UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

Institute of Sound and Vibration Research

Models of binaural hearing for sound lateralisation and localisation

by

Munhum Park

Thesis for the degree of Doctor of Philosophy

October 2007

"The hearing ear, and the seeing eye, the LORD hath made even both of them."

Proverb 20:12

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
INSTITUTE OF SOUND AND VIBRATION RESEARCH

Doctor of Philosophy

MODELS OF BINAURAL HEARING FOR SOUND LATERALISATION AND LOCALISATION

by Munhum Park

The current study suggests two models of binaural hearing, which aim to make predictions for inside- and outside-head localisation of a single sound source in the horizontal plane. Both models consider free-field ITDs and ILDs as the memory of sound localisation to which the target interaural disparity is compared. The first model, the characteristic-curve (CC) model acquires the best estimate of a source location by finding the nearest-neighbour of the target ITD and ILD in the characteristic curve of free-field interaural disparities. On the other hand, the second model, the pattern-matching (PM) model, assumes that the excitation-inhibition cell activity pattern suggested by Breebaart et al. [J. Acoust. S. Am., 110(2):1074-1088, 2001] provides the internal representation of the sound localisation cues. Given the uniqueness of EI-patterns, the pattern-matching process operates in each auditory frequency band to give an estimate of the sound source position, which is then frequency-weighted to finally establish the probability function of target location. In the two listening tests presented in the current study, it has been found that both models are capable of predicting many important features of human sound localisation. For example, the inside-head localisation (laterality) of dichotic pure tones has been reasonably well predicted at low source frequencies, 600 Hz and 1200 Hz, by the CC model individualised for each participant. In addition, the prediction of the PM model has been successfully compared to listening test results where the outside-head localisation of the participants was investigated for real and virtual acoustic sources. Given the simplicity and the originality in modelling the central processes of auditory spatial hearing, particularly in handling the ILD information of binaural signals, the predictive scope of the models is regarded as being worthy of further investigation. Furthermore, considering the reasonable predictions made for both lateralisation and localisation of acoustic stimuli, the models developed appear also to be well-suited to the computational evaluation of spatial audio systems.

Contents

List of Figures

List of Tables

Author's Declaration

I, Munhum Park, declare that this thesis entitled, 'Models of binaural hearing for sound lateralisation and localisation' 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 the 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;
- parts of this work have been published as:
  - M. Park, P. A. Nelson, and Y. Kim. (2005). An auditory process model for sound localization. IEEE WASPAA, New Paltz, New York.
  - M. Park, P. A. Nelson, and Y. Kim. (2006). An auditory process model for the evaluation of virtual acoustic imaging systems. 120th AES convention. Paris, France.
  - M. Park, P. A. Nelson, and Y. Kim. (2006). An auditory process model for the evaluation of virtual acoustic imaging systems. IOA, Southampton, U.K.

Signed: _____

Date: _____

Acknowledgements

First of all, I would like to thank God. His guidance has been truly amazing in my life! My sincere gratitude is then to my supervisor, Philip Nelson for his advice and encouragement, and to Mark Lutman, Ben Lineton and Daniel Rowan in the ISVR for their specialist help on the non-engineering aspects of the project. I would also like to thank 14 friendly participants in the experiments who willingly became acoustic dummies and patient listeners. (It is very much of pity not to be able to name each of them here.) For the financial support, I am grateful to the sponsors of the project, Samsung Advanced Institute of Technology (Youngtae Kim) and Electronics and Telecommunications Research Institute (Kyeongok Kang). Also, I am much in debt of prayer and love to Southampton Korean Church (Hyunin Moon), Jeongreung Jungang Presbyterian Church (Sanghyun Kwon) and Daedeok Presbyterian Church (Jungsam Lee). Finally, this thesis is dedicated to my family.

Abbreviations

AIM	**A**uditory **I**mage **M**odel
CC model	**C**haracteristic **C**urve model
CPP	**C**onstant **P**ower **P**anning
EC	**E**qualisation and **C**ancellation
EI	**E**xcitation and **I**nhibition
EID	**E**xcitation-**I**nhibition **D**ifference
ERB	**E**quivalent **R**ectangular **B**andwidth
FFT	**F**ast **F**ourier **T**ransform
GUI	**G**raphic **U**ser **I**nterface
HRIR	**H**ead-**R**elated **I**mpulse **R**esponse
HRTF	**H**ead-**R**elated **T**ransfer **F**unction
IAD	**I**nter-channel **A**mplitude **D**ifference
IHL	**I**nside-**H**ead **L**ocalisation
ILD	**I**nteraural **L**evel **D**ifference
ITD	**I**nteraural **T**ime **D**ifference
JND	**J**ust **N**oticeable **D**ifference
OHL	**O**utside-**H**ead **L**ocalisation
PM model	**P**attern-**M**atching model
PRMSE	**P**ercentage **R**oot **M**ean **S**quare **E**rror

Symbols

$'$	Unless otherwise stated, $'$ indicates variables in the converted ITD-ILD space
$D(\theta)$	Probability function of source position
EI''_{tg}	Target EI-pattern
EI''_{T}	Template of EI-pattern
H_θ	Original HRTFs (or HRIR) in time domain
H'_θ	Interpolated HRTFs (or HRIR) in time domain
$\mathbf{L}$, $\mathbf{R}$	Amplitude gains given to the left/right channels of stereophony
L_i, R_i	Left and right channels in the ith auditory frequency band
M	Magnitude response of HRTF in linear scale
M'	Magnitude response of HRTF in dB
T	Period of pure tone signal
$W(f)$	Frequency weighting function
c	Speed of sound
d	Distance between loudspeaker and subject (center of interaural axis)
$\mathbf{e}$	Least square error in the CC model
e_L	Signal energy in left channel
e_R	Signal energy in right channel
f	Frequency in Hz
f_s	Sampling frequency in Hz
g_1, g_2	Amplitude gains given to channel 1 and 2 of non-symmetric stereophony system
h	Distance between ears
k_α	Conversion factor for ILD axis
k_τ	Conversion factor for ITD axis
n	Noise mask applied to EI-pattern
$p(\tau)$	Neural sensitivity function
t	Time sequence
w	Time window applied for the computation of EI-patterns
w_1	Sound pressure at receiver channel 1
w_2	Sound pressure at receiver channel 2

Δt	Time delay between direct and reflected sound waves
Δx	Displacement of subject in left-right direction
Δy	Displacement of subject in forward-backward direction
$\Delta\alpha$	Level difference between adjacent attenuation taps
$\Delta\theta$	Rotational displacement of subject in horizontal plane
$\Delta\tau$	Time difference between adjacent delay taps
Φ	Phase response of HRTF
Φ'	Phase response of HRTF in exponential form
Ψ, $\widehat{\Psi}$	Auto-correlation and normalised auto-correlation
δ	Time disturbance given in binaural processor
ε	Gain disturbance given in binaural processor
ζ	PRMSE between original and interpolated HRTFs
θ	general notation for angular location in azimuth
θ_1, θ_2	Angular location of loudspeaker 1 and 2 in stereophony system
θ_D, θ_R	Direction of incidence for direct and reflected sound waves
θ_a	Target image location
θ_c	Angular location of the midpoint of two loudspeakers
θ_m	Intermediate argument used to derive the CPP law
θ_p	Source location either predicted by model or reported by subject
θ'_p	Angular location indicated by head-tracker
$\widetilde{\theta_p}$	Source location reported by mispositioned subject
θ_t	Targeted position or location associated with free-field ITD and ILD
θ_{kem}, θ_{sbj}	Angular location associated with KEMAR and subjects' HRTFs
σ_δ, σ_ε	Standard deviation of random variables, δ and ε
σ_n	Standard deviation of noise mask n
τ, α	General notation for ITD and ILD
τ_n, α_n	Free-field ITD and ILD
τ_{tg}, α_{tg}	Target ITD and ILD
τ_0, α_0	Approximate time and level differences between binaural signals
ϕ_a	Interaural phase difference given by stereophony system
ϕ_r	Interaural phase difference given by a free-field sound source
χ	Cross-correlation function between EI-patterns
ψ	Angular aperture of two loudspeakers in stereophony system
ω	Angular frequency in rads^{-1}

Dedicated to my family

Chapter 1

Introduction

It is well known that the location of receptors on the retina has a 1:1 correspondence to the 2D projection of our 3D space, and that such a relationship is maintained at higher levels of processing along the visual pathways [2]. However, there is no similar point-to-point correspondence between a spatial location and the perceived locus of an acoustic image at lower peripheral stages of our hearing system. Instead, it is believed that the localisation of sound stimuli occurs entirely as a consequence of neural processing of monaural or binaural signals [3]. Although spatial orientation by audition is a purely computation-based perception, relevant listening tests have demonstrated that humans are quite accurate in localising a single sound source: On the horizontal plane, the mean error for a stimulus directly in front is approximately 1°, although it can be up to 10° for sound sources to the sides [4].

Because of this nature of spatial hearing, many computational models have been developed that simulate the hearing processes. These have particularly been focused on explaining the results of listening tests where subjective judgements of an acoustic image position have been investigated [5–11]. In these models, the peripheral hearing processes have been represented by simple signal processing modules that reflect the experimental findings of auditory physiology, for example, regarding the transfer characteristics of the basilar membrane followed by the generation of neural impulses at the inner hair cells in the organ of Corti [1, 8, 12–14]. However, only a little is known about the binaural processes where the two monaural neural signals from each peripheral processor are combined for further computation. Therefore, most of the models have been based on Jeffress' model of coincidence detector [5], which has remained only hypothetical until the recent discovery of brain cells in birds that exclusively respond to simultaneous neural inputs from both channels [15, 16].

Jeffress' model [5] has been successful in describing possible neural structure for computing interaural cross-correlation, and thus the interaural time difference (ITD), but it has been modified in later studies in order to handle the interaural level difference (ILD) information, which is believed to be one of the important localisation cues on the horizontal plane [4]. For example, Lindemann [8] extended Jeffress' delay line with modular elements accepting the static and dynamic inhibitions from the contralateral channel, while Stern and Colburn [11] multiplied the interaural cross-correlation with a weighting function which reflects the influence of the ILD on the lateral position of acoustic image. In the meantime, the importance of free-field localisation cues has been incorporated in Gaik's model [6] where the ITDs with naturally associated ILD values were given more weight.

Nevertheless, the underlying neurological process that obtains the ILDs has yet to be discovered and therefore, the above introduced weighting schemes that incorporate the ILD information are perhaps too complicated without solid evidence from neuroscientific findings.

In the current study, two models of sound localisation are suggested which handle the ITD and ILD information simultaneously without the requirement of additional weighting schemes. Best categorised as the central processor or the decision device of a binaural hearing model, the first model investigates the target interaural disparities on the 2-dimensional ITD-ILD space, where position estimates are given simply by finding the closest match to the function describing the relationship of free-field ITDs and ILDs. The natural combinations of the free-field ITDs and ILDs are, in this study, described as comprising the **characteristic curve** in a certain auditory frequency band.

Being consistent with the **characteristic-curve** (CC) model in the emphasis on free-field ITDs and ILDs, the second suggested model is slightly more sophisticated than the first, and the model includes all of the peripheral, binaural and central processes in the auditory pathways [4]. Especially for the binaural and the central processes, an EI-cell activity pattern [1] over ITD-ILD space (instead of a single point on the ITD-ILD space in the CC model) is considered to be the internal representation of sound localisation cues. The best estimate for each frequency band is obtained by a pattern-matching procedure with reference to the collection of EI-patterns predefined for all possible azimuthal directions. Overcoming a few issues in the CC model, for example, the handling of the waveform and the envelope ITDs, the **pattern-matching** (PM) model can give a single estimate of target location for a broadband sound source by combining the predictions in each auditory frequency band according to an experimental frequency weighting scheme.

The main purpose of the current study is to present the structures of the two binaural hearing models and to validate their predictions in actual listening tests where subjects perform inside- and outside-head localisation [17]. Compared to the previous studies reported in the literature, however, experimental arrangements in the current listening tests are advanced with up-to-date measurement techniques, while a wider range of target conditions will be dealt with in a comprehensive investigation of subjective responses. In particular, the results of the localisation listening tests are expected to be meaningful not only for the assessment of the established model, but also for the evaluation of different arrangements of multi-channel sound reproduction systems, providing useful insights into optimal loudspeaker configurations.

The details of the model structures will be presented in chapters 2 and 5 for the CC and the PM models, respectively. While their implications for various features of human sound localisation and lateralisation will be described, preliminary simulation results will be compared to the relevant listening test results reported in the literature.

Ideally, head-related transfer functions (HRTFs) contain all information related to spatial hearing for a certain source-receiver configuration [4]. Therefore, if the HRTFs are measured for the participants in the listening tests, the interaural disparities (the characteristic curve) and the EI-patterns arising in the free-field listening environment can be obtained across frequency to establish personalised hearing models. In chapter 3, the procedures and the results of the HRTF measurement are presented for 6 subjects, who participated in the later listening tests. Among many new features of the current HRTF measurement, both proximal-region (source-receiver distance: 0.3 m) and distal-region (source-receiver distance: 1.5 m) measurements will be presented for comparison.

The results of listening tests will be presented in chapters 4 and 6 respectively, for the lateralisation of low-frequency dichotic pure tones and the localisation of broadband real and virtual sources, respectively. In particular, chapter 4 presents the predictions of the simple characteristic-curve model where the comparison with subjective responses is made at two low frequencies, 600 Hz and 1200 Hz. On the other hand, in chapter 6, the pattern-matching model is applied to explain the features found in localisation listening tests, where subjective responses to various stereophonic arrangements, as well as to a single source, will be investigated.

The chapters have been arranged so that the structure of a model and the associated listening test results can be found in sequence, except that the chapter of the HRTF measurement has been placed following the description of the CC model. Therefore, chapters 2, 3 and 4 may be considered to be the first part of this thesis, while chapters 5 and 6 can be regarded as constituting the second. Finally, conclusions will be presented in chapter 7.

Chapter 2

A characteristic-curve model of sound lateralisation and localisation

2.1 Introduction

There have been many models of human spatial hearing which aim to explain the results of subjective listening tests concerning, for example, the localisation and lateralisation of sound images [5–11] and the binaural masking level difference [1, 18]. Since the interaural time difference (ITD) and the interaural level difference (ILD) are believed to be important cues in spatial hearing, especially in the horizontal plane [4], all of those models have devices to process one or both of the interaural disparities. Jeffress' binaural coincidence detector [5] has been extensively employed by most of the relevant models as a device to process the ITD. In the meantime, ILD information has been dealt with in various ways. According to the ILD processing method, spatial hearing models working with both ITD and ILD belong to one of the following three categories.

(1) *Models considering ILD as supplementary input to the coincidence detector.* The 'position-variable' model of Stern and Colburn [11] belongs to this category where ILD has been accounted for by multiplying the activity of the coincidence detector with a Gaussian-shaped intensity-weighting function. About a decade later, Lindemann [8] also modified Jeffress' delay lines [5] by employing an inhibitory weighting function determined mainly by the intensity of the contralateral signal. Gaik [6] extended Lindemann's model [8] by considering the naturalness of ITD and ILD combination as an additional weighting factor.

(2) *Models combining separate estimates from ITD and ILD.* Models in this category find the best matching azimuth angles for a given ITD and ILD separately, then perform a weighted summation across parameters as well as frequency. Macpherson [9] followed this approach where the estimates from the ITD and ILD have been weighted according to the reliability and weighting factors determined during the pre-process. A similar approach has been taken by Braasch [19] who focused on the influence of the target-distracter ratio rather than a conclusive prediction of source location. Pulkki et al. [10] made qualitative comparisons between the interaural disparities across frequency which were given by a real sound source and those by a virtual sound image.

(3) *Models considering a single parameter from ITD and ILD.* Some models of spatial hearing have particularly emphasised that ITD and ILD are closely related, investigating a single estimate obtained from a combined pair of ITD and ILD. Hafter [18] used the time-intensity trading ratio to obtain a single parameter from the interaural disparities. Lim and Duda [7] arbitrarily scaled ITD and converted this to an equivalent ILD and made a vector of interaural parameters for all relevant frequencies, which is then compared with vectors obtained in free-field conditions to give a prediction of sound source location.

In this study, a simple model of auditory central processing is suggested. As classified in category (3), this model considers the natural combinations of ITD and ILD as a reference for spatial hearing, while the nearest-neighbour matching technique [7] is employed to make predictions of the image location. By investigating the ITD-ILD space in a single auditory frequency band, the current model, the characteristic-curve model, attempts to explain some representative features found in the lateralisation of dichotic tones as well as the localisation of free-field stimuli.

Section 2.2 will describe the CC model and its principle assumptions. Section 2.3 will be dedicated to the implications of the model for sound lateralisation, followed by simulation results compared with subjective test data in literature. In section 2.4, the predictive scope of the current model on the localisation will be investigated, and other aspects of the model will be briefly discussed in section 2.5 followed by some conclusion in section 2.6.

