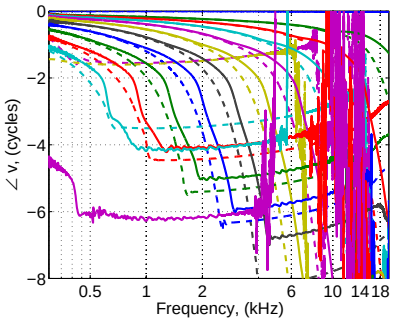
(e) $\frac{\Delta_M}{\Delta_T}$=0.42



(f) $\frac{\Delta_M}{\Delta_T}$=0.42

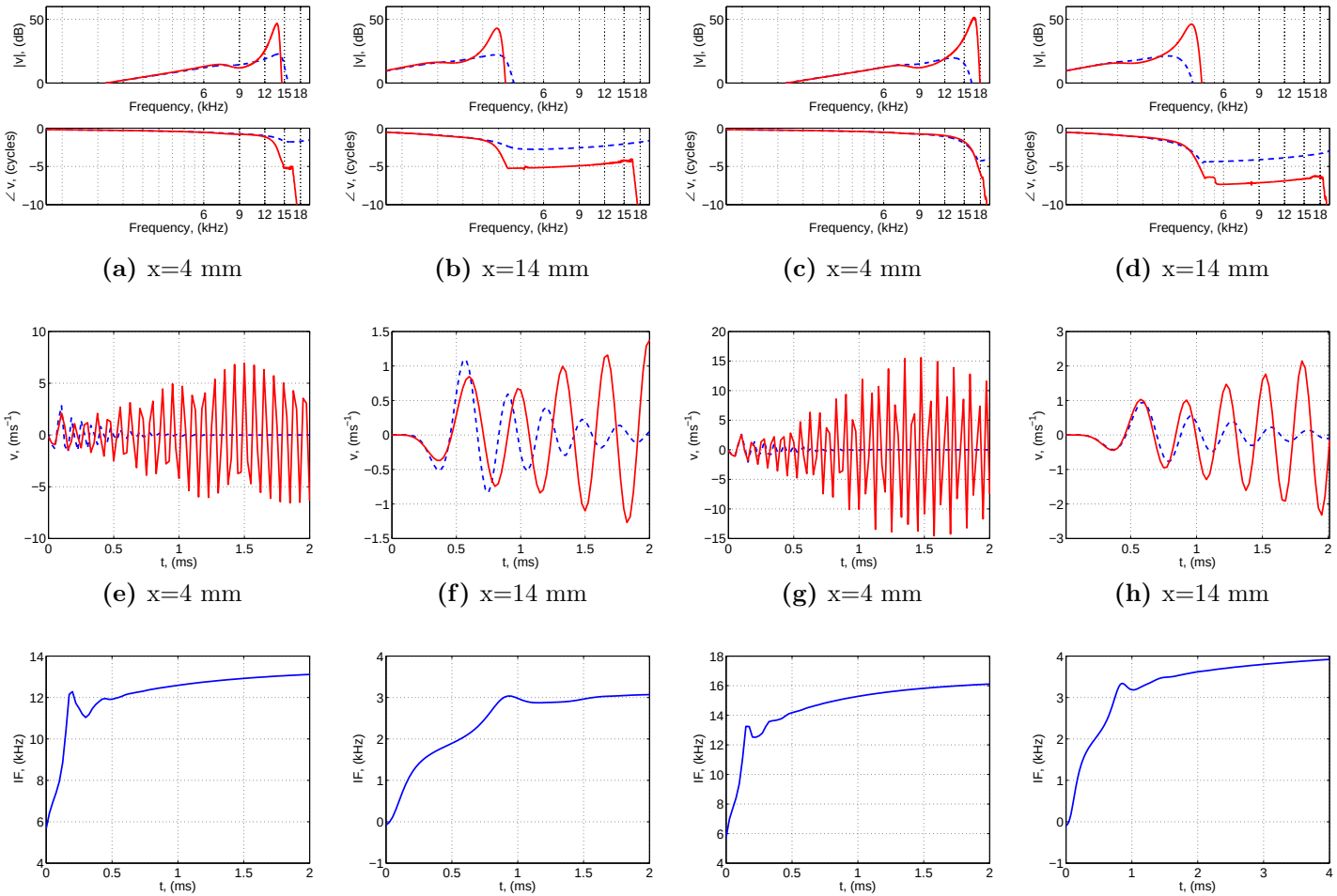


(g) 3D



(h) 3D

Figure H.12: Response of the coupled passive (dashed lines) and active (solid lines) shoe box model of the cochlea using different fluid coupling configurations. Responses shown between x=2.1 mm until x=34.9 mm, in increments of 2.1 mm.