2.2 The model of auditory central processing

Most hearing models include a peripheral processor as well as a binaural processor [1, 8, 13]. In the peripheral processor, the acousto-mechanical transfer characteristics and the neural transduction in middle and inner ears are simulated by corresponding signal processing modules, whilst the information relating to ITD and ILD is obtained in the binaural processor. The current model, however, assumes that the ITDs and ILDs are already available for all auditory frequency bands of interest, after the processes in peripheral and binaural devices. Suggested methods to compute the interaural disparities can be the widely used interaural cross-correlation [5] combined with an ILD detector such as the EID (excitation-inhibition difference) processor [19] or the delay and attenuation network introduced by Breebaart et al. [1]. However, the main concern of the current model is not the detailed mechanism of low-level computation, but the central stages of pattern recognition and auditory image formation [4]. Furthermore, no consideration of time-integration or time-varying interaction of ITD and ILD has been made in this study, which implies that the current model simply examines the interaural disparities given by low-level processes at a certain moment of interest.

Another important assumption of the current model is the human's use of natural combinations of ITD and ILD as a reference for spatial hearing. The term of "natural combinations" of interaural disparities is equivalent to "free-field cues" appearing in literature [20]. It has been previously used by Gaik [6], where he experimentally studied the relation between the diffuseness of the auditory image and the deviation of a target combination of ITD and ILD from the curve of natural combinations in ITD-ILD space. Many other models contain similar implications of naturalness of ITD and ILD combinations although, in many cases, without being described explicitly [7, 9, 10, 19–21]. There are some good reasons for this assumption to be acceptable. First of all, it is widely accepted that human beings acquire the sensation of various types of stimuli by learning and adaptation, and the spatial hearing is not an exception, supported by recent studies regarding its plasticity [22–24]. Since a majority of the acoustic signals perceived during the lifetime are from real sources in space, it is likely that our auditory central processor has been being trained mostly by natural combinations of interaural disparities, and it is reasonable to assume that unnatural pairs of ITD and ILD that are rarely experienced will be perceived with reference to the natural combinations. Also, it is perhaps unlikely that there are individual neurons that act as feature detectors for each of the different possible pairs of ITD and ILD in the higher level of the auditory pathway. Instead, it is reasonable to consider that neurons for the natural combinations of ITD and ILD form a basis for so-called distributed coding [2] in higher-level cognition, which will work more flexibly even with a new target ITD and ILD never experienced by

listeners. Relevant physiological evidence has yet to be fully revealed, but the presence of receptive fields dedicated to a single region of space in the barn owl's optic tectum [25] can be reasonably linked to the existence of reference neurons in spatial hearing.

Since the head-related transfer functions (HRTF) contain all the free-field cues available to the corresponding individual, the natural combinations of ITD and ILD can be obtained by analysing those impulse responses frequency by frequency. However, particular care is required in dealing with ITDs in the frequency range above approximately 1500 Hz, since it is well known that the fine structure of the signals are lost during the neural transduction in the organ of Corti, thus leaving only the envelopes [26]. This implies that only the envelope ITDs are available to the higher level of auditory processing, which, based on the above assumption, will serve as reference ITDs.

Natural ITD, τ_n and ILD, α_n obtained from free-field stimulus in the horizontal plane are functions of frequency f and azimuth angle θ, i.e. they can be written as $\tau_n(\theta, f)$ and $\alpha_n(\theta, f)$. These functions of interaural disparities can be displayed in $\tau - \alpha$ space for each frequency band, which forms a closed curve if drawn for sources in the whole range of azimuth angles from 0° to 360°. Since the HRTFs are different from person to person depending on the anthropometry, the shape of this curve at a certain frequency is unique for each individual, and so it can be called the **characteristic curve** of natural combinations of ITD and ILD.

The solid line in Figure 2.1(a) is an example of such a curve obtained from a KEMAR HRTF [27] at 600 Hz, where positive values for ITD and ILD indicate that the signal at left ear is louder and arriving earlier than the signal at the right ear. Since the HRTFs used here have been manipulated to be symmetric with respect to the median plane, the characteristic curve passes exactly through the origin of the coordinate system, which might not be the case in general.

Except for the high-frequency bands where the characteristic curve is established by the envelope ITDs, the current model assumes secondary curves [dashed lines in Figure 2.1(a)] to be available within the perceptual window of ITD, which are the versions of the primary curve shifted by the period corresponding to the band-centre frequency. Many models of spatial hearing, particularly those based on the coincidence detector, assume that the size of perceptual window for ITD can be larger than the largest possible delays for free-field stimuli [8, 28]. Such models accommodate a multiple number of peaks within the correlation window so that predictions of multiple images can be made. For a similar reason, the current model assumes the presence of secondary curves.

Having established the 'memory' of sound localisation on the $\tau - \alpha$ plane by the characteristic curve, the best estimate θ_p of the angular location is found for a target pair of

ITD, τ_{tg} and ILD, α_{tg} on the basis of the least squared error. First, since the units for ITD and ILD are different, the $\tau - \alpha$ space has to be converted to a non-dimensional space, say, $\tau' - \alpha'$. Thus

$$\tau' = \frac{\tau}{k_\tau(\tau, \alpha, f)}, \quad \alpha' = \frac{\alpha}{k_\alpha(\tau, \alpha, f)} \tag{2.1}$$

where k_τ and k_α are the conversion factors for ITD and ILD, respectively. (Unless notified otherwise, a prime will, hereinafter, indicate corresponding variables in $\tau' - \alpha'$ space.)

Then, the 'distance,' $\mathbf{e}(\theta, f)$ between the target disparity and the characteristic curve in $\tau' - \alpha'$ space is given by [see figure 2.1(b)]

$$\mathbf{e}(\theta, f) = \sqrt{\Big(\tau'_{tg} - \tau'_n(\theta, f)\Big)^2 + \Big(\alpha'_{tg} - \alpha'_n(\theta, f)\Big)^2} \tag{2.2}$$

and the model prediction, $\theta_p(f)$ can be finally obtained by finding the minimum of $\mathbf{e}(\theta, f)$:

$$\theta_p(f) = \arg\min_\theta \mathbf{e}(\theta, f) \tag{2.3}$$

It is noteworthy that the target ITD found within the perceptual window will not necessarily be the true ITD, since all the quasi-periodic ITDs will give the same estimate thanks to the secondary characteristic curves.

The presence of the conversion factors in Eq. (2.1) is essential. As an indication of relative influence of the interaural parameters on the displacement of an auditory image, it is reasonable to relate these factors to the time-intensity trading ratio, where the amount of ITD and ILD inducing an equivalent image shift is considered. Many psychophysical experiments revealed that this trading ratio depends on the frequency as well as the ITD and ILD [4], and it is, therefore, obvious that the conversion factors, k_τ and k_α should contain the arguments of τ, α and f. Another possibility is to relate the conversion factors to the just-noticeable differences of ITD and ILD, since the neural resolution in the detection of ITD and ILD could reflect the actual spacings between ITD/ILD detectors. Regardless of psychophysical background, it appears that the conversion factors in Eq. (2.1) should depend on frequency as well as the given ITD and ILD. In particular, the dependence of neural selectivity on ITD has been found in physiological studies as summarised by Stern and Trahiotis [28], and has been implemented in previous models of spatial hearing, for example, by Stern and Colburn [11]. However, it is unlikely to be possible in this study to determine the conversion factors, k_τ and k_α as definite functions of those arguments, since, as suggested by many experimental results derived under different conditions [29], the interactions between interaural time

and level differences are not yet clearly known, and they are "too complicated to be describable by any one-number criterion" as Blauert pointed out [4]. Simulations in the following sections will use constant values for k_τ and k_α that have been found to give the best fit to the listening test results.

In order to imitate the whole process in spatial hearing, the estimates $\theta_p(f)$ in Eq. (2.3) have to be somehow integrated across frequency to give a single estimate of the location of the sound image. Although there are some findings regarding the tonotopic organisation within the primary auditory cortex [30], the interaction between different auditory frequency channels is not yet known. Therefore, in this study the current model will be applied only to a single auditory frequency band.

2.3 Implication of the model for lateralisation

In this section, the model described above will be investigated in terms of the laterality prediction of a dichotic pure tone at low frequency. Since the lateral displacement from the mid line is the issue, distinction between the front and back is unnecessary in this case. In the simulation study discussed in the later part of this section, the final prediction of the current model has been converted to the range between $-90°$ (left) and $+90°$ (right). Meanwhile, for the convenience of explanation in the following paragraphs, the characteristic curves have been simplified to be single straight lines as shown in Figs. 2.2 and 2.3.

Combinations of ITD and ILD used in relevant listening tests have been reported to give intracranial auditory images [4, 29, 31], and so it is also necessary to make an additional assumption for the current model to deal with these internal auditory images and the relationship to their output in azimuth angle. It is known that virtual acoustic images created by a non-individualised HRTF can give internalised images in the case of headphone reproduction, and the externalisation is reported to be accomplished only when the correct interaural and spectral cues are presented to both ears across frequency [32]. According to this observation, it may be reasonable to assume that a pair of ITD and ILD off the characteristic curve is internalised, while external images are perceived when the interaural disparities are exactly on, or in the very vicinity of, the characteristic curves, consistently across auditory frequency bands. With these assumptions, the model output in azimuth angle can be related to the relative lateral displacement of an internalised auditory image.

Figures 2.2 and 2.3 show how the characteristic curve is referenced to give a prediction of laterality at low frequency. First, a target signal without ILD is considered in Fig. 2.2 where the trajectory of target ITD and ILD can be represented by $\alpha' = 0$. When the target ITD is 0 (point o), an auditory image is obviously located at the centre. As τ'_{tg} increases (point a), the corresponding image found by the nearest neighbour matching moves to the left-hand side (remember that positive ITD and ILD indicate louder signals arriving to the left ear earlier, and positive θ_p represents a sound source in the right hemisphere.) Reaching the point where ITD becomes equal to the half-period (the point marked by $T'/2$), the image suddenly migrates to the lower leg of the secondary characteristic curve, which implies that the perceived image is now located on the right-hand side. Then, the image laterality decreases approaching the intracranial centre from the right (point b). The same matching procedure takes place for the negative ITDs (points c and d), and a completed model prediction can be plotted as shown in panel (b) of Fig. 2.2. Relevant listening test data confirm this model prediction, where the

auditory image is reported to move to the side favoured by the increasing target ITD until it makes a sudden transition to the contralateral side [29, 31].

In the meantime, Fig. 2.3 shows the predictions for the non-zero target ILD. Due to the given interaural level difference, the auditory image is not located at the centre when the ITD is 0. Instead, it is located to the left-hand side favouring the positive ILD, and moves further left as the ITD increases (point a). A sudden transition to the contralateral side is also observed. However, the critical ITD in this case is greater than it was for zero ILD [compare with Fig. 2.2(b)]. In addition, the maximum absolute laterality in the direction not favoured by the target ILD is less than that to the favoured side, which implies that the extent of the lateral position is reduced for a conflicting pair of ITD and ILD. This model prediction is also confirmed by the listening test results reported by Sayers [31] and Domnitz and Colburn [29].

Some models of spatial hearing assume that the amplitude and time-delay errors can be introduced prior to the binaural processing [33], which result from the limited accuracy of neural coding and processing as well as the internal physiological noise from the ears and other parts of the human body [4]. These errors could misplace the target ITD and ILD in the $\tau' - \alpha'$ space. In addition to the internal noise, relevant psychophysical tests are exposed to measurement errors, especially because the subject's auditory space has to be represented quantitatively. For example, previous listening tests employed a visual chart or equivalents [31, 34, 35] or an acoustic pointer [29, 36] to quantify the subjects' perception, which inevitably involves some degree of error.

Figure 2.4 shows the influence of the internal errors on the model predictions. A characteristic curve (with waveform ITD) has been obtained from the KEMAR HRTF [27] at 600 Hz, and the laterality has been predicted for ITDs at every $50\mu s$ with 0-dB ILD. For a 600-Hz pure tone, the current model has been found to best explain relevant listening test results when the conversion factors k_τ and k_α are $\sim 44\,\mu s$ and 1 dB, respectively. Internal error has been introduced assuming independent zero-mean Gaussian random processes for target ITD and ILD (see Fig. 2.10) with standard deviation of σ_δ=10 μs and σ_ε=1 dB, respectively. These values have been approximated from the representative data for the just-noticeable difference of ITD and ILD reported in the literature [4]. Since the model output ranges from 0° to 360°, predictions have been converted to be between −90° and +90° by considering the mirror images for the estimates found in the rear hemisphere. The contrast shown for each point in Fig. 2.4 represents the probability of model prediction for each estimate along the vertical axis, obtained from 500 samples of random processes.

It is obvious that the model prediction now appears to be distributed in the θ_p direction, instead of following a single curve as schematically shown in Fig. 2.2(b). The degree of

spread given by, for example, the standard deviation of the model prediction for each target ITD and ILD may be indicative of the diffuseness of the auditory image. In particular, there is a bimodal distribution found in the region where the auditory image suddenly migrates to the contralateral side (see the dotted boxes in Fig. 2.4), and this is closely related to the existence of so-called 'dual images.' Sayers [31] reported that his subjects involved in the laterality measurements were found to give three types of judgements for signals with 0 ILD and approximately half-period ITD: far left, far right or centre. Although the current model does give similar judgements at extreme left and right, no centre image is predicted even with random errors in target ITD and ILD. However, it is still possible to obtain a centre image if the centroid of the judgements in this region is considered. Shackleton et al. [37] and Lindemann [8] have taken a similar approach where they examined either the centroid or the individual peaks in the cross-correlation function depending on the nature of the data to be explained, especially because the centroid could not give predictions of dual images. It is tentatively suggested that subject's attention could switch his or her judgement either to the left or to the right side, and could even fuse the two extreme images within short time interval, thus internally creating a virtual centre image.

It is also noteworthy in Fig. 2.4 that the estimates around the half-period ITD appear to be concentrated at $\pm 90^\circ$, which resulted from restricting the model predictions to be only in the frontal hemisphere. This is inconsistent with listening test results presented by Sayers [31] where a linear increase of absolute lateral displacement has been observed even in the region of bimodal distribution. Perhaps the azimuth angles corresponding to the maximal lateral displacement could be greater than $\pm 90^\circ$. However, further simulations with a new arbitrary choice of possible range of prediction are beyond the scope of this study.

Figure 2.5 compares the model predictions with the listening test results reported by Sayers [31]. Panel (a) shows Sayers' data [31] where he used a 600-Hz pure tone with various combinations of external ITD and ILD, and obtained subjects' perception of lateral position by means of a visual chart. The model predictions in panel (b) have been prepared with KEMAR HRTF [27] databases under the same conditions as described in relation to Fig. 2.4.

Since the units in the two plots in Fig. 2.5 are different, point-by-point comparison appears to be inadequate. Nevertheless, the model prediction in panel (b) is reasonably consistent with the listening test data in panel (a) at least in terms of the shape of the laterality curves. In addition, the current model gives relatively accurate predictions for the critical ITD values where the swift transition to the far sides takes place. It is also observed in both model simulation and subjective data that those critical ITD values

found in the region of conflicting ITD and ILD are smaller when the target ILD becomes greater. This earlier transition to the contralateral side for greater ILD is also observed and reported in Domnitz and Colburn [29].

2.4 Implication of the model for localisation

Application of the current model to the prediction of sound localisation is rather straightforward. A sound source in space presents pairs of ITD and ILD which should be on the characteristic curves across frequency since those pairs were the stimuli that have formed the curves. However, the presence of internal noise discussed in section 2.3 can cause the target pair of interaural disparity to spread over the true values, while the measurement error will also disrupt the accurate quantification of subject's perception. Measurement error in sound localisation tests can be harder to control than that in laterality measurement since the amount of error could depend on the source location. In relevant listening tests, subjects are often asked to turn their head to the location of sound source where the direction of head is automatically detected by an electromagnetic device [38, 39]. The accuracy of this method is limited by the extent of body movement, and so the measurement error could increase for sound sources in the rear hemisphere.

As was the case in lateralisation modelling, the internal errors can be accounted for by considering independent random processes for ITD and ILD, the standard deviation of which can be approximated from the just noticeable difference (JND) of each interaural disparity. The measurement error is not easy to include in the current model, and the best way seems to be to attempt to increase the standard deviation of internal noise, and examine the consistency with listening test results.

Fig. 2.6 schematically shows the influence of the error introduced to simulate the internal noise and the measurement uncertainty. Similar to the lateralisation case, the target ITD and ILD are now random processes with mean values at actual target ITD and ILD but spread over a region depicted by circles in Fig. 2.6. Consequently, the model prediction now becomes a distribution instead of a single definite value as is the case in actual listening tests, and the mean and the variance of the predictions depend not only on the actual target ITD and ILD but also on the adjacent pairs on the characteristic curve.

It is also interesting to see that the spread of target ITD and ILD is able to account for the phenomenon of front-back confusion [4]. For instance, the boundary A in Fig. 2.6 represents the size of confidence interval for a pair of ITD and ILD which originally corresponds to an azimuth angle, for example, θ_t which is front right. Since some samples within the boundary are closer to the other leg of the characteristic curve corresponding to the rear hemisphere (darker area in boundary A), the corresponding estimates are found near $180^\circ - \theta_t$. The probability of front-back confusion is therefore affected by the spacing between the two legs of the characteristic curve and the amount of error introduced into the target ITD and ILD.

Figure 2.7 shows the simulation results of the localisation by the current model, where a 600-Hz pure tone is assumed to be the source. The KEMAR HRTF [27] has been used to generate both the characteristic curve and the target pair of ITD and ILD at every 5° in the horizontal plane. The model parameters, k_τ, k_α, σ_ε and σ_δ were identical to those used in the case of lateralisation in section 2.3. The grey-scale level of each point in Fig. 2.7 represents the probability of model predictions. (The histograms shown in Figs. 2.8(a) through (c) can be regarded as the vertical slices of Fig. 2.7 for each target angle.) It is clearly shown that the Gaussian random processes employed for the internal noise caused the response angles to spread out, which, otherwise, should have been found only on a straight line from the bottom-left to the top-right indicating perfect localisation. In addition, the front-back confusion is also clearly observed as the response angles are found bimodally for each target angle, one area of the responses showing the correct matches while the other for the mirror images with respect to the frontal plane. In particular, it is noteworthy that this bimodal distribution is found for most of the target angles even for lateral angles such as 90° and 270°.

Another interesting feature resulting from the noisy target of the source ITD and ILD is that the mean error and the variability of the model predictions are dependent on the source location. As each point marked on the characteristic curve in Fig. 2.6 corresponds to source locations at every 5°, it is observed that more points are populated closer to 90° and 270° where ITD and ILD slowly vary with source azimuth angle. If the amount of internal noise introduced to the target is assumed to be independent of the source location, the boundary B of the same size as A will contain more prospective estimates, which will result in greater errors and standard deviation in the model prediction. This expectation is consistent with most of the listening test results reported in the literature [4, 38, 39].

Simulation results shown in Fig. 2.7 also confirm this dependence of the model statistics on the source location. It is observed that the local range of responses for each target azimuth angle which represents the variances, becomes greater as the target angle approaches 90° (or 270°) from both positive and negative directions. The individual histograms presented in Fig. 2.8 give a clearer comparison between target angles in terms of the model responses where horizontal axes have been scaled identically. The location estimates form two distinctive peaks [the secondary peak is out of range in panel (a)] one within the correct hemisphere and the other the mirror-imaged. As the source location approaches 90° (or 270°), these peaks become lower, broader, and closer to each other, which implies that the variability of the model predictions becomes greater. Note that there is no definitive relation between the broadened peak and the mean error of the response angles. However, the shape of the histogram seems to change very little around 90° (and 270°) which can be easily observed from the comparison between panels (b)

and (c) in Fig. 2.8, and accordingly, the mean error of the model predictions for lateral target angles increases. In addition, for this region of 'constant' responses, resolving the front-back confusion with a single critical angle, for example, 90° could lead to greater unwanted errors as is the case with actual listening tests.

The above presented simulation results are qualitatively consistent with the experimental data in many previous studies of human sound localisation. However, a comparative analysis cannot be made here, mainly because the current model concerns spatial hearing at a single frequency only, while the mainstream of the previously published data have been acquired by presenting relatively broadband noise to subjects. In order to deal with signals with broader bandwidths, it is necessary to accommodate a form of frequency weighting, properly combining all the local model predictions made in each auditory frequency band, for which, however, there is yet insufficient evidence from physiology or neurology. More importantly, as mentioned in section 2.2, both of the two types of ITD, envelope and waveform, have to be considered and carefully incorporated into the current model depending on the signal frequency. However, the critical frequency between the ranges where a particular type of ITD is effective is difficult to define, and thus, the use of a single characteristic curve assumed in the current model is not suitable. This matter will be returned to in future work.

2.5 Other aspects of the current model

2.5.1 Implication for mid- and high-frequency lateralisation

Having found that the model predictions are qualitatively consistent with experimental data in literature, it is of further interest to see what other implications the current model has for human spatial hearing. Whereas Figs. 2.2 and 2.3 schematically show how the nearest-neighbour matching works at relatively low frequencies, Fig. 2.9 illustrates the implication of the current model at higher frequencies. It is known that ILD could be negligible at very low frequency, while it may be as large as 20 dB at high frequencies due to the increased acoustic shadow effect caused by the listener's head [26]. On the other hand, the ITD is not as heavily affected by the frequency as the ILD, where the maximum ratio of the ITDs at high and low frequencies is only about 2/3 [40]. Therefore, assuming that the conversion factors, k_τ and k_α are constant across frequency, the slope of the simplified characteristic curve shown in Fig. 2.9 becomes steeper as frequency increases, while the spacing between the primary and secondary characteristic curves becomes smaller. In the meantime, the secondary curves become ambiguous and disappear over about 2 kHz due to the loss of waveform ITD, and the ITD of the primary characteristic curve start indicating the envelope ITD, although the transition between these two phases is difficult to define. Accordingly, three solid lines in Fig. 2.9 represent the characteristic curves at low frequencies, and the dashed lines indicate those at mid and high frequencies, where the secondary lines are disregarded at high frequencies.

Keeping in mind that the two ends of the simplified characteristic curves in Fig. 2.9 approximately correspond to $\pm 90^\circ$ which, in terms of lateralisation, are the far right and left sides, it is interesting to see how the relative influence of each interaural disparity varies with frequency. For example, a target of ITD, τ'_{tg} without ILD in Fig. 2.9 is matched on the characteristic curve more to the left at low frequency (θ_{1L}) compared to the estimate found at higher frequencies (θ_{1H}), i.e. $|\theta_{1L}| > |\theta_{1H}|$. On the contrary, another target with a pure ILD of α'_{tg} is detected more to the left at higher frequencies ($|\theta_{2H}| > |\theta_{2L}|$). In addition, the waveform ITD does not have any influence at high frequency, obviously because the characteristic curve is matched only for envelope ITD. This observation made from the slope of the characteristic curves across frequency is consistent with what is suggested by the duplex theory [41] modified with the role of envelope ITD [42]. Thus the waveform ITD and the envelope ITD are effective at low and high frequencies, respectively, whilst the ILD is effective over all the frequency range despite the variance of its importance relative to ITD [4].

The additional implication of the characteristic curves in Fig. 2.9 is that the maximum laterality normally found at around half-period ITD for a zero target ILD should decrease

with frequency due to the narrower spacing between the curves and their increased slope. Sayers [31] reported that listeners seem to perceive a dichotic pure tone within a narrower range of laterality at higher frequency. Decades later, Schiano et al. [36] confirmed this observation by laterality tests across frequency, where he and his colleagues found that the laterality of pure tones with moderate amount of ITDs ($100\,\mu s$ or $150\,\mu s$) is more or less constant up to 1 kHz where it suddenly decreases. Although the current model generally predicts a greater laterality for lower frequency, the sudden collapse to the centre at 1 kHz seems to be difficult to explain, partly due to the increased irregularity of characteristic curve at higher frequency.

2.5.2 Diffuseness of the perceived image

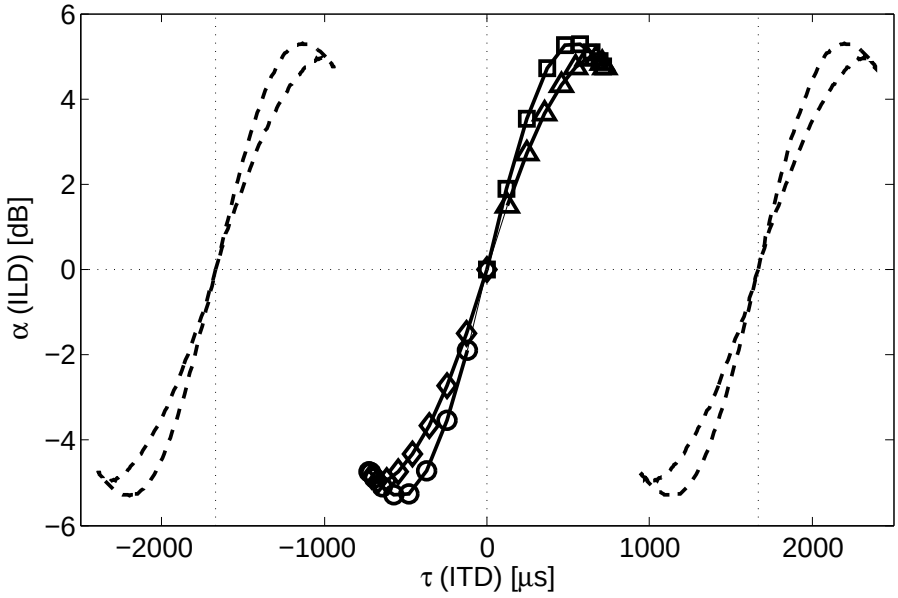
As discussed in section 2.2, conversion factors included in the current model are unlikely to be constant over $\tau - \alpha$ space as assumed in the simulations, but should depend at least on the ITD, similar to the neural sensitivity function $p(\tau)$ suggested by Stern and Colburn [11]. At the same time, the amount of internal error introduced to the target ITD and ILD may vary over $\tau - \alpha$ space. For example, Domnitz and Colburn [29] found that the JND of ITD increases when baseline ITD and ILD conflict with each other. If the internal error in the current model is based on the JND of interaural disparities, the values of σ_ε and σ_δ should also increase in this conflicting region. If the above arguments are considered, it might be expected that there is a greater uncertainty in a target ITD and ILD that is located more distant from the characteristic curve, hence resulting in a greater spread in the distribution of laterality estimates (see Fig. 2.10). Considering that image diffuseness is often related to the variance of the judgements in similar models, this observation of the current model can be regarded as implying that a greater image diffuseness is perceived for target ITD and ILDs that are farther from a natural combination. Such a relationship between the image diffuseness and the distance from the natural combination has been experimentally observed by Gaik [6].

2.6 Conclusion

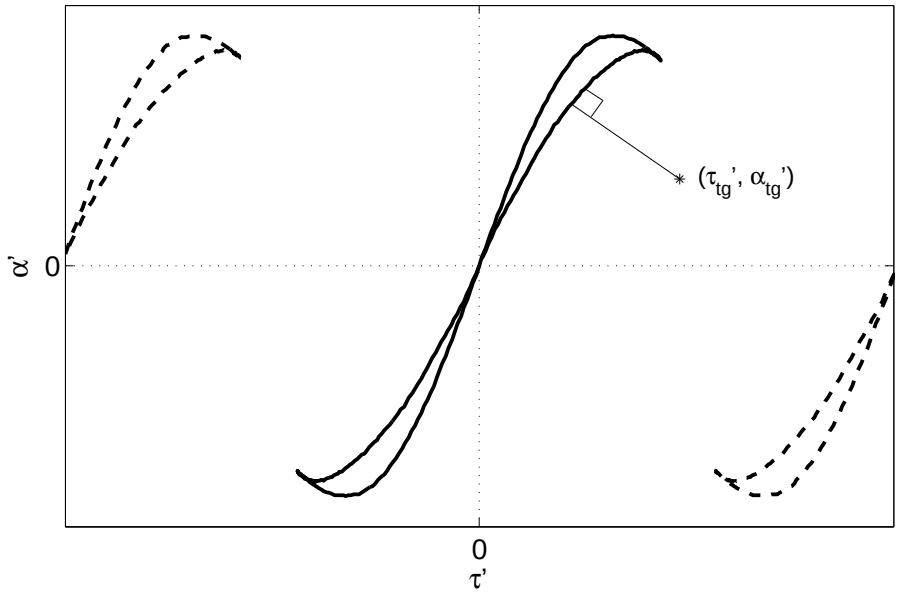
The model of spatial hearing presented in this study is mainly concerned with the auditory central processes where the source location decision is made based upon the acquired sound localisation cues. A simulation study showed that the nearest-neighbour matching technique applied in the $\tau - \alpha$ domain is simple and efficient in predicting the intracranial or extracranial location of a sound image. The comparison with relevant listening test results has given good qualitative confirmation of the feasibility of the current model.

Some of the simulation results described in this study, especially the prediction of laterality of dichotic pure tones, can be obtained with similar models of spatial hearing such as those presented by Lindemann [8] and Stern and Colburn [11]. However, it is remarkable that the current model is especially simple in dealing with ILD, unlike those which used additional weighting and mapping schemes. It is also noteworthy that this model attempts to explain the localisation and the lateralisation of sound signal within a single framework, which is the widely accepted importance of 'free-field cues' [6, 7, 9, 10, 19–21].

Despite the relatively weak support from physiological evidence, the current model is worthy of further investigation, considering the simplicity and the relatively satisfactory predictions regarding the features of human spatial hearing.



(a)



(b)

FIGURE 2.1: (a) Characteristic curves representing the natural combinations of ITD and ILD at 600 Hz obtained from a KEMAR HRTF. The solid line marked every 10° is the primary curve while the dashed lines are secondary curves. ($\diamondsuit$: $0° \sim 90°$, $\circ$: $90° \sim 180°$, $\square$: $180° \sim 270°$, $\triangle$: $270° \sim 360°$). (b) The characteristic curve is shown in $\tau' - \alpha'$ space, where the target ITD and ILD find their nearest-neighbour on the curve.

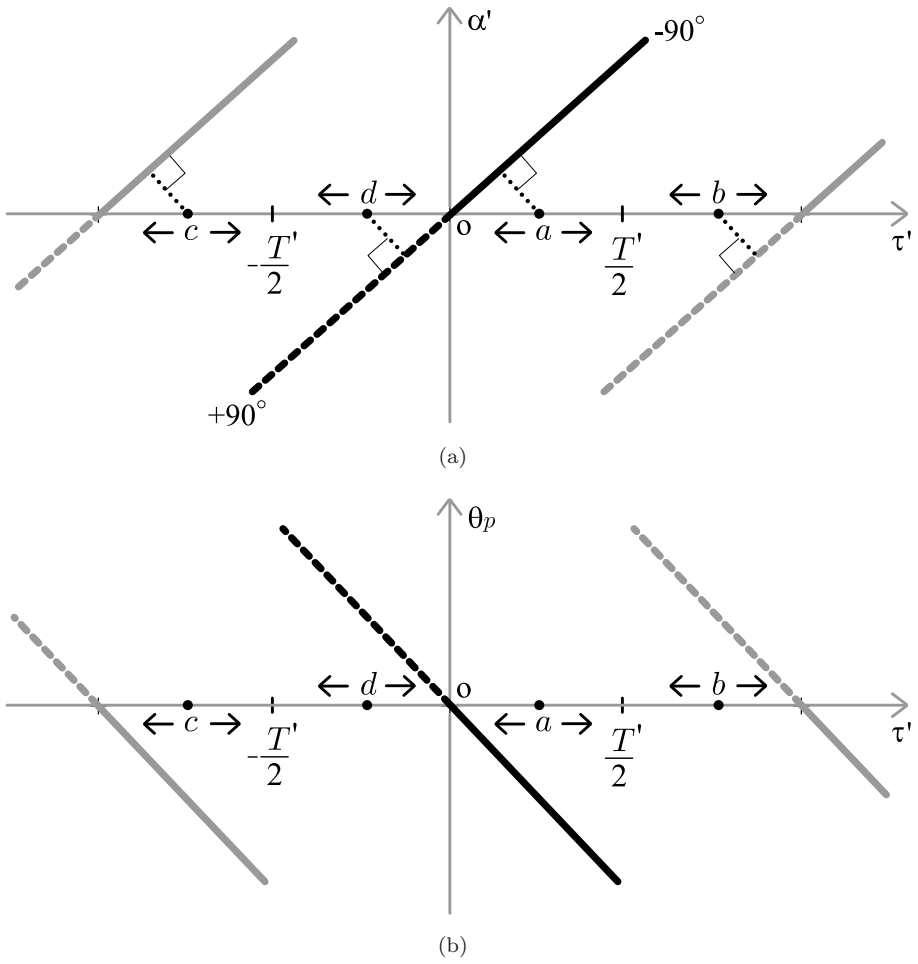


FIGURE 2.2: The procedure of nearest-neighbour matching is shown schematically. (a) Target is on the τ' axis, which means that the auditory image is created only by ITD without ILD. (The thick dashed half of the simplified characteristic curve represents the source locations in the right hemisphere whereas the solid for the left.) (b) Model prediction is plotted corresponding to the matching procedure in panel (a). Note that the laterality plot is periodic, repeating with every T'.