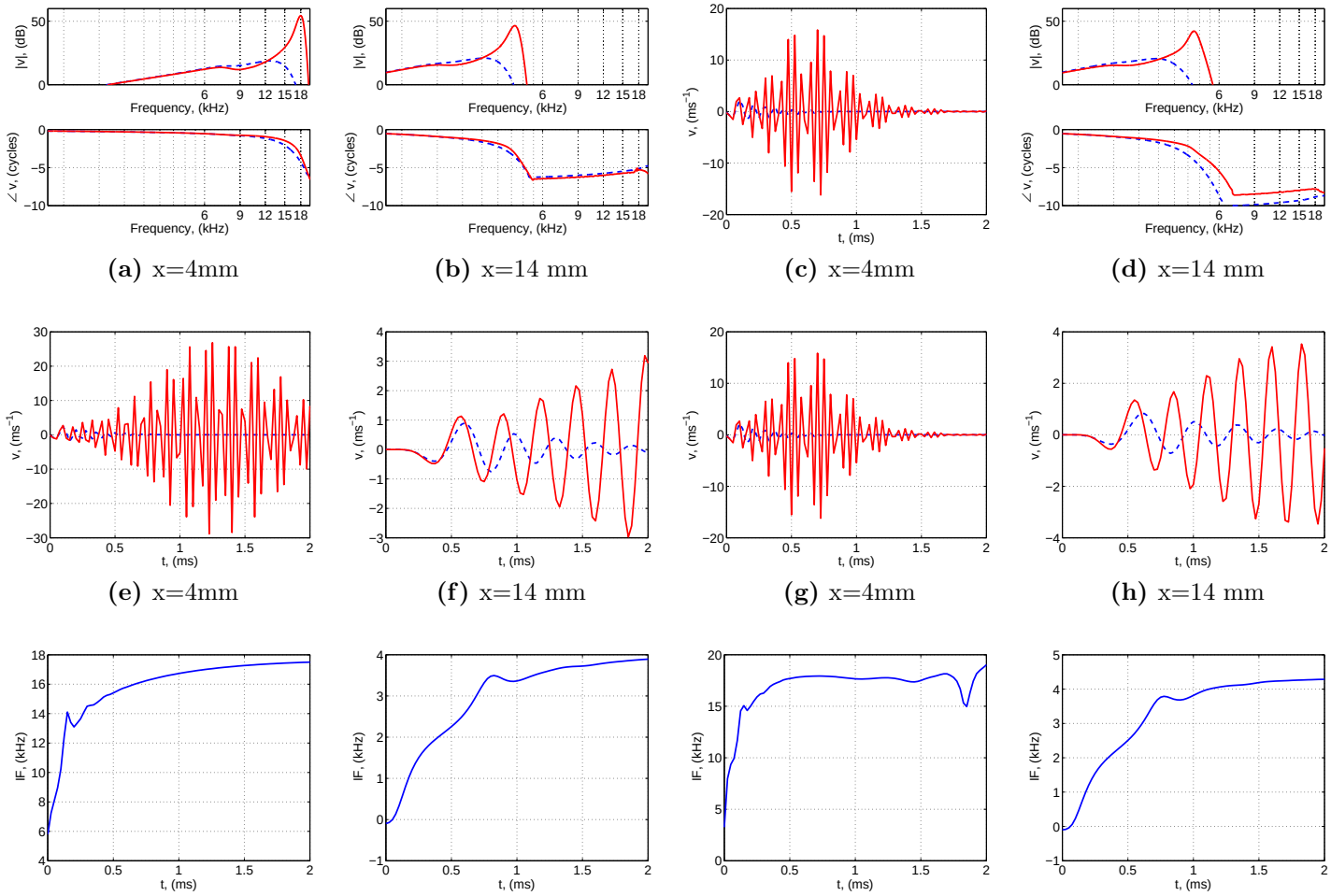


(a) x=4 mm **(b)** x=14 mm **(c)** x=4 mm **(d)** x=14 mm

(e) x=4 mm **(f)** x=14 mm **(g)** x=4 mm **(h)** x=14 mm

(i) x=4 mm **(j)** x=14 mm **(k)** x=4 mm **(l)** x=14 mm

Figure H.13: Details of the passive (dashed blue) and active (solid red) coupled response of a shoe box cochlea using a 1D formulation of the fluid coupling (left and middle left plots) and a diminished fluid coupling configuration where $\frac{\Delta_M}{\Delta_T} = 0.5$ (middle right and right plots). Upper plots show the coupled velocity response when excited at the stapes by a sinusoidal of 1 ms^{-1}, the medium plots show the impulse responses and the lower plots show the frequency glides.



(a) x=4mm **(b)** x=14 mm **(c)** x=4mm **(d)** x=14 mm

(e) x=4mm **(f)** x=14 mm **(g)** x=4mm **(h)** x=14 mm

(i) x=4mm **(j)** x=14 mm **(k)** x=4mm **(l)** x=14 mm

Figure H.14: Details of the passive (dashed blue line) and active (solid red line) coupled response of a shoe box cochlea using a diminished fluid coupling configuration where $\frac{\Delta_M}{\Delta_T} = 0.42$ (left and middle left plots) and a 3D fluid coupling (middle right and right plots). Upper plots show the coupled velocity response when excited at the stapes by a sinusoidal of 1 ms^{-1}, the medium plots show the impulse responses and the lower plots show the frequency glides.

Fig. H.12 shows magnitude and phase of the frequency responses at some positions along the length of the cochlea for a passive cochlea, $\gamma = 0$, and for an active cochlea, $\gamma = 1$. Several fluid coupling configurations are studied. For the case of a 1D fluid coupling case, the results show narrow peaks, with a maximum amplification of 24 dB. As observed, the phase lags after the CF are placed between 2 and 6 cycles. If the near field fluid coupling is increased, the peaks become broader, and the quality factor of the responses decreases, however, a greater amplification is achieved. As the near field fluid increases, the phase lag after the CF also does. If the fluid coupling is selected to be fully 3D, the shape of the responses varies largely. The 3D fluid configuration gives a large amplification at low frequency, which is not believed to be found in a human cochlea, as discussed by Robles and Ruggero [187].

Fig. H.13 and Fig. H.14 show more details of the coupled response at 4 mm and 14 mm along the length of the cochlea, for passive and active operation modes using different fluid coupling configurations. Along these figures there are shown the Bode plots of the velocity response and the impulse responses, obtained by inverse Fourier transformation. The frequency glides are also presented, which show the instantaneous frequency at a given time of the impulse response.

Theoretically, the BM has the same total mass per element independently of the fluid configuration, as explained in Section H.5.4. It can be observed in the Bode plots, however, how the peaks of the passive response are lowered in frequency if more 3D fluid coupling is used, which suggests that loading of the effective mass added by T is lower than the equivalent loading by the near field fluid. This has also been observed in the results shown by Elliot *et al.* [260]. In the formulation of the model, the response of the feedback loop providing the action of the CA is out of the influence of the fluid coupling configuration, and so the peaks of the active responses. By observing the Bode plots of Fig. H.13, it can be observed how the CF of the active response are lowered in frequency with respect to the CF of the lower response. For other fluid coupling configurations this aspect is not observed, which suggests that for the values for the parameters used in the model a 1D fluid configuration can not be applied, as a half-octave shift in the CF is expected to be found between active and passive responses [264]. This is also seen in the IRs, in where the active response zeros are more spaced than those of the passive response, as they correspond to a lower frequency, which is not consistent with the IRs observed in measurements, in where the period of the

[2]This parameters are shown in Table H.1

active response is always reduced from that of the passive responses [194, 192].

H.6.2 Responses based on model with varying cochlear chambers

Although a straight box cochlea provides a good approximation, a more refined model can be obtained by using the dimensions of a real cochlea to calculate the far- and near-field fluid coupling. By using the formulation developed by Elliott *et al.* [257], it is possible to take into account the variation of area that the different scalaes found in the cochlea present in a real specimen. The left hand side plot of Fig. H.15 shows the assumed variation of area of the upper fluid chamber (scala media and scala vestibuli) and the lower fluid chamber (scala tympani). These areas were published by Elliott *et al.* [260], based on figures obtained by Zakis and Witte [265], and posteriorly interpolated with a spline function. The BM width, B, is assumed to increase linearly from 0.2 mm at the base to 0.5 mm at the apex and the cochlear partition width, W, to be 1 mm all along the cochlea.

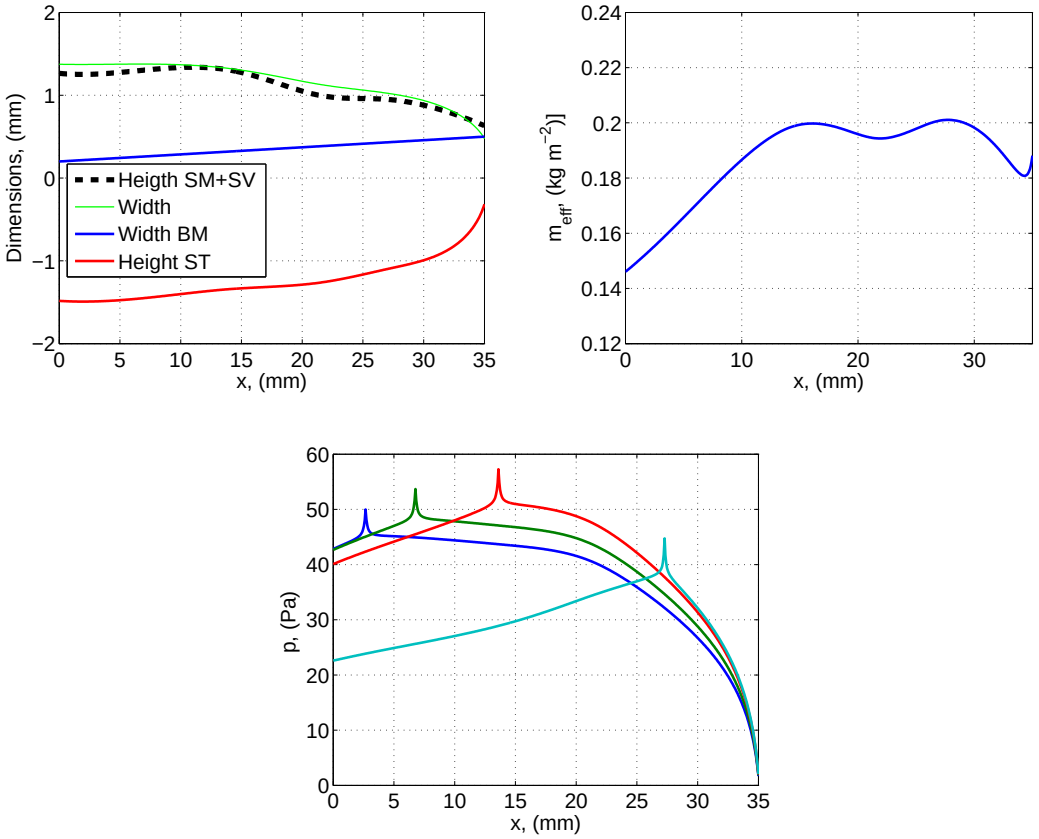


Figure H.15: Assumed variation in the cochlear chamber dimensions and the corresponding effective mass added to the BM due to the near field fluid. The lower image shows the calculated mean pressure due to far- and near-field components in the fluid matrix when only a single element of the discrete BM at x= 2.7. 7, 14 or 27 mm is driven sinusoidally with a velocity of 10 mms^{-}1 at 1 kHz.