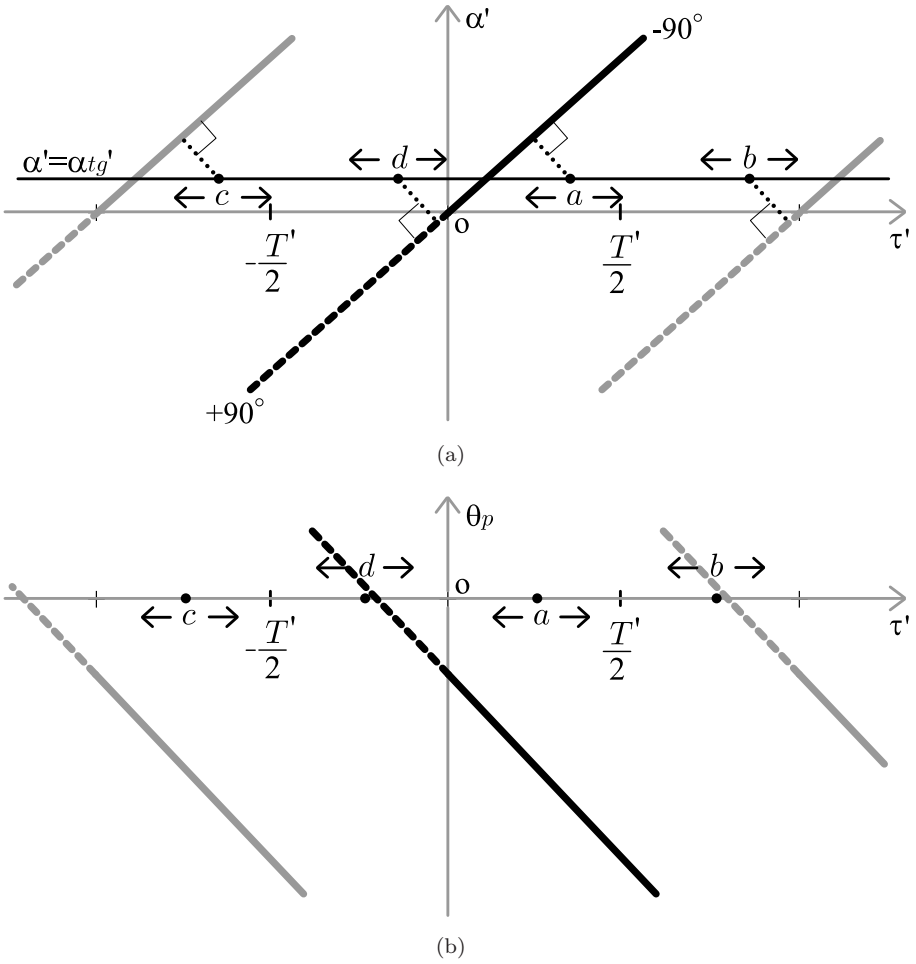


FIGURE 2.3: (a) The procedure of nearest-neighbour matching is shown schematically. (a) The target now moves on the line represented by $\alpha' = \alpha'_{tg}$. (The thick dashed half of the simplified characteristic curve represents the source locations in the right hemisphere whereas the solid for the left.) (b) Compared to panel (b) in Fig. 2.2, the laterality plot is shifted to positive ITD, images are found more often on the left side (negative azimuth) that is favoured by the given ILD. The periodicity of laterality plot is still maintained.

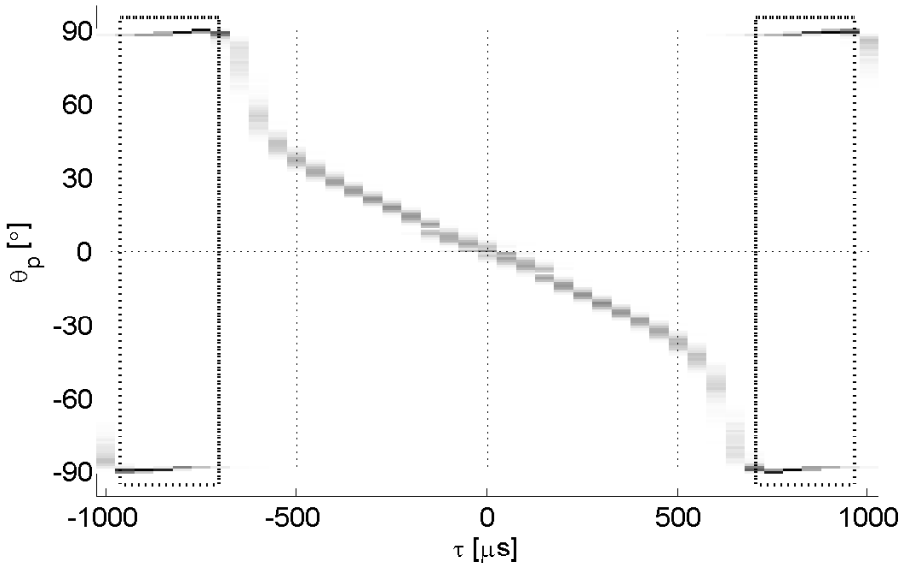
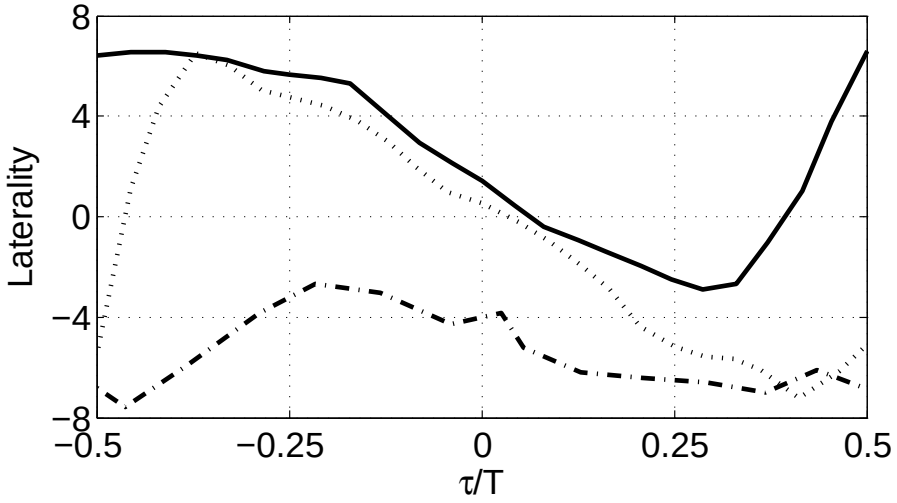
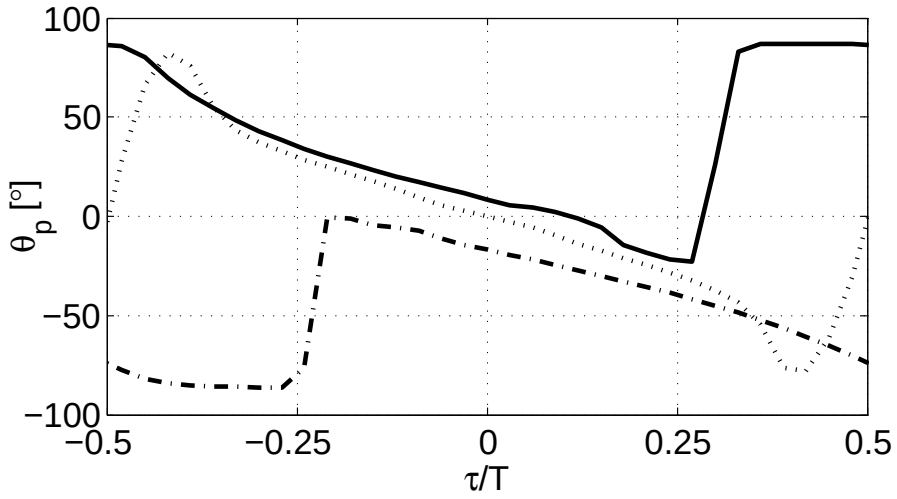


FIGURE 2.4: Predictions of the current model with internal error are shown for dichotic tone without ILD. (500 model runs with KEMAR HRTF [27] at 600 Hz) The contrast of each point represents the probability of model prediction at corresponding response angle along the vertical axis. When τ_{tg} becomes equivalent to half-period of signal, dual images are found on each far side (see the dotted boxes). It is also clear that the laterality plot will be smoothed out with internal error if the centroid is considered.



(a)



(b)

FIGURE 2.5: Laterality curves obtained for -6 dB (solid), 0 dB (dotted), and 12 dB ILD (dash-dotted) are shown where the τ axis has been scaled by the period of the signal. (600 Hz pure tone) (a) The mean values of the reported image positions are reproduced from the listening tests by Sayers [31], where the 6-dB curve has been symmetrically modified to correspond to -6 dB. (b) Laterality predictions are shown as data have been averaged for 500 model runs on the characteristic curve obtained from the KEMAR HRTF [27].

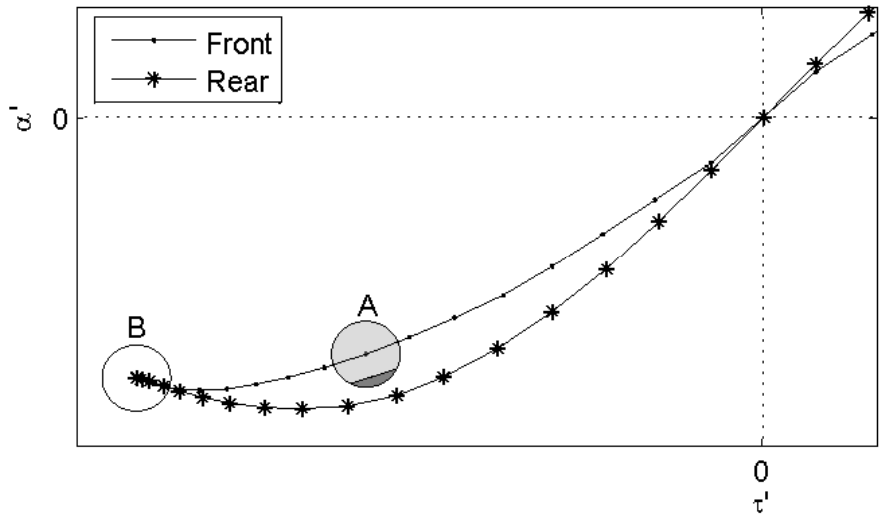


FIGURE 2.6: Source localisation by the model is illustrated schematically where the characteristic curve is marked at every 5° of the target angle. A target ITD and ILD exactly on the characteristic curve easily finds its matching azimuth estimate. However, with internal error and measurement error, it is misplaced within a certain boundary, and the azimuth estimate now becomes a random process similar to the actual listening tests. Some points within the boundary are matched to the opposite side with respect to the frontal plane due to the shape of the characteristic curve, which implies so-called 'front-back confusion' (the darker area within boundary A). Since azimuth estimates are more densely populated around ±90°, more localisation error is expected in those regions (boundary B).

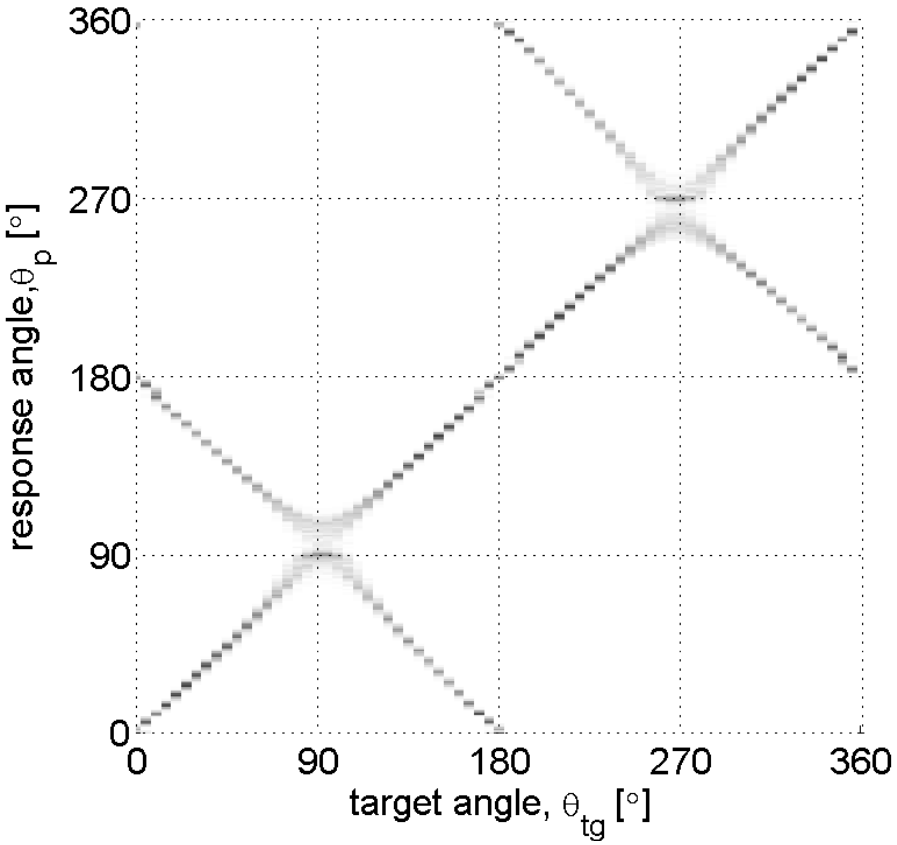
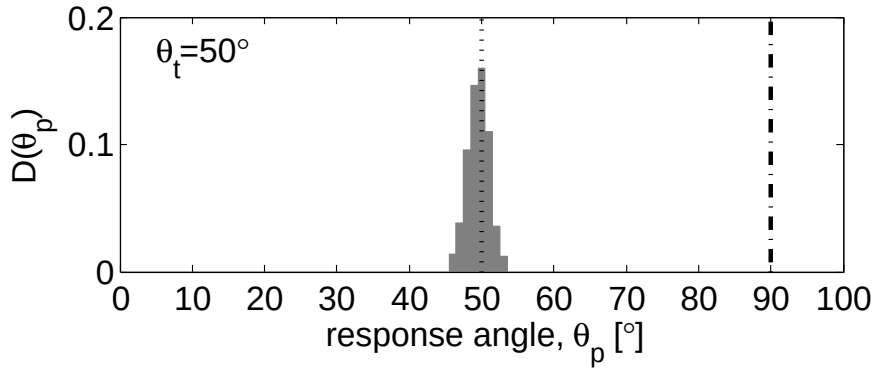
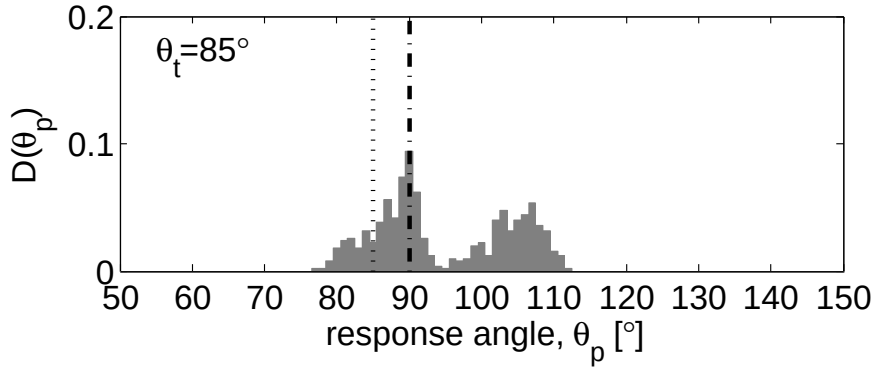


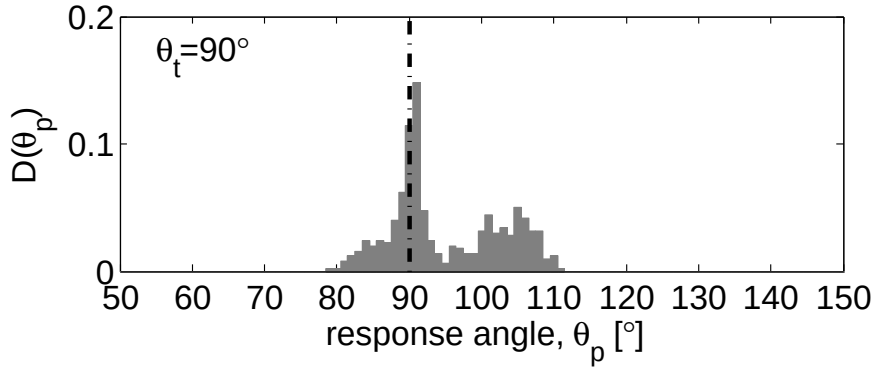
FIGURE 2.7: Model predictions made for free-field stimuli. Target angles range from 0° to 355° at every 5°, while the histogram of corresponding responses are shown with resolution of 1° along the vertical axis. Front-back confusion is clearly shown as there are two shorter legs running against the main leg that represents responses made around the exact target.



(a)



(b)



(c)

FIGURE 2.8: Individual probability functions $D(\theta_p)$ of model predictions are shown for target angles at (a) 50°, (b) 85° and (c) 90°. The dotted lines indicate the target positions while the dash-dotted line is for 90°. $D(\theta_p)$ can be regarded as frequency in each bin of histogram, normalised between 0 and 1. The peak in panel (a) is sharper and narrower than those in panels (b) and (c) which implies the more variability of model predictions for lateral angles.

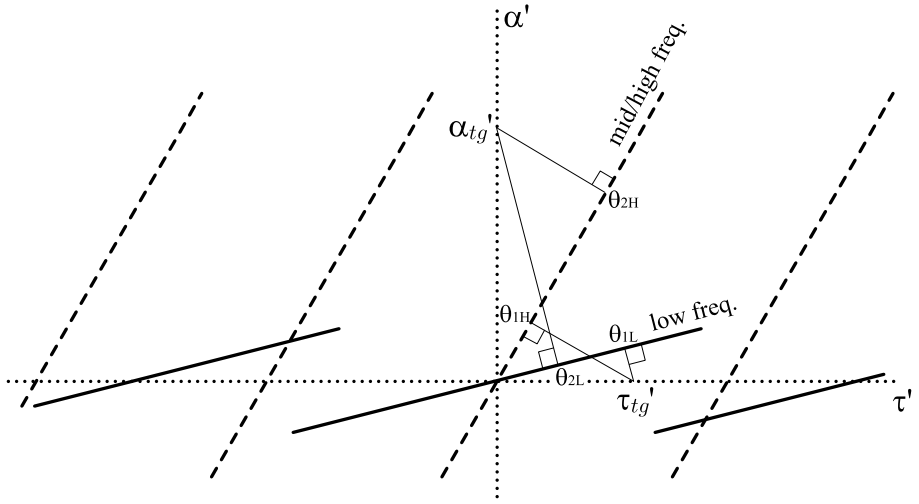


FIGURE 2.9: Simplified characteristic curves are schematically shown for different frequency ranges. The influence of ITD and ILD for each range is illustrated by estimates corresponding to target points $(\tau'_{tg}, 0)$ and $(0, \alpha'_{tg})$.