The near-field component of the fluid is calculated by inverse Fourier transforma-

tion of the wavenumber description of the total fluid and posterior subtraction of the part correspondent to the far field fluid coupling [260]. The effective mass which is added to that of the BM in this configuration is shown in the right hand side plot of Fig. H.15. As it is observed, this is similar in magnitude to that obtained for a straight box cochlea, with some ondulations due to the variations in the fluid chambers. The distribution of pressure difference created with these fluid chambers is shown in the lower plot of Fig. H.15. The pressure magnitude is similar to that found in a straight box cochlea, but the shape of the curves is slightly different.

Fig. H.16 introduces the coupled responses calculated from the model, together with the impulse responses and the frequency glides. The coupled responses resemble those shown by Ramamoorhty *et al.* [190]. Prior to the peak of the active response the magnitude of the response is slightly decreased, followed by a large enhancement, due to the CA action. The IRs show a shape similar to that observed in measurements [194], and so the glides for the active responses, which stabilise around the frequency of the peak of the BM a few ms after the start of the response [251].

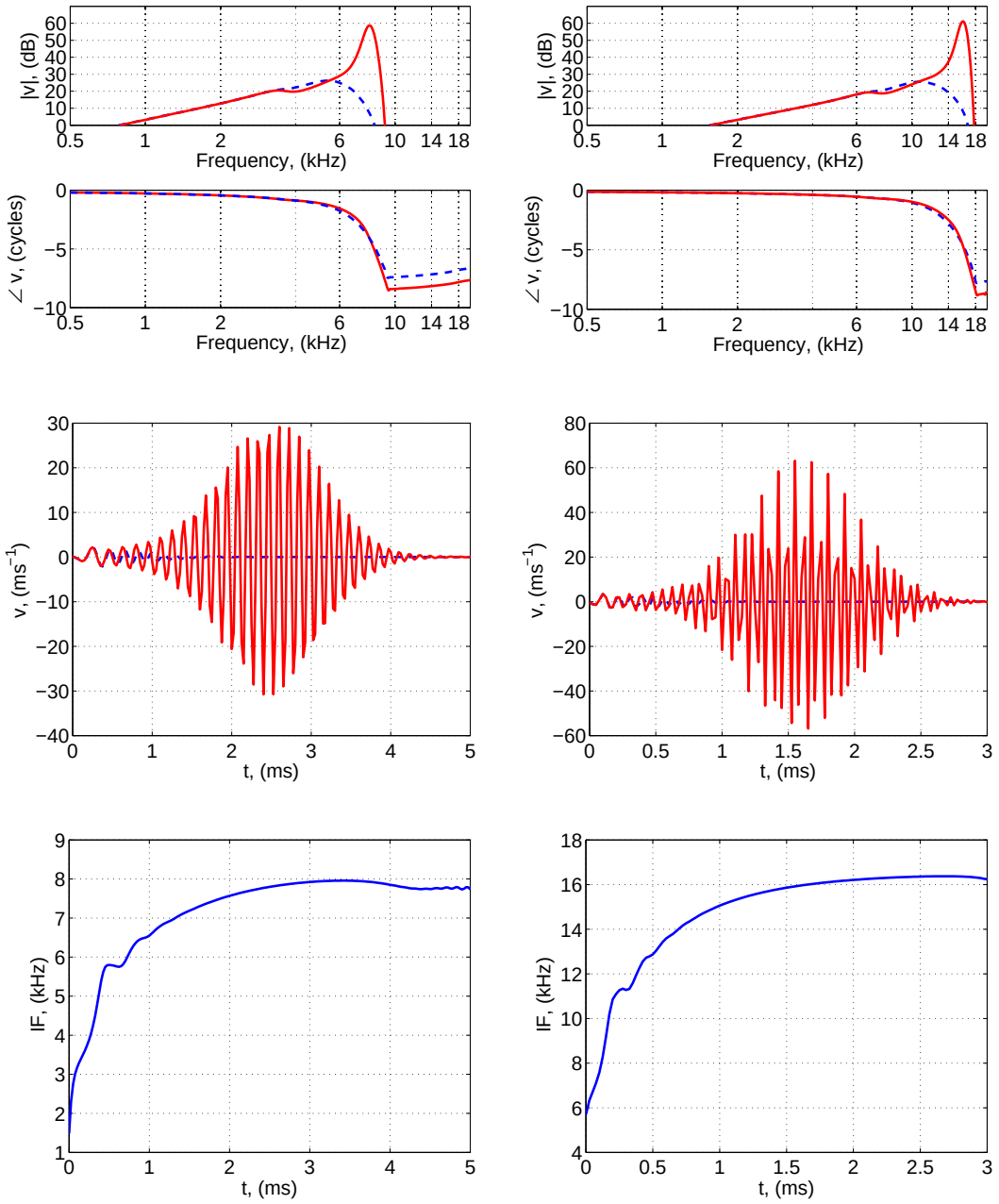


Figure H.16: Details of the passive (dashed blue) and active (solid red) coupled response of the model with varying cochlear chambers using a diminished fluid coupling configuration where $\frac{\Delta_M}{\Delta_T} = 0.42$. Upper plots show the coupled velocity response when excited at the stapes by a sinusoidal of 1 ms^{-1}., the medium plots show the impulse responses and the lower plots show the frequency glides. The left hand side plots show results for x=5.2 mm and the right hand side plots show results for x=10.2 mm.

H.6.3 Calibration of the model

Since there is little work on relating the number of cochlear sections in the mdoel to the number of rows of outer hair cells, $N = 512$ elements have been used to simulate a human cochlea. Several configurations and sets of parameters have been tried, until the best balance has been obtained. Attention has been put to try to obtain responses which accurately replicate a human cochlea, aiming to obtain:

- A maximum gain which simulates that of human cochlear filters.
- A half octave frequency shift between passive and active responses.
- A quality factor comparable to that observed in human cochlear tuning estimates.
- A very small variance of the placement of the zeros of the impulse responses with active gain.
- A growing glide that stabilises at a frequency after a certain time.

Fig. H.17 shows the amplification and the quality factor obtained with the fluid coupling and cochlear chambers presented in Sections H.6.1 and H.6.2. It can be observed how these two parameters are greatly affected by the amount of near field fluid coupling used.

As it can be observed in Fig. H.17, a configuration which provides a good balance between amplification and quality factor, is that found when varying cochlear chambers and a fluid coupling of $\frac{\Delta_M}{\Delta_T} = 0.42$ are used, which suggests the use of such set up as the final model. This configuration provides around 35 dB of amplification at high frequency, which allows to replicate the hearing loss of the standard ISO [12] and simulate hearing losses, as previously done with the Neely and Kim model [201].

Fig. H.18 shows the magnitude and phase of the velocity of the cochlea when this is excited by a sinusoidal velocity distribution at the stapes. Another way of visualising the dynamics of the cochlea is by plotting the frequency response of a certain point along its length, as shown in Fig. H.19. As observed in the figure, the CF shifts in frequency about a half an octave between passive and active cases, as observed in *in vivo* measurements [264] and about 33 dB of amplification are found around 16 kHz.

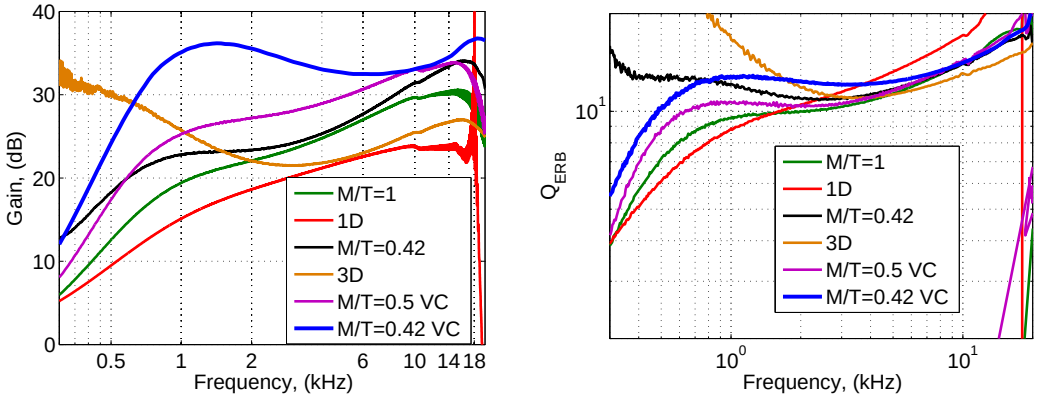


Figure H.17: Comparison of maximum gain and equivalent rectangular bandwidth quality factor, Q_{ERB}, given by the model for straight and varying cochlear chambers and various fluid coupling configurations.

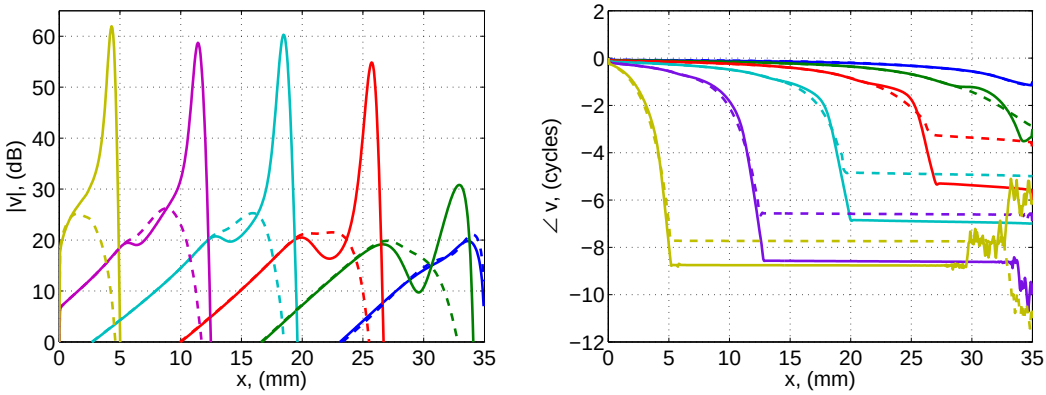


Figure H.18: Coupled active (solid lines) and passive (dashed line) BM velocity distribution of the cochlea with varying chambers at excitation frequencies of 500 Hz, 1 kHz, 2 kHz, 5 kHz, 10 kHz and 20 kHz.