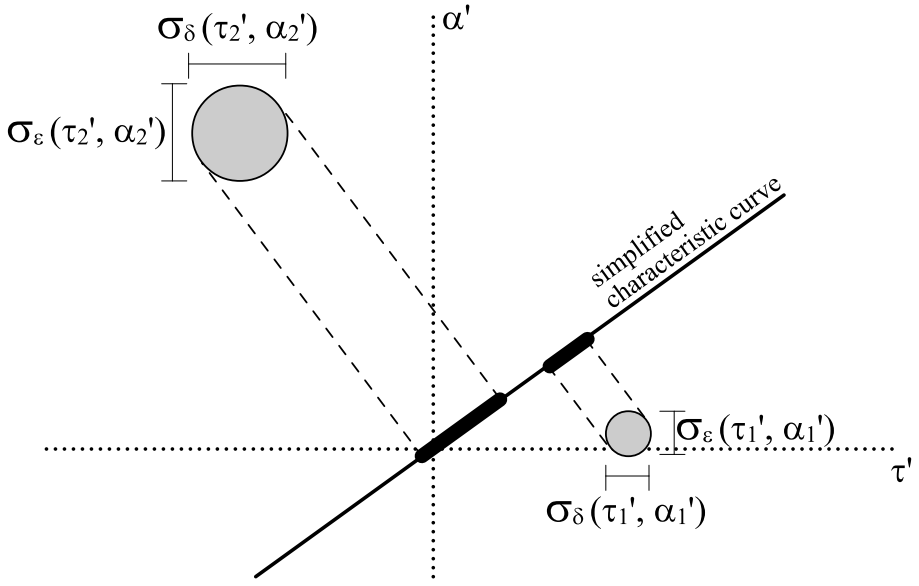


FIGURE 2.10: Assuming greater internal errors for conflicting pairs of ITD and ILD (τ'_2, α'_2), corresponding model estimates are more broadly distributed on the characteristic curve, compared to the consistent pair closer to the curve (τ'_1, α'_1). This relation between image diffuseness and the nature of ITD/ILD combination is also reported by Gaik [6] from his listening tests.

Chapter 3

HRTF measurements

3.1 Introduction

A head-related transfer function (HRTF) refers to the frequency response function from a sound source in space to the two ears [43]. The equivalent time-domain representation is referred to as a head-related impulse response (HRIR), and it contains, ideally, all of the information relating to the acoustic transmission between the source and the ears, which depends on the direction and the distance of the head relative to the source. Since the signals received at the ear drums are the primary inputs to the hearing system, knowing the HRTFs can be considered as the first step to understand how humans and animals are able to perceive the locations of acoustic stimulus. In addition to its importance in psychoacoustics and relevant hearing research, a database of the HRTFs is the essential element in the development of virtual acoustic imaging systems based on binaural technology [4]. In using this technology, subjects may perceive a realistic illusion of auditory scene if signals are presented over headphones after being processed by a pair of filters made of the HRIRs corresponding to the designated channels, directions and distances.

There have been many methods developed to model the HRTFs numerically, which include simple geometrical modelling of head and torso [44], application of the boundary element method [45, 46], an infinite-impulse-response filter approximation [47, 48], application of principal component analysis [49], the use of surface spherical harmonics [50] and so forth. Except for the first two methods in the above list, however, these techniques are generally intended to reduce the storage size of existing HRTFs, and most implementations of binaural technology still require actual measurements of the transfer functions.

In past decades, HRTFs have been measured in many different ways, and the conditions under which each measurement has been made vary depending on the purpose of the application [43, 51–54]. The type of source signal, the measuring technique and the number of source locations can all be important parameters distinguishing one measurement from another. Nevertheless, the choice of subject, whether human or acoustic manikin is regarded as the primary factor which consequently determines the equipment required and the measurement procedures. For example, in case of human subjects, the selection of microphones has to be handled with priority, as they have to be wearable. Also, positioning of the subjects becomes a very critical issue [43, 51], since they have to remain at designated positions in the initially set posture throughout the recordings, and a monitoring system as well as a backrest or a headrest may be needed, which should minimally interfere with the sound field.

The second important condition to be determined for a measurement of HRTFs is, in author's opinion, the distance between loudspeaker(s) and the subject (or the manikin). The majority of recordings have been made for relatively distant sound sources, located $1\,m$ or more from the subject, where the wavefront arriving at the subject location is assumed to be planar. Since the solid angle corresponding to the subject's head seen from the sound source is relatively small at this '**distal-region** [53],' the directivity of the loudspeaker is of less importance than other characteristics of the transducer. However, as the distance between the subject and the loudspeaker becomes smaller, the reasonable near-field characteristics and the relatively uniform directivity patterns become critical requirements for the sound source, which should approximate a point monopole [53]. In addition, the subject positioning becomes an even more important issue for this '**proximal-region**' HRTFs [53], since the measurement accuracy is more vulnerable to positioning error at shorter distances due to the increased influence of the acoustic parallax. A few representative experimental studies of HRTFs have been summarised in Table 3.1 according to the conditions and specifications of the measurements.

In the following sections, measurement of HRTF databases will be presented. The intention of this experimental study is to provide individualised hearing models (see chapters 2 and 5) based on the HRTF database measured for the participants of the listening tests to be reported in the later chapters. Therefore, recordings will be made for source locations at every 5° azimuth angle, only on the horizontal plane. Furthermore, the current measurement will be a 'blocked-ear-canal' measurement where the microphones bundled with spongy ear plugs will be inserted into the meatus. Regarding the issue of positioning, an automated voice-feedback system aided by the electromagnetic head-tracking device has been established, which will guide subjects throughout the measurements. The distance between the subject and the loudspeaker has been designed to be $1.5\,m$ and $0.3\,m$, for the distal- and the proximal-region measurements, respectively. The

	Gardner [52]	**Moller** [43]	**Algazi** [51]	**IRCAM** [54]	**Brungart** [53]	**Mannerheim** [27]	**Cho** [55]	**Current**
Subject(no.)	KEMAR	Human(40)	Human(43) & KEMAR	Human(51)	KEMAR	KEMAR	KEMAR	Human(6)
Distance, m	1.4	2	1	Not known	0.125∼1	2	0.5	0.3 & 1.5
Ear canal type	Open	Open & blocked	Blocked	Blocked	Open	Open & blocked	Open & blocked	Blocked
No. of source locations	710	97	1250	187	Not known	1008	1008	72
Resolution in Elev.	10°	22.5°	5.625°	15°	Not known	10°	10°	n/a
Range in Elev.	-40° ∼90°	-90° ∼90°	-45° ∼230.625°	-40° ∼90°	Not known	-40° ∼90°	-40° ∼90°	n/a
Resolution in Azim.	5°	∼22.5°	∼5° (-90° ∼90°)	∼15°	Not known	5°	5°	5°
Sampling rate	44.1 kHz	48 kHz	44.1 kHz	44.1 kHz	Not known	48 kHz	48 kHz	48 kHz
Input signal	MLS	MLS	Golay-code	Sweep	Periodic sweep	Pink noise	Pink noise	Pink noise
SNR	65 dB	70 dB	Not known	Not known	Not known	50∼70 dB	53 dB	Not known
Head positioning	n/a	Paper marker on top of subject's head displayed on the monitor.	Same as Moller	Headrest and head-tracker monitoring	n/a	n/a	n/a	Voice-feedback aided by head-tracker
Microphone	Etymotic ER-11, Neumann KMi 84	Sehnheiser KE 4-211-2, B&K 4182/4136	Etymotic ER-7C	Knowles FG3329	B&K 4134/4165	B&K 4132/4136	B&K 4134/4136	Panasonic WM-60A
Loudspeaker	Realistic Optimum pro 7	Vifa M10MD-39	Bose Acoustimass	Tannoy system 600	Custom-made	Custom-made	Micro-speaker (Bujoen, BMS-1709SL08C)	Celestion AVC102, Micro-speaker (Bujoen, BMS-1709SL08C)
Data processing	Minimum phase inverse filtering	20 kHz low-pass filter	Hanning window	Not known	Frequency domain measurements from 100 Hz to 19.2 kHz	Rectangular window & inverse filter	Rectangular window & inverse filter	Hanning window & minimum phase inverse filtering (distal-region only)

TABLE 3.1: A few representative measurements of the head-related transfer functions reported in the literature.

distal-region HRTFs are expected to be the primary data which will be used to build individualised decision-making models for each subject, but the models based on the proximal-region HRTFs will be also considered for comparison purposes in future work.

Section 3.2 describes the details of measurement design and procedures with the specifications of equipment used, where the characteristics of the transducers and the interference of the equipment with the sound field will be investigated in particular. In section 3.3, HRTFs and HRIRs at a few representative source locations will be presented, following the detailed account for the post-processing of the raw data. In addition, the procedures to acquire ITDs and ILDs at a single frequency will be briefly introduced, and the discussion on the obtained characteristic curves will be also made. In section 3.5, possible measurement error, mainly the positioning error, will be investigated, and the simulation results will be compared with the experimental data. Finally, section 3.6 will present some conclusion for this chapter.

3.2 Measurement

The HRTF measurement has been carried out in the large anechoic chamber at the Institute of Sound and Vibration Research (ISVR), University of Southampton, which measures $9.15\,m \times 9.15\,m \times 7.32\,m$ with the lower cut-off frequency at approximately 80 Hz. Subjects were sitting on a chair which incrementally rotates (see Figs. 3.1 and 3.2), and on each rotation, binaural signals transmitted from a single loudspeaker to the subject's ears have been recorded simultaneously by in-ear microphones. The following subsections will describe the details of the measurement procedure and arrangement.

3.2.1 Measurement specifications and equipment

The choice of microphone is one of the important issues in the HRTF measurement. The overall size of the transducer should be small in both length and diameter, since the microphone has to be safely inserted but deep enough for the diaphragm to be positioned flush to the concha. Also, the microphone response has to be acceptable in the frequency range under investigation, and the reliability and the cost are also of general concern. Among many candidates, the Panasonic WM-60A omnidirectional electret condenser microphone has been selected, which is 6.0 *mm* in diameter and 5.0 *mm* in length [see Fig. 3.3(a)]. At a very low cost per unit, it is reasonable in size and characteristic response, which will be shown later in this section.

Custom treatments were necessary to make the microphones wearable [see Figs. 3.3(b) through (d)]. Very thin wires have been used for cabling to make sure that they may interfere as little as possible with the sound field in the vicinity of the ears. In addition, disposable earplugs designed for clinical purposes have been customised to completely surround the microphone periphery so that the subjects may feel comfortable, and at the same time, the ear canals may be completely blocked in accordance with the measurement design. Finally, a safety string was attached to the microphone in order to make sure a safe and easy removal of the insert after experiment.

Two types of loudspeakers have been employed depending on the distance from the subject to the sound source. In case of the distal-region HRTF measurement, the sound wave incident on the head of a subject is assumed to be planar, and it is more important to ensure a high quality frequency response of the loudspeaker than other spatial characteristics such as the directivity. The selected loudspeaker for the distance of 1.5 *m* was Celestion AVC102 [see Fig. 3.4(a)], the dimension of which is 15(L)×20(H)×9(W) in *cm*. On the contrary, for the proximal-region, which is 30-*cm* distance in this study, a loudspeaker to approximate a point monopole source is required, giving a wider beamwidth

with a relatively small dimension [53]. Among the limited selections, a micro-speaker unit (BMS-1709SL08C manufactured by Bujeon Co. Ltd., Korea) has been chosen, which was originally designed for mobile phone handsets. In the previous measurement of proximal HRTFs (at 50 *cm*) using KEMAR [55], it has been shown that this unit in a custom-made plastic cabinet measuring 4(L)×4(H)×1.7(W) in *cm* [see Fig. 3.4(b)] has reasonable frequency and time responses.

As for the signal amplifiers, a custom-built 4-channel device has been used for the microphones, while a YAMAHA H5000 power amplifier has been employed for the loudspeaker. The Huron 2.0 Digital Audio Convolution Workstation (Lake Technology) both generated and captured signals to the loudspeakers and from the microphones, respectively. A 8192-sample pink noise (frozen) was used for the source signal, and the raw 2-channel signals from microphones were captured and saved at 48-kHz sampling frequency. All operations involved in the measurement including the movement of the motorised chair were controlled by Matlab 7.0 (The Mathworks, Inc.). See Fig. 3.1 for the arrangement of equipment in the anechoic chamber and the control room.

The characteristics of the above equipment have been measured and plotted in Figs. 3.5 through 3.7 where the responses with Celestion AVC102 ('Celestion' from this point) are shown in panels (a) while those with the micro-speaker ('Bujeon' from this point) in panels (b). From Fig. 3.5 where the windowed free-field responses are presented in the frequency domain, the upper cut-off frequency appears to be roughly 10 kHz for both loudspeakers. (Note that there is a dip in the Celestion at around 13.5 kHz.) The response for the Celestion is observed to be relatively flat below this boundary, and in the low frequency range it rolls off slowly even down to 100 Hz. However, the response of Bujeon rapidly decreases below 1 kHz, reaching its floor response at around 300 Hz.

Fig. 3.6 shows directivity patterns for the two loudspeakers. It is remarkable that Bujeon maintains a uniform directivity pattern throughout the frequency range of interest. On the contrary, the pattern for Celestion starts to deviate above 1 kHz, and becomes very irregular at high frequencies above 8 kHz. It is noteworthy that the measurement conditions for the two loudspeakers were slightly different: The sound field created by the Bujeon was sampled at every 15° at 30-*cm* distance which is the same condition to be configured for this loudspeaker in actual measurements. Meanwhile, the Celestion was sampled at every 5° at 80 *cm*, which is shorter than the designated distance (150 *cm*). These differences in the measurement conditions were due to some technical difficulties, but the above discussed results appear to reasonably reflect the device characteristics in actual measurements.

Having compared the directivity patterns of the two loudspeakers, Fig. 3.7 shows 3-dB beamwidth giving the estimate of angular ranges across frequency within which responses

deviate no more than 3 dB in magnitude compared to the response at 0°. It is obvious that Bujeon has a wider beamwidth of approximately 90° or greater from about 500 Hz up to 15 kHz. The Celestion also has a good angular range of uniform response at low frequencies, but, as frequency increases, it quickly reduces to under 90° and appears to taper off above 7 kHz.

Despite the individual characteristics discussed above for the two transducers, both loudspeakers may be regarded as suitable for the measurement at their designated distances. (Note that the required angular ranges for 30-*cm* and 150-*cm* measurements are only about 40° and 8°, respectively, considering the 'aperture angle' corresponding to the normal size of human head.) After all, the above discussion was not only about the loudspeakers but also about the frequency responses of all the equipment employed, and they are expected to provide reliable output signals in a relatively wide frequency range up to about 10 kHz, which is very reasonable with the small low-cost microphones.

Apart from the equipment directly employed for the task of recording, there were facilities to help the subject maintain the correct position and direction. Firstly, a backrest attached to the custom-made seat supported subject's upper body. Meanwhile, a head-tracking device has been used to monitor the position of the subject's head during the measurement (Refer to section 3.2.2 for the procedure). This head-tracker, Polhemus FASTRAK, consists of an electromagnetic wave transmitter and a receiver (see Fig. 3.8), giving relative position and direction to a high level of accuracy. In the current measurement, the receiver was attached to a thin piece of soft plastic linked to a flexible headband which was worn by the subject on the head [see Figs. 3.8(b) and (c)]. The transmitter was located beneath the custom-made seat so that it may function as the origin of the coordinate system which rotates along with the chair (see Fig. 3.1).

Although the size of the backrest and the plastic piece for the head-tracker have been minimised, these facilities attached relatively close to the microphones may interfere with the sound field. Therefore, some preliminary measurements have been made with a KEMAR to show the influence of those devices. The distance between the KEMAR and the loudspeaker (Celestion AVC102) was 150 *cm* and impulse responses were measured at every 5° by Ear Simulator RA0045 (GRAS). The recorded impulse responses were post-processed with a 200-point Hanning window applied around the peaks, while the first 250 samples were zeroed. Measurements have been made with and without the headband and/or the backrest, and the frequency responses were compared in terms of the *log* magnitude with reference to the recordings made without any device worn or attached.

Fig. 3.9 shows the result of the comparison for the left channel where the grey-scale level indicates the degree of deviation from the reference response; white for less than

1 dB, greys from 1 dB to 3 dB, black for more than 3 dB. With the headband worn, it is observed in panel (a) that response deviates mostly in the high frequency range above 5 kHz, since the sound waves at low frequencies with longer wavelengths are readily diffracted around small objects. It is also interesting to note that the deviation is found more at the angular locations where the corresponding ear is shadowed (remember that 90° indicates the sound source to the subject's right). It is tentatively attributed to the weak signal strength in this region, the signal thus being more vulnerable to interference.