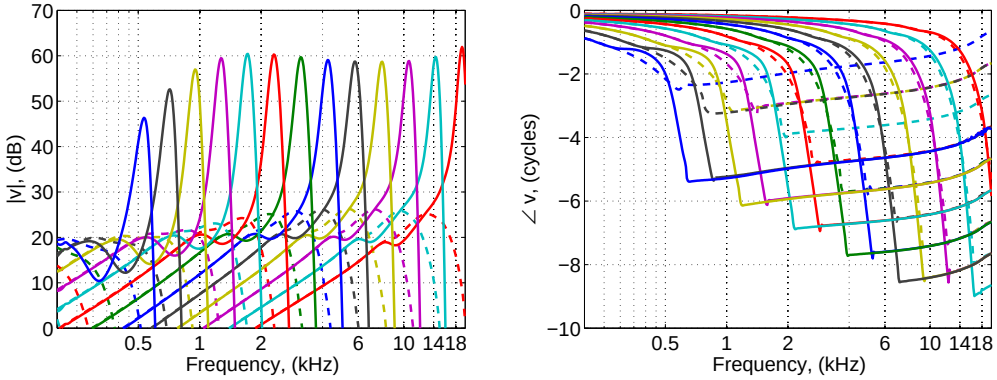


Figure H.19: Response of the coupled passive (dashed lines) and active (solid lines) cochlea with varying chambers using a diminished 3D fluid configuration with $\frac{\Delta_M}{\Delta_T} = 0.42$. Response shown between x=2.1 mm until x=34.9 mm, in increments of 2.1 mm.

Appendix I

Outer and middle ear models

I.1 Stapes input impedance

The presented outer ear model is based on the ear canal model defined by Egolf *et al.* [266] and the middle ear model of Kringlebotn [267]. The middle ear and ear canal model are the same as the ones used by Emery Ku [268] and Jackie Young [269].

In order to check if a middle ear model can be used, it is needed to see if the stapes input impedance of our cochlear model is similar to the impedance used for the stapes by Kringlebotn's [267]. The cochlear input impedance as defined by Emery Ku [268] is

$$Z_{IN} = \frac{P_{ST}}{u_{ST} A_{ST}}, \tag{I.1}$$

with $A_{st} = 0.00001$ m ^{2} to obtain a volume velocity. The behaviour of Z_{IN} against frequency is shown in Fig. I.1.

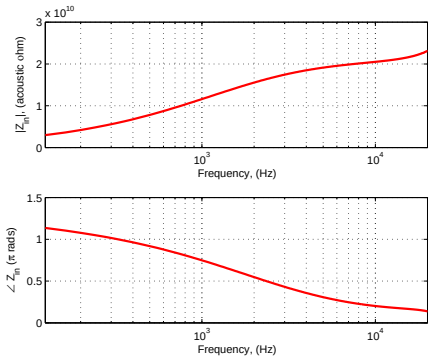


Figure I.1: Coupled cochlea input impedance as in Eq. I.1.

The value obtained for Z_{IN} is plotted in acoustic ohms, giving a result close to measurements performed by Puria [270], being Z_{IN} the impedance value used to

calculate other outer ear parameters.

I.2 Middle ear and ear canal models

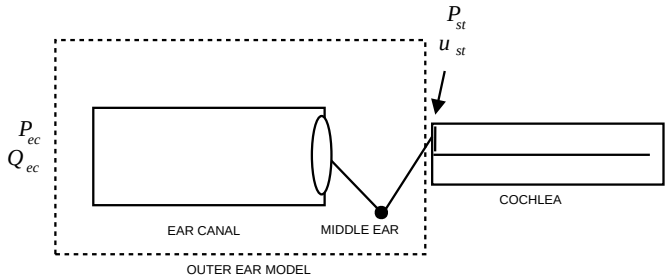


Figure I.2: Representation of the whole outer ear model.

The middle ear model used can be observed in Fig. I.2. By using a two port network model the transfer function between the ear canal input and the staples can be obtained. The whole two port network used can be observed in Eq. I.2. The matrix that includes R is used to simulate an earplug.

$$\begin{bmatrix} P_{EC} \\ Q_{EC} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/R & 1 \end{bmatrix} \begin{bmatrix} T_{ECED11} & T_{ECED12} \\ T_{ECED21} & T_{ECED22} \end{bmatrix} \begin{bmatrix} T_{EDST11} & T_{EDST12} \\ T_{EDST21} & T_{EDST22} \end{bmatrix} \begin{bmatrix} P_{ST} \\ Q_{ST} \end{bmatrix}. \quad \text{(I.2)}$$

By multiplying the three transfer function matrices, a new transfer function matrix that characterizes the whole system is obtained

$$\begin{bmatrix} 1 & 0 \\ 1/R & 1 \end{bmatrix} \begin{bmatrix} T_{ECED11} & T_{ECED12} \\ T_{ECED21} & T_{ECED22} \end{bmatrix} \begin{bmatrix} T_{EDST11} & T_{EDST12} \\ T_{EDST21} & T_{EDST22} \end{bmatrix} = \mathbf{T}_{ecst} \quad \text{(I.3)}$$

so that,

$$\mathbf{T}_{ecst} = \begin{bmatrix} T_{ECST11} & T_{ECST12} \\ T_{ECST21} & T_{ECST22} \end{bmatrix}. \quad \text{(I.4)}$$

The transfer function of interest is the one that occurs between stapes velocity and the ear canal input pressure. Being P_{EC} defined as

$$P_{EC} = Z_{EC} Q_{EC}. \quad \text{(I.5)}$$

The input impedance to the ear canal, Z_{EC}, is then defined as

$$Z_{EC} = \frac{T_{ECST11} Z_{st} + T_{ECST12}}{T_{ECST21} Z_{st} + T_{ECST22}}. \quad \text{(I.6)}$$

The result obtained for Z_{EC} can be observed in Fig. I.3, being similar to the

ones obtained by Shaw & Stinson 1981 (Fig.7 of Kringlebotn's model [267]) until 6 kHz. Up to this frequency there can be observed that the response is quite peaky. There are not well known measurements up to these frequencies, as the response varies largely and depends on each individual.