On the other hand, the backrest attached to the seat appears to have little influence on the sound field [see Fig. 3.9(b)], perhaps because it is relatively distant from the measurement positions with its size probably insignificant compared to the nearby objects, the torso of the KEMAR or the subject. Fig. 3.9(c) shows the combined effect, from which the measurements with both headband and backrest attached to the subject appear to be reliable only up to 5 kHz. Considering, however, the presence of sound absorptive materials found with ordinary subjects such as hair and clothes, the deviation observed above in the frequency response is expected to reduce in actual experiments, thus possibly extending the effective frequency range.

3.2.2 Measurement procedure

A total of 6 paid subjects participated in the measurements. They are all male in their 30's and late 20's. Before the experiments, subjects' ear canals were examined by using an otoscope in order to make sure that there is no obstructive material, e.g. ear wax, inside the ears to the depth to which the microphones will be inserted. This experimental study has been approved by the Safety and Ethics Committee of the Institute of Sound and Vibration Research (ISVR), University of Southampton (Approval number: 777).

The geometrical configuration of the subject on the platform and the loudspeaker was of most concern, since the reliability of the recorded data depends on how accurately a subject is initially positioned relative to the sound source and how accurately that position is kept during each recording. It was also important to make sure that subjects feel comfortable during the experiment which took approximately 40 minutes up to 1 hour. For the latter requirement, the chair was equipped with a backrest on which subjects can sit back. Furthermore, the head-tracker played an important role, monitoring the position and the direction of subject's head to update the voice-feedback system.

The actual measurement procedure started with positioning the subject's head so that the midpoint of the interaural axis was aligned with the axis of chair rotation, and the centre of the loudspeaker was also aligned with the level of the subject's ears. For this, a laser level was used from the subject's left side for a visual alignment (see Fig. 3.10).

Then, the head-tracker recorded this initial position and direction of the subject's head as a reference. The last step of the alignment was to make sure that the loudspeaker is actually located at 0° relative to the subject's head. The loudspeaker position giving a zero ITD has been regarded as a practical 0° reference, and a few preliminary measurements of binaural impulse responses followed by the experimenter's feedback to the motorised turntable could achieve this to within a resolution of 1°, which is the programmed resolution of the step motor.

Once the measurement started from 0°, the head-tracker continuously read the subject's position, based on which one of the automated voice-feedbacks has been played over the loudspeaker to help the subject correct his/her posture. There were 10 types of voice messages recorded for the 5 degrees of freedom as shown in Fig. 3.11, and the preliminary test of this automated voice-feedback positioning system showed that subjects can keep and get back to the initial position and direction within a reasonable time period ($\leq$ 30 sec.), when the tolerances for each translational and rotational degree of freedom are $\pm 0.5\,cm$ and $\pm 0.75°$, respectively. These values for each tolerance appear to give the maximum possible accuracy within the limited measurement time, where the repositioning task turned out to be relatively time-consuming with less tolerances.

Subjects quickly learned to react to the given voice-feedback, and could easily maintain the reference posture after a few trials. The recording procedure has been automatically triggered when the subject's head came within the tolerance, and the time taken for a single measurement was about 1 second. On completion of each recording, the subject's head position and direction were monitored once again, and, if out of tolerance, the data recorded for that trial were discarded and recorded again. Two successful recordings were made for each azimuth angle at every 5°, which took 40 minutes to 1 hour for the whole measurement of 360°, depending on the subject. A break was given every 10–20 minutes, and the distal- and the proximal-region HRTFs were measured in separate sessions on different days.

Finally, the free-field measurement was made without the subject in position, where reflective surfaces of the chair and the platform were covered with sound-absorptive wedges as effectively as possible. The impulse responses were recorded at the two different distances, when the pair of microphones used in the measurements were positioned very close to each other. These individual free-field responses will be used for post-processing separately in each channel (see Fig. 3.5 for the free-field responses in the left channel).

3.3 Data processing and results

The 8192-sample impulse responses between loudspeaker and in-ear microphones can be further processed and equalised with respect to the free-field responses. As for the distal-region measurements, minimum phase inverse filters have been obtained from the free-field responses, which were then applied to the measured HRIRs [52]. In order to acquire the inverse filters, the left- and right-channel free-field responses are first windowed in the time domain by 200-point Hanning windows, the maximum of which are aligned with the absolute peaks of the responses. It is further observed that the responses before the 250th sample can be zeroed, which contain no meaningful data. These windowing and zeroing processes can disregard the unnecessarily long tails of the impulse responses, thus suppressing unwanted noise and reflection [see Fig. 3.12(a)]. Then, a 8192-point fast Fourier transform (FFT) is applied to the impulse responses to give magnitude and phase responses in the frequency domain. In order to prevent the final inverse filter from having a 'ringing tail' due to any excessively low amplitude in the high frequency range, the magnitude responses over 9.5 kHz have been flattened as depicted by the dashed lines in Fig. 3.12(b) [52]. Considering that the effective frequency range of the measurement is already limited by the microphone and loudspeaker responses, this equalisation process does not significantly influence the reliability of the measurement any further. The modified magnitude responses are recombined with the corresponding phase responses, and these frequency responses are inverted, inverse-Fourier-transformed, and FFT-shifted. As shown in Fig. 3.12(c), the inverse filters at this stage are mixed-phased with non-causal responses. Finally, minimum phase inverse filters are acquired by taking real cepstra using the *rceps* function in Matlab 7.0 [see Fig. 3.12(d)].

The raw recordings of HRIRs are also windowed with 200-point Hanning windows and zeroed in the same way that the free-field responses are processed. These treated HRIRs are then convolved with the inverse filters acquired above. Finally, the data sequences from 200th to 455th points are only taken as 256-point equalised HRIRs.

In contrast to the post-processing for the distal-region data, it has not been possible to acquire usable inverse filters from the proximal-region free-field responses due to the limited and unreliable transducer responses at low frequencies. Therefore, no further process has been implemented in time-domain except that both HRIRs and free-field responses were windowed (200-sample Hanning window as above) and zeroed (until the 85th sample). On the other hand, in the frequency domain, the HRTFs obtained by FFT have been equalised by the free-field responses only in the magnitude responses [53].

Fig. 3.13 shows the distal-region HRIRs of subject SF at a few representative azimuth angles before and after the post-processing. In general, the impulse responses presented in this figure contain some known features of directional transfer functions: the greater interchannel differences in the attack times and the peak amplitudes at lateral angles and the well-aligned and almost identical responses at 0° [27, 52, 55]. It is also observed that, after the equalisation, unwanted reflections and high-frequency noises have been relatively well controlled to result in smoother impulse responses.

The distal-region frequency domain responses shown in Fig. 3.14 can give clearer pictures of the impact of the post-processes including the free-field equalisation. (In Fig. 3.14, responses in full 8192-samples rather than the 256-sample truncated version are shown for discussion purpose.) First of all, the Hanning windows applied to both the free-field responses and the raw HRIRs effectively removed the high-frequency variability, particularly from the contralateral channels as shown in panels (c) through (d). In addition, flat and smooth responses at low frequencies have been also achieved, which can be directly attributed to the free-field equalisation. In both frequency responses before and after the post-processing, some well-known features of HRTFs can be clearly observed, which include the pinna notch at about 9 kHz [panels (a) and (b)] and the greater interaural level difference at higher frequencies [panels (c) through (f)] [27, 52, 55].

It is known that equalised HRTFs at 0° converge approximately to 0 dB at very low frequencies, since the presence of the human head hardly affects the sound field in this range, thus giving responses nearly identical to the free-field responses [27]. However, in the current data shown in Figs. 3.14(b) [and 3.16(b)], such a convergence is not always observed, which perhaps resulted from the less satisfactory microphone responses in the very low frequency range. Considering this limited reliability at low frequencies particularly below 100 Hz, further post-processing such as bandpass filtering can be carried out depending on the nature of actual application.

In the time domain, it is difficult to observe differences between the distal- and the proximal-region HRIRs when the responses in Fig. 3.15 are compared to those in Fig. 3.13. (It is recalled that there is no equalised data for the proximal-region due to the absence of the inverse filtering process.) The time-domain features mentioned above for the distal-region HRIRs are also found in the proximal-region data in terms of the interchannel differences in attack times and amplitudes. However, in the frequency domain a clear contrast can be made as shown in Fig. 3.16, where the greater interaural level difference is observed for the proximal-region HRTFs than the distal-region. The increased ILD in the HRTFs measured at shorter distance is commonly reported in similar studies [53, 55], which have been explained well both in theory and numerical

simulations by the emphasised head-shadowing in the near-field. The difference between the distal- and the proximal-region HRTFs will be further discussed in relation to the characteristic curve in the following sections. Finally, it is noteworthy that HRIRs and HRTFs of the participants other than the subject SF were in common in showing the above discussed features, although they will not be presented here in detail.

3.4 Computation of characteristic curves

Having found that the HRTFs acquired in the current measurement are qualitatively comparable to those reported in the literature, the acquired HRIRs can be further processed to give characteristic curves. In order to obtain ITDs [see Fig. 3.17(a)], a 100-ms pure tone signal at frequency, f is first modulated at an envelope frequency of 20 Hz and zero-padded to give a target signal. Then, this signal is convolved with the post-processed HRIRs for a certain azimuth angle, producing synthesised binaural signals. (Post-processed HRIRs indicate the equalised HRIRs in case of the distal-region data, but the windowed HRIRs for the proximal-region.) The resolution of the final ITD depends on the sampling frequency, and the binaural signals can be oversampled at a higher sampling frequency which has been shown to give a smoother ITD curve. The peak of the cross-correlation function can be found for these interpolated signals to give the ITD, where it is necessary to correct the quasi-periodic ITDs by adding or subtracting multiple numbers of signal periods. On the other hand, ILD can be obtained simply by comparing the magnitude responses of the HRTFs at the designated frequency f as shown in Fig. 3.17(b). The ITDs and ILDs at this stage are true values reflecting the shape of subject's head and torso and the distance from the loudspeaker. However, there can be a few data points away from the expected 'trajectory' of each interaural disparity, possibly due to the measurement error or the tolerance of the positioning error. Therefore at the final stage, a curve fitting process has been additionally carried out to find smooth functions for the ITDs and the ILDs [see the last processes in Figs. 3.17(a) and (b)].

Figs. 3.18(a) and (b) show ITDs and ILDs at 600 Hz obtained from the distal-region HRIRs (Subject SF) where raw ITDs and ILDs have been fitted with polynomials at the order of 11 (using Matlab 7.0 built-in functions, *polyfit* and *polyval*). The order of curve-fitting has been set relatively high in order to make sure that no significant curve shape is lost. It is obvious that the features of the ITD and the ILD functions have been well preserved while irregular data points especially at lateral angles have been smoothed out. As expected from the average values found in previous studies in the literature [27], the ITD ranges from $\sim -800\,\mu s$ to $\sim +800\,\mu s$ [40], and the ILD from ~ -7.5 dB to $\sim +7.5$ dB.

Combining ITD and ILD functions in Figs. 3.18(a) and (b) can give a characteristic curve shown in Fig. 3.18(c) where it has been marked at every 10° of azimuth angle. Features discussed in section 2.2 can be found. Firstly, the two legs of the curve representing the frontal and the rear areas are not overlapped but are distinctive from each other, which implies that source localisation in the horizontal plane is even possible based only on the ITD and ILD information, but can be vulnerable to front-back confusion if there are

errors or noise in processing the binaural input signals. It is also apparent that sound sources at lateral angles can give similar combinations of ITD and ILD as there are more marks around the turning points of the characteristic curve in Fig. 3.18(c), which has been related to the more localisation error for these source positions in section 2.4.

The characteristic curve shown in Fig. 3.18(c) is reasonably symmetric with respect to the origin where ITD and ILD are zero. However, it is noteworthy that the ITDs and the ILDs are not necessarily identical for sound sources at 0° and 180°, which can be attributed partly to the asymmetry in the shape of head, but also to the random error in positioning the subject by the head-tracking device. (Remember that there were, inevitably, certain tolerances allowed for translational and rotational degrees of freedom. See section 3.2.2.) The characteristic curves obtained for all other subjects have similar features to those as shown in Fig. 3.18(d), where the mismatch between 0° and 180° can be found in most of the curves, and, depending on the subject, the shapes of the left and right 'lobes' of the characteristic curves have been found to be different from each other, again, due to the left-right asymmetry of the head. A detailed inspection of Fig. 3.18(d) illustrates that the width of the lobes and the degree of left-right asymmetry in individual curves may vary from subject to subject, and it is reasonable to say that the characteristic curve is as unique for each subject as the individual HRTFs at a single frequency.

In contrast to the distal-region results presented above, ITDs and ILDs obtained from the proximal-region HRIRs have been found to be mostly asymmetric with respect to the median plane as shown for the 600-Hz pure tone signal in Figs. 3.19(a) and (b) (subject SF). (The order of curve-fitting is 11 as was the case for the distal-region.) Although ITD is relatively close to $0\,\mu s$ at 0° and 360 ° in Fig. 3.19(a) thanks to the initial alignment procedure inspecting the arrival times of the signals (see section 3.2.2), it is about $200\,\mu s$ at 180°, far away from its 'home' position, when the subject faces backward. This mismatch severely disrupted the symmetry of the ITD curve, broadening the positive peak at around 270°. A similar observation can be made for the ILD function shown in Fig. 3.19(b), and it appears that there have been some systematic errors in the measurement which became more prominent in case of the proximal-region. As a result, the characteristic curve shown in Fig. 3.19(c) is significantly distorted, and particularly, the shapes of the two turning points approximately at 90° and 270° appear to be very different from each other. This distortion of the characteristic curve has been also found in the results for other subjects as illustrated in Fig. 3.19(d), which makes it difficult to determine whether the proximal-region characteristic curves are uniquely shaped for each person. The errors responsible for the distorted characteristic curves will be discussed in more detail in section 3.5.

Apart from the more prominent asymmetry resulting in the distorted curves, the proximal-region results are also distinguished from the distal-region data by the greater range of ILD. As clearly depicted in Fig. 3.19(b), the maximum of the absolute level difference is now about 12 dB, greater approximately by 5 dB than that for the distal-region. Attributed to the increased influence of the head-shadowing at a shorter distance, this expanded ILD range is also reported in the literature [53, 55].

3.5 Analysis of the positioning errors

Even though the accurate geometrical configuration of both transducer and subject has been regarded as the most important issue in the design of the current measurement, it is unlikely that relative positioning has been perfectly maintained throughout the experiment. In particular, regarding the subject's self-positioning procedure facilitated by the head-tracking device, random errors can be introduced into some or all of the 5 degrees of freedom within the given tolerances (see Fig. 3.11). In addition, other factors in the design of the measurement might possibly result in certain types of systematic errors. For example, the accuracy in the initial positioning procedure or the build-quality of the platform can be associated issues.

In Fig. 3.20, the two points marked by L and R indicate ideal positions of ears, and S indicates the location of sound source at 0°. Assuming a free-field propagation, ITD can be computed by considering the difference between the path lengths from the source to the left and right ears, when the source location changes from 0° to 360° (or when the subject rotates with respect to the origin). The simulated ITD is slightly less than the actual ITD (approximately by $200\,\mu s$) due to the absence of the head, but can be a reasonable estimate. Having obtained an estimate of the ITD at the ideal centre position, the subject's head can now be assumed to be initially misplaced by Δx and rotated by $\Delta\theta$ (see Fig. 3.20), and the interaural time difference in this case can be recalculated to show the influence of the mispositioning. (Only the misplacement in the lateral direction, Δx and the azimuthal rotation, $\Delta\theta$ will be considered for the purpose of presentation.) The deviation of the simulated ITDs with respect to the values obtained at the ideal position has been computed and plotted in Figs. 3.21(a) and (b) for the distal-region and the proximal-region measurements, respectively. Parameters of $\Delta x = +0.5\,cm$ (misplacement to the right) and $\Delta\theta = -0.75°$ (head turning to the right) have been used, which are the maximum tolerance for each direction, resulting in deviation in a consistent way. As the dashed and the dash-dotted lines in Fig. 3.21 indicate deviations in the ITD introduced by non-zero Δx and $\Delta\theta$, respectively, it is noteworthy that the translational misplacement gives a greater error for the proximal-region than for the distal-region measurement, whereas the influence of the rotational misorientation is invariant to the distance. Considering that the maximum absolute error is about $10\,\mu s$ and $17\,\mu s$ for the distal- and the proximal-region, respectively, it is also apparent that the proximal-region HRIR measurement is more vulnerable to head-movement, which is entirely attributed to the increased errors caused by the translational movement. Since the other extreme movement where $\Delta x = -0.5\,cm$ (misplacement to the left) and $\Delta\theta = +0.75°$ (head turning to the left) will induce the same magnitude of error but in the opposite direction, the range of ITD variation within the given tolerance

can be double the figures suggested above, and even greater variation may be observed if possible positioning errors in all the 5 degrees of freedom are accounted for.

Even though significant efforts have been made with the aid of a laser level to control the subject's initial position in front-back and up-down directions, there was no device to correctly position the subject in the left-right direction, where they were only instructed to sit at the centre of the seat by estimating the unoccupied widths at the sides. For example, the subject could be initially misplaced in the positive lateral direction by Δx as shown in Fig. 3.22. Consequently, the new ear positions L$'$ and R$'$ become off the left-right axis in the next step of the 0-degree alignment where ITD is controlled to be approximately zero. So, during the measurement for all azimuth angles, the new, misplaced positions for the left and right ears, L$'$ and R$'$ will draw non-overlapping circular trajectories with respect to the desired centre of head, O, which is the axis of rotation (see the dashed circular paths in Fig. 3.22).

The path lengths from the source S to the misplaced ear positions L$'$ and R$'$ can be computed for different Δx's in the range from $0\,cm$ to $5\,cm$ with the source location varying from 0° to 360°. Then, both ITD and ILD can be obtained from the path length differences for the distal- ($d = 1.5\,m$) and the proximal-region cases ($d = 0.3\,m$), while, particularly, the ILD can be computed by assuming the sound pressure inversely proportional to the path length. The results are shown in Fig. 3.23 where the darker lines indicate a larger Δx that is a displacement to the right. From panel (a), it is obvious that the distal-region measurement is very robust to the initial positioning errors, as the ITD function has been hardly influenced. However, there are significant deviations for the proximal-region measurement as illustrated in panel (b), which are particularly prominent for the target angles in the rear hemisphere. When the source is located in these positions around 180°, positive Δx causes the signal to arrive to the left ear earlier, thus increasing ITD (remember that positive ITD indicates earlier arrival to the left ear), which, on the other hand, would be reduced if Δx were negative. By comparing Figs. 3.23(c) and (d), a similar argument based on the path length difference can be made for the ILD. Whereas the influence of the positioning error on the interaural level difference is insignificant for the distal-region measurement [panel (c)], the proximal-region data have been found to be very vulnerable to the lateral misplacement in the initial positioning procedure [panel (d)]. In particular, the deviation in ILD is not only observed for the target angles around 180° but across all range of source locations, systematically distorting even the maximum and the minimum, respectively at around 90° and 270° as denoted by the arrows. In fact, the errors made during the initial positioning procedure influence the travelling distances for binaural signals, and therefore, both absolute and relative sound levels and signal arrival times are affected [e.g. compare attack times in

Figs. 3.15(c) and (d)], but further discussion on this issue appears to be beyond the scope of this study.

It is interesting to see that the distortion of ITDs and ILDs observed in Fig. 3.19 for the actual proximal-region data are very similar to the simulation results presented above. Although the simulated ITDs and ILDs are slightly lower in an absolute sense for the lack of consideration of the subject's head, the increase in ITDs as well as in ILDs for target angles in the rear hemisphere is comparable between the simulation and the measurement results. In addition, the above presented error analysis suggests that all subjects in the actual experiments could have been seated slightly to the right-hand side with respect to the axis of rotation, the influence of which is particularly prominent in the proximal-region case. Such a systematic displacement can be perhaps attributed to the geometrical (in)accuracy of the platform.

3.6 Conclusion

With the primary interest of establishing individualised hearing models described in chapters 2 and 5, the head-related transfer functions for 6 subjects have been measured in an anechoic chamber. Two types of loudspeakers have been used exclusively at two different distances to the subject location, $1.5\,m$ for the distal-region and $0.3\,m$ for the proximal-region, where the subject's seat was incrementally rotated to give a total of 72 recordings at every 5° in the horizontal plane. In particular, the use of an automated voice-feedback system aided by the head-tracker has been found successful in resolving the issue of subject positioning, providing a reasonable level of accuracy without the use of the headrest which possibly interferes more intensely with the sound field than the small head-tracking device.

In both time and frequency domains, the measured HRTFs have been examined and compared with those obtained in similar studies reported in the literature. Some known features have been also found in the current data, where the effective frequency range appeared to be limited by the responses of the in-ear microphone and the loudspeakers. The HRTFs have been further processed to produce ITDs and ILDs, and thus the characteristic curves, which have been found in the distal-region case to be unique for each subject in terms of the width of the 'lobe.'

The influence of possible positioning errors has been analysed in two ways. Firstly, the random errors within the tolerance given by the automated positioning system have been simulated for the selected degrees of freedom, and the maximum deviation in the ITDs has been found to be about $10\,\mu s$ and $17\,\mu s$ for the distal- and the proximal-region measurements, respectively. This is an inevitable variation in the measured data, resulting from the use of head-tracking system with no physical means to maintain the head position. On the other hand, the systematic error associated with the lateral misplacement (and the consequent misorientation) in the initial referencing procedure is an undesirable error, which has been also simulated to show its influence on the ITDs and the ILDs. This error analysis suggested that there might have been a consistent misplacement of subjects to the right, leading to an increase both in the ITDs and the ILDs for the rear hemisphere, which is especially prominent in case of the proximal-region measurement. The greater vulnerability of the proximal-region data to the positioning errors is readily understood by considering the greater influence of acoustic parallax, and the application of the measured HRTFs in the current study will be made only within such an understanding.

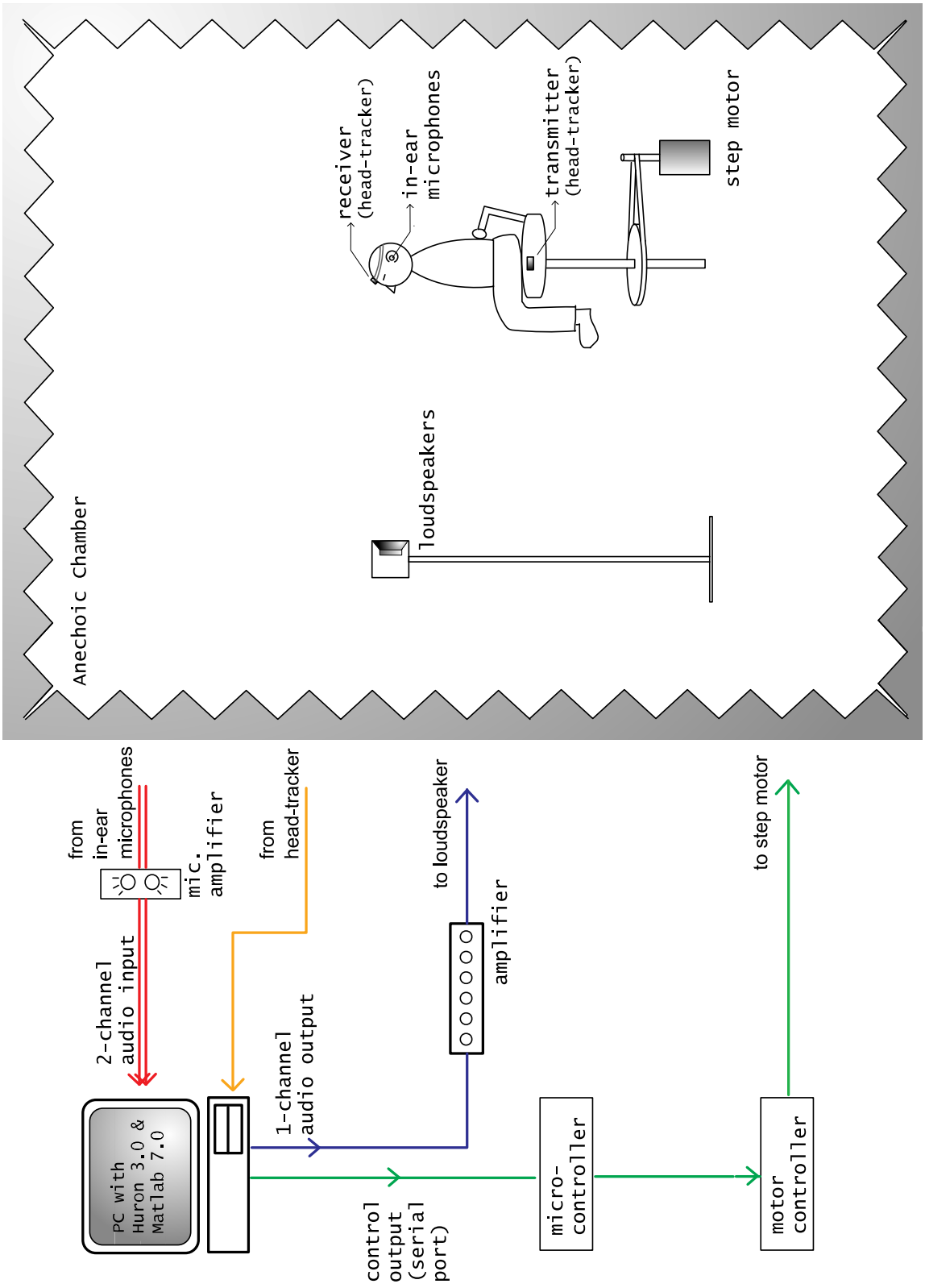


FIGURE 3.1: Sketch of measurement arrangement in the anechoic chamber and the control room.

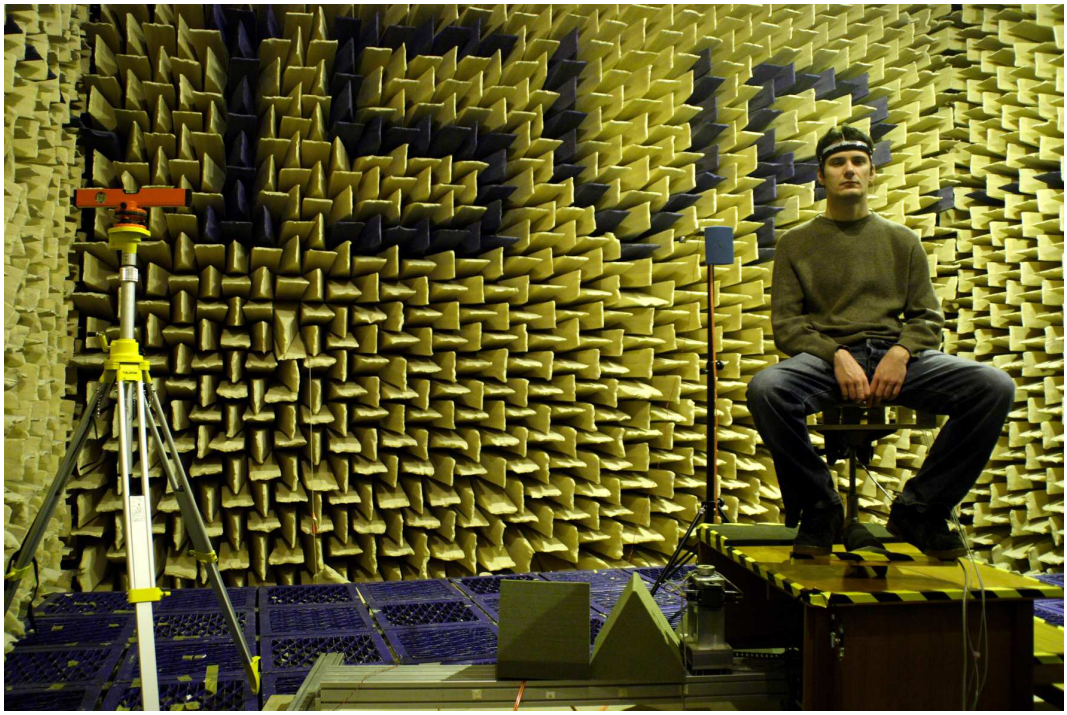
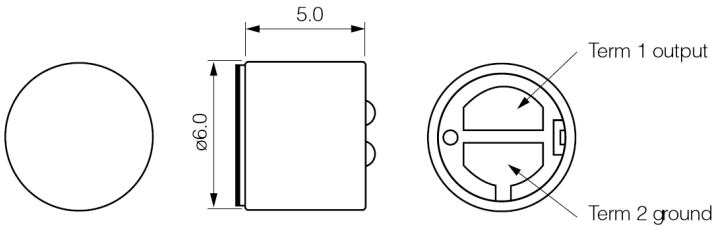
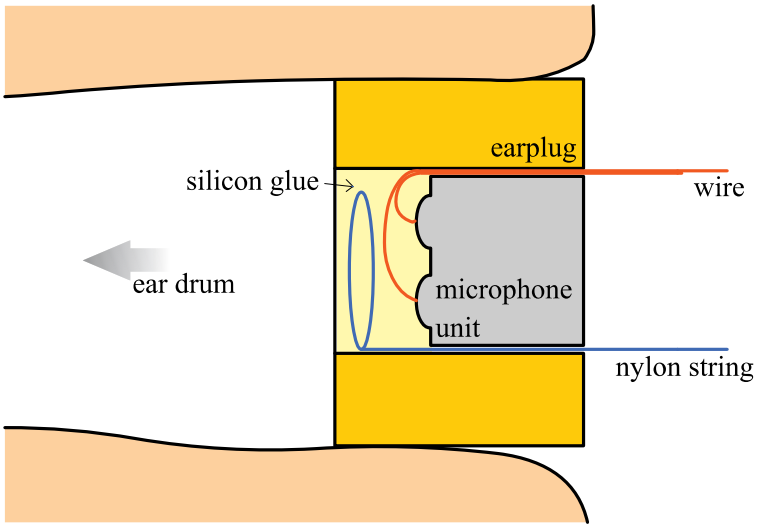


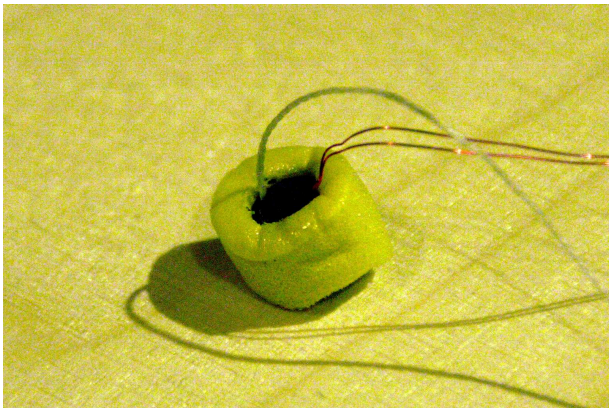
FIGURE 3.2: Photograph taken on the measurement site.



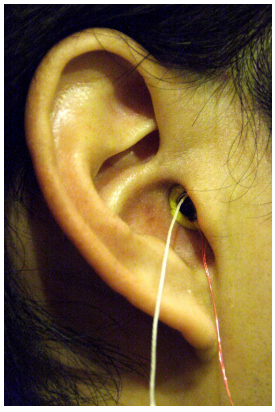
(a)



(b)



(c)



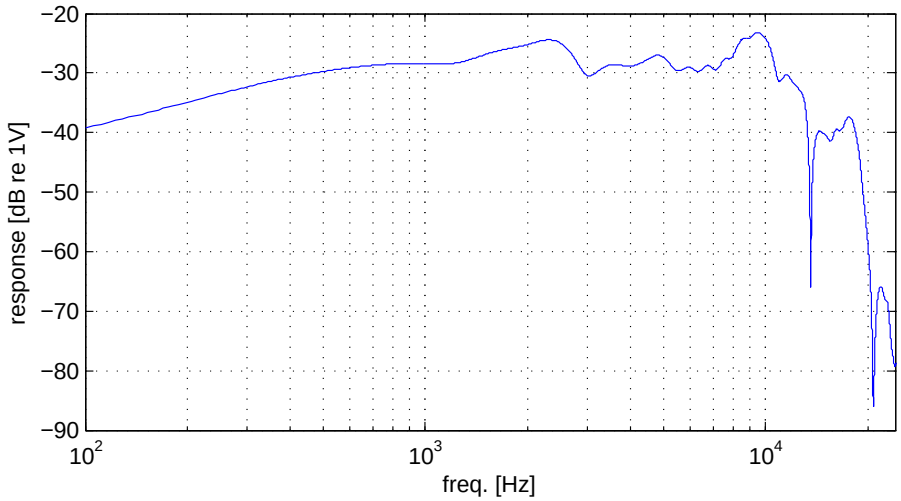
(d)

FIGURE 3.3: (a) Diagram of Panasonic WM-60A, taken from the manufacturer's data sheet. (b) Diagram of custom treatment. The microphone unit has been inserted to a spongy ear plug where a nylon string has been attached for easy and safe removal. The right side of the microphone unit in this diagram faces out of the ear. (c) Photograph of actual treated microphone unit. (d) Microphone unit inserted to the subject's ear.

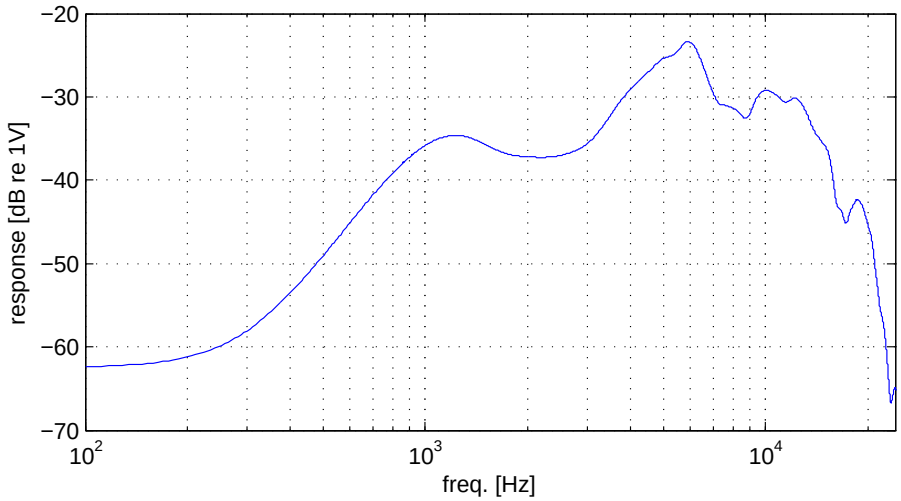
(a)

(b)

FIGURE 3.4: Photographs of (a) Celestion AVC102 (taken from manufacturer's website) and (b) Bujeon BMS-1709SL08C in a custom-made plastic cabinet (taken from Cho et al. [55]).