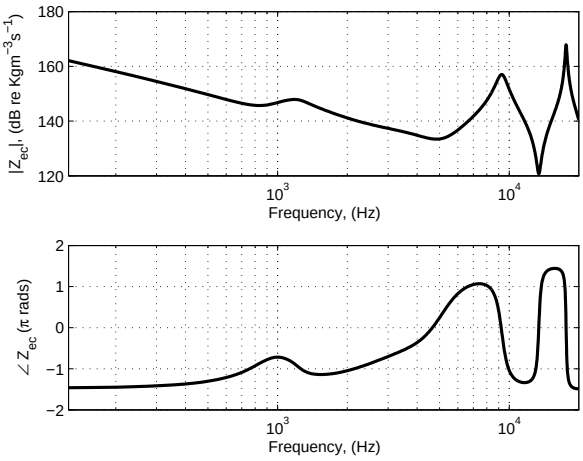


Figure I.3: Acoustical impedance seen at the input of the ear.

By using Eq. I.5 the forward transfer function between input pressure and stapes velocity is obtained

$$Q_{EC} = T_{ECST21} P_{ST} + T_{ECST22} Q_{ST} \tag{I.7}$$

$$\frac{Q_{ST}}{Q_{EC}} = (T_{ECST21} Z_{ST} + T_{ECST22})^{-1} \tag{I.8}$$

$$\frac{u_{ST}}{P_{EC}} = [Z_{EC} A_{ST} (T_{ECST21} Z_{ST} + T_{ECST22})]^{-1} . \tag{I.9}$$

As Z_{ST}, Emery used in his thesis to be equal to 1.10^{10} acoustics ohms. Here it is used the impedance plotted in Fig. I.1. The transfer function of interest is that between the input pressure at the ear canal and the stapes linear velocity, which is the input to the cochlear model we are using. This transfer function is denominated as, $\frac{u_{ST}}{P_{EC}}$, which can be observed in Fig. I.4.

Following, there are shown other forward transfer functions used by different authors. Kringlebotn uses the Forward Pressure Gain, FPG, that he defines as "sound pressure difference across the cochlear partition at the basal end, divided by sound pressure at the ear drum". It can be obtained by performing

$$FPG = \frac{P_{ST}}{P_{ED}} = T_{EDST11} + \frac{T_{EDST12}}{Z_{ST}}. \tag{I.10}$$

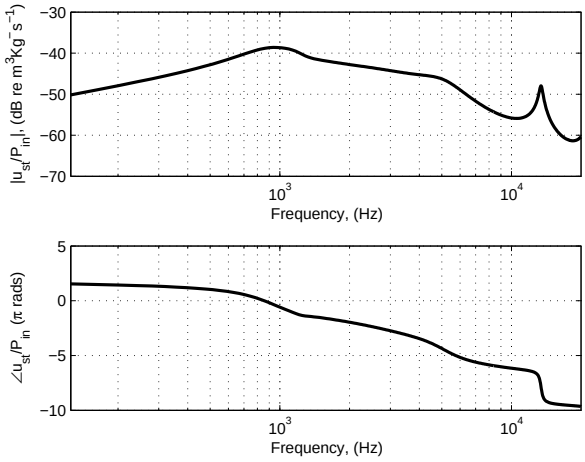


Figure I.4: Forward transfer function between the input pressure to the ear canal and the velocity of stapes $\frac{u_{ST}}{P_{EC}}$.

The result can be observed in Fig. I.5, using Z_{st} as the one defined in Eq. I.1 or using Z_{ST} 1.1e10 acoustic ohms. Another parameter that we introduce

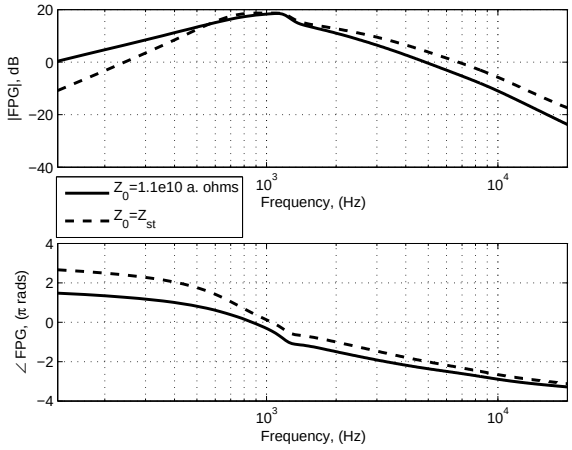


Figure I.5: Forward pressure transfer function between the eardrum and the stapes area.

here is the Overall Forward Pressure Gain, $OFPG$, which takes into account the amplifying effect of the ear canal. $OFPG$ is defined as

$$OFPG = \frac{u_{ST}}{P_{EC}} Z_{ST}, \tag{I.11}$$

which result can be seen in Fig. I.6.

Emery Ku defines the forward transfer function as the ratio between the acceleration of the stapes footplate, a_{ST} and the volume displacement in the ear canal

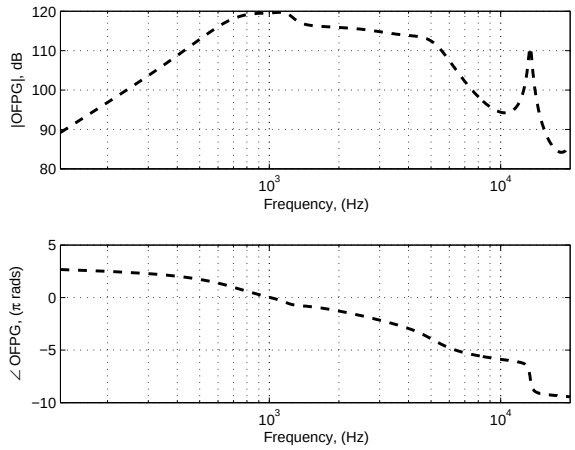


Figure I.6: Overall forward pressure transfer function between the eardrum and the stapes area.

$\int Q_{ST}$, i.e.,

$$\frac{a_{ST}}{\int Q_{EC}} = \frac{-\omega^2}{A_{ST}} \frac{Q_{ST}}{Q_{EC}} \tag{I.12}$$

The result obtained for this transfer function can be observed in Fig. I.7.

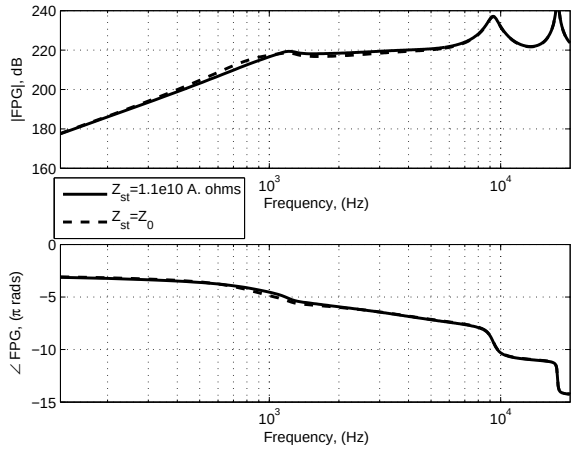


Figure I.7: Transfer function between stapes acceleration and volume velocity at the ear canal.

Appendix J

List of the equipment used for the real-time measurements

J.1 Measurements procedure

J.1.1 Transfer function and directivity estimation

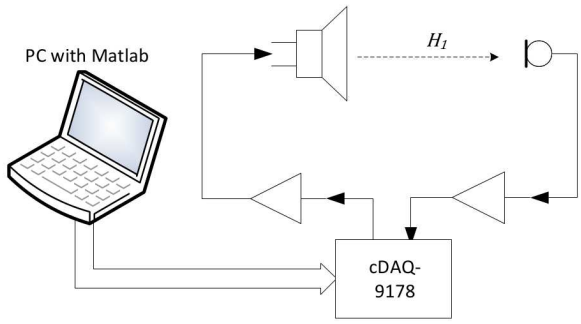


Figure J.1: Measurement set-up used for estimating a transfer function, H_1 between a source and a microphone.

The transfer functions of the loudspeakers used in this thesis have been obtained using the H_1 estimator, which is defined by [120]

$$H_1(j\omega) = \frac{S_{xy}(j\omega)}{S_{xx}(j\omega)} \tag{J.1}$$

where $S_{xx}(j\omega)$ and $S_{xy}(j\omega)$ represent the auto spectra of the input voltages to each source and the cross spectra between the input voltage to each source and the output voltage of each microphone . These two parameters have been obtained by performing measurements, using white noise as excitation, of the input voltage to the loudspeaker and the output voltage from the microphone, which can be observed in Fig. J.1. Such voltages were registered using a computer using

Matlab together with a National Instruments cDAQ-9178 acquisition card. The quality of the measurements has been assessed by using the coherence estimator, $\gamma^2_{x,y}(j\omega)$, which is defined as

$$\gamma^2_{x,y}(j\omega) = \frac{|S_{xy}(j\omega)|^2}{S_{yy}(j\omega)S_{xx}(j\omega)}, \tag{J.2}$$

and which must be close to 1 for the estimation to be accurate.

The estimation of the directivities has been performed by measuring the transfer functions of a source by a surrounding circumference of microphones. For the measured real time directivity plots, the measured acoustical pressures are shown instead of the measured transfer functions.

J.2 Measurement of real time performance of arrays

- Laptop or PC running Matlab 2011 a with data acquisition toolbox
- National Instruments cDAQ-9178, Compact DAQ 8-slot USB Chassis
- 2x National Instruments 9263, 4-channel output module
- 4 National Instruments 9234, 4-channel input module
- 8 channel power amplifier, custom built in the ISVR workshops
- 7 PCB Piezotronic 130A24 microphones
- 1/2" microphone calibrator

J.3 Measurement of loudspeaker directivities

- Laptop or PC running Matlab 2011 a with data acquisition toolbox
- National Instruments cDAQ-9178, Compact DAQ 8-slot USB Chassis
- National Instruments 9263, 4-channel output module
- 3 National Instruments 9234, 4-channel input module
- 12 Panasonic WM-063T electret condenser microphones
- 12 electret microphone pre-amplifiers (custom built)

- 1/4 Microphone calibrator
- Power amplifier: HH Electronic VX-200