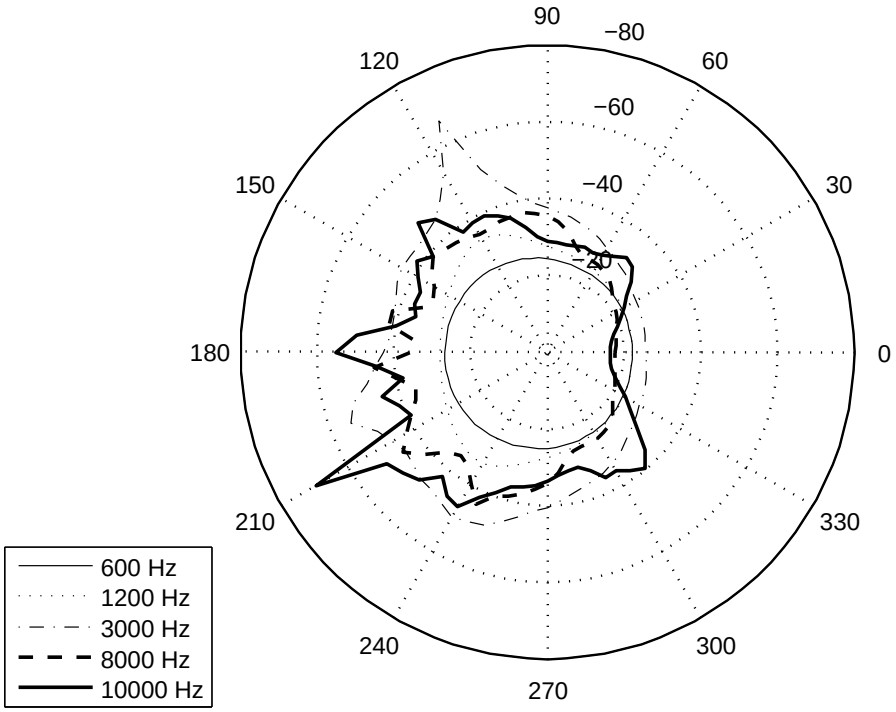


(a)

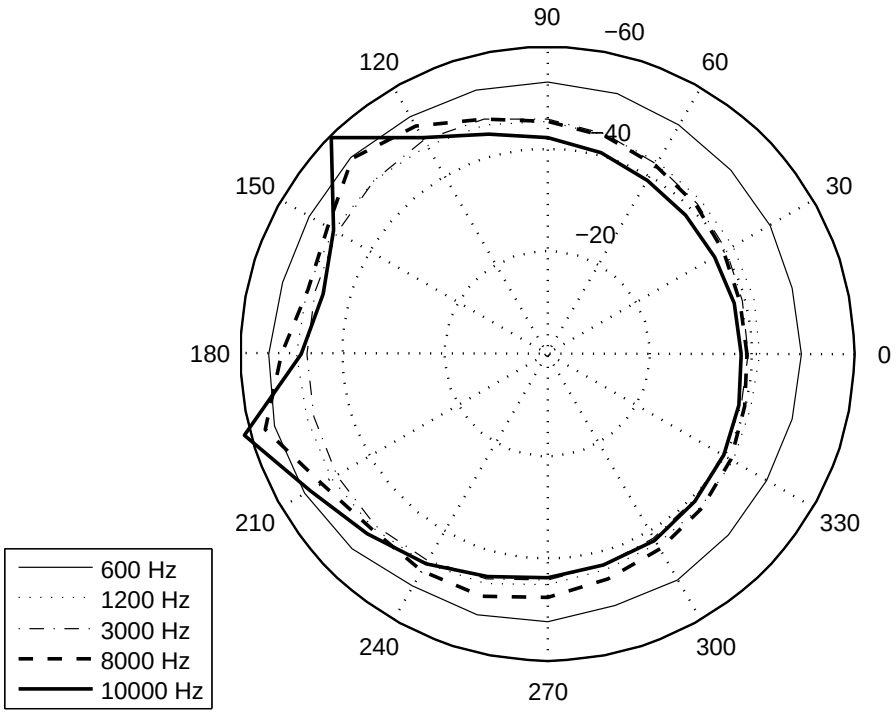


(b)

FIGURE 3.5: Free-field responses measured with (a) Celestion AVC102 (1.5 m) and (b) Bujeon BMS-1709SL08C (0.3 m). 200-point Hanning windows have been applied to the responses at their peaks in the time domain.

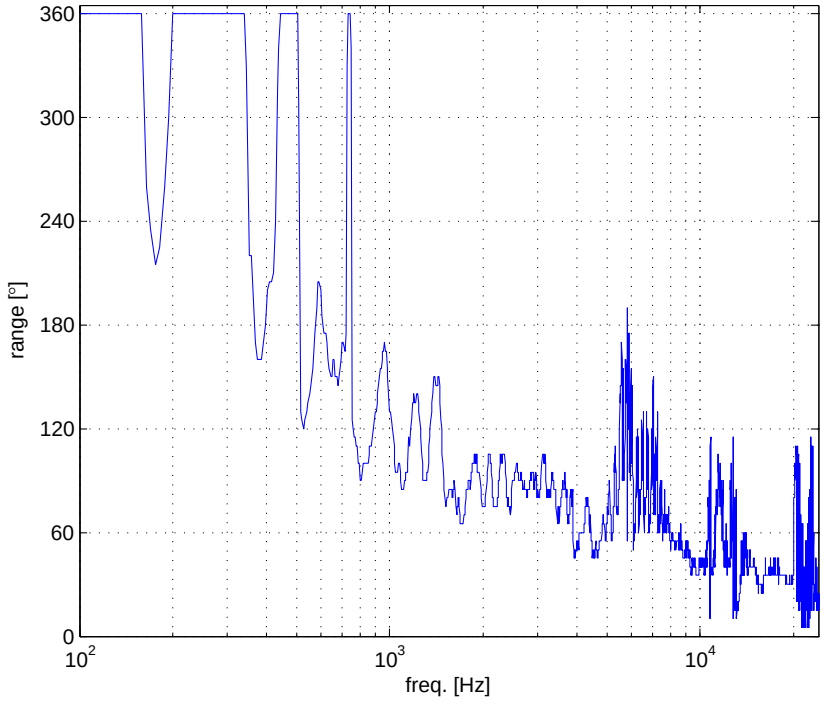


(a)

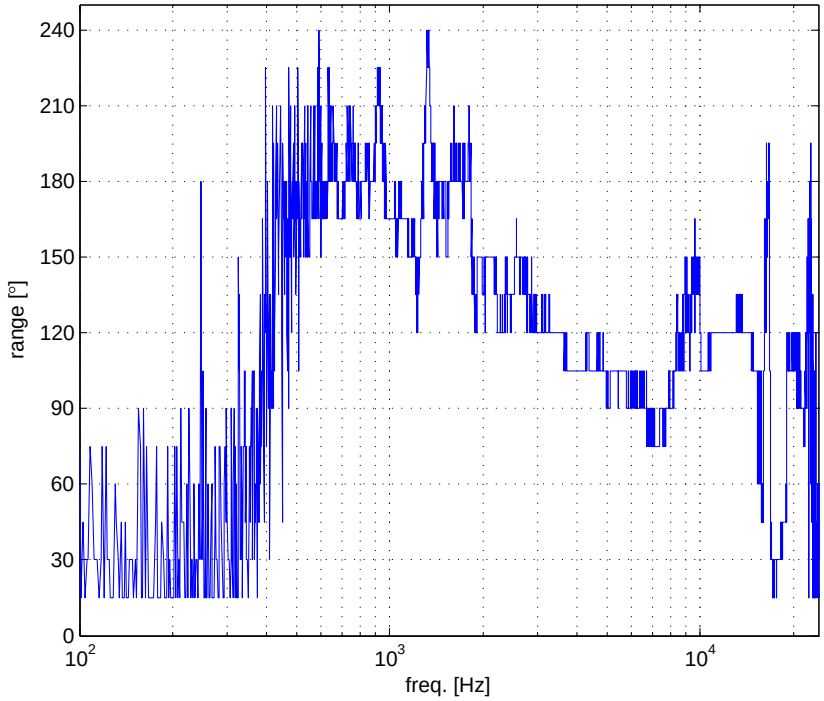


(b)

FIGURE 3.6: Directivity patterns (dB) obtained for (a) Celestion AVC102 at every 5° at 80 *cm* and (b) Bujeon BMS-1709SL08C at every 15° at 30 *cm*.

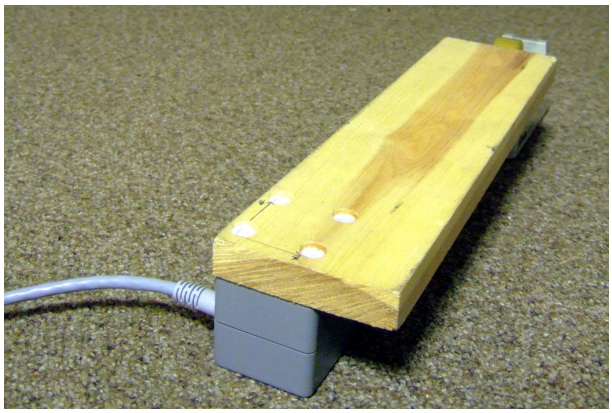


(a)

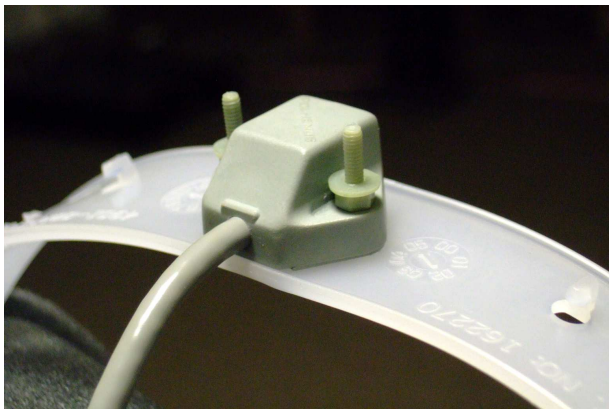


(b)

FIGURE 3.7: 3-dB beamwidths computed from the directivity patterns for (a) Celestion AVC102 and (b) Bujeon BMS-1709SL08C. Note that the spatial resolutions were (a) 5° and (b) 15°.



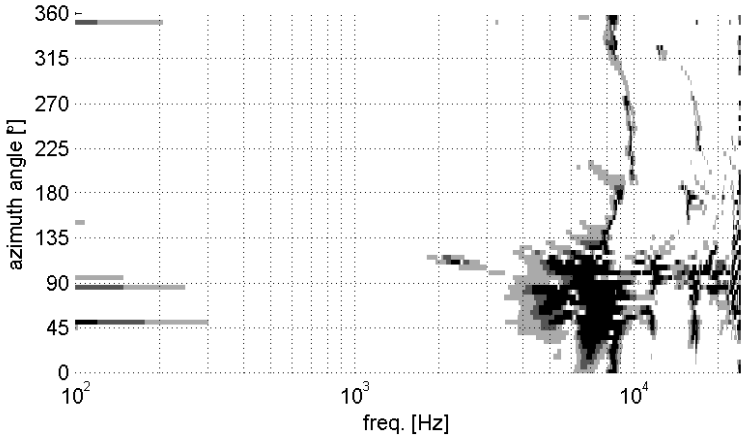
(a)



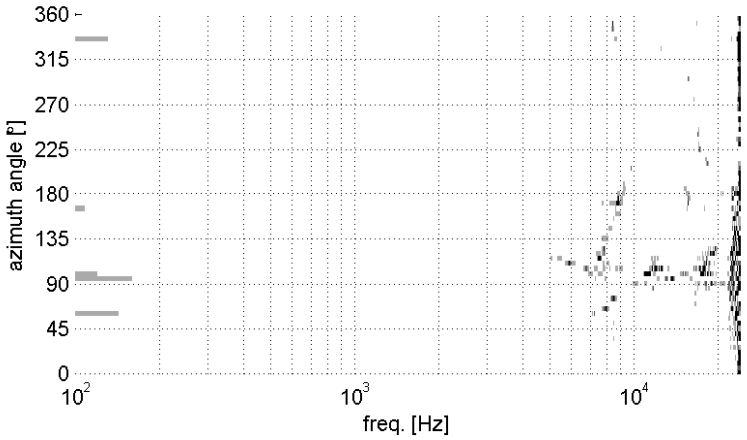
(b)

(c)

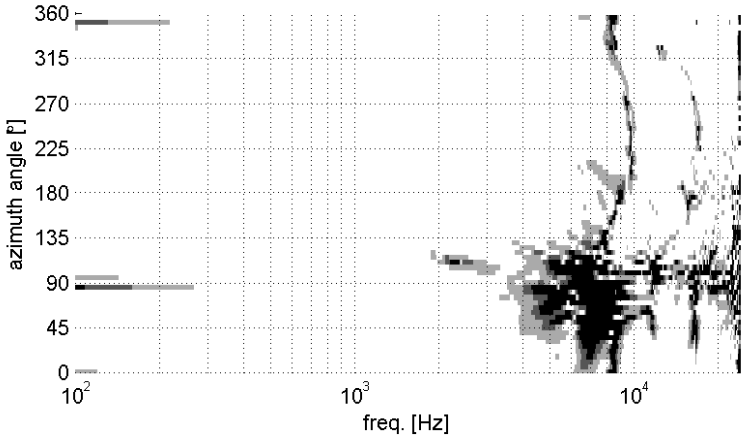
FIGURE 3.8: Polhemus FASTRAK use for subject positioning. (a) Transmitter unit temporarily attached to wooden panel. (b) Receiver unit attached to the safety helmet lining. (c) Photograph of the headband with the receiver unit worn by subject.



(a)

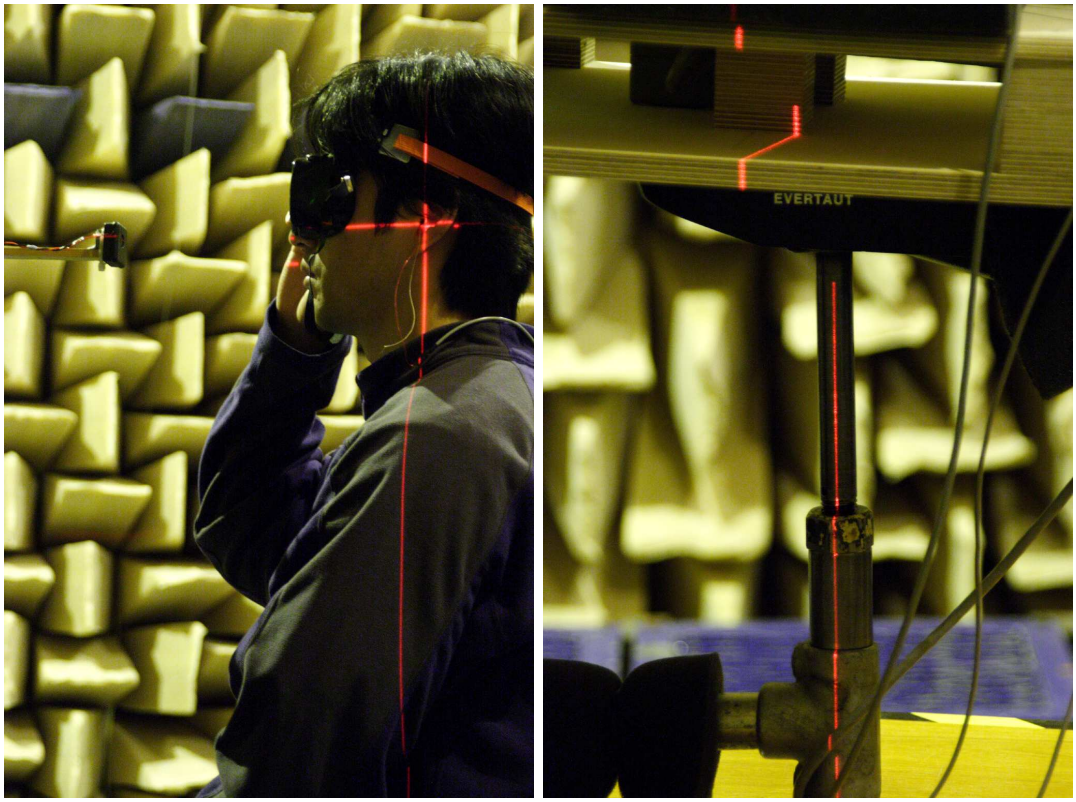


(b)



(c)

FIGURE 3.9: Deviation in HRTFs frequency response when (a) the headband, (b) the backrest and (c) both headband and backrest have been used. The four-step grey-scale level indicate the increase in deviation by 1 dB from less than 1 dB (white) to more than 3 dB (black).



(a) (b)

FIGURE 3.10: Laser level used to adjust the subject's ear position with reference to (a) the height of the loudspeaker and (b) the axis of rotation. For safety, a laser protective goggle has been used during the initial positioning procedure.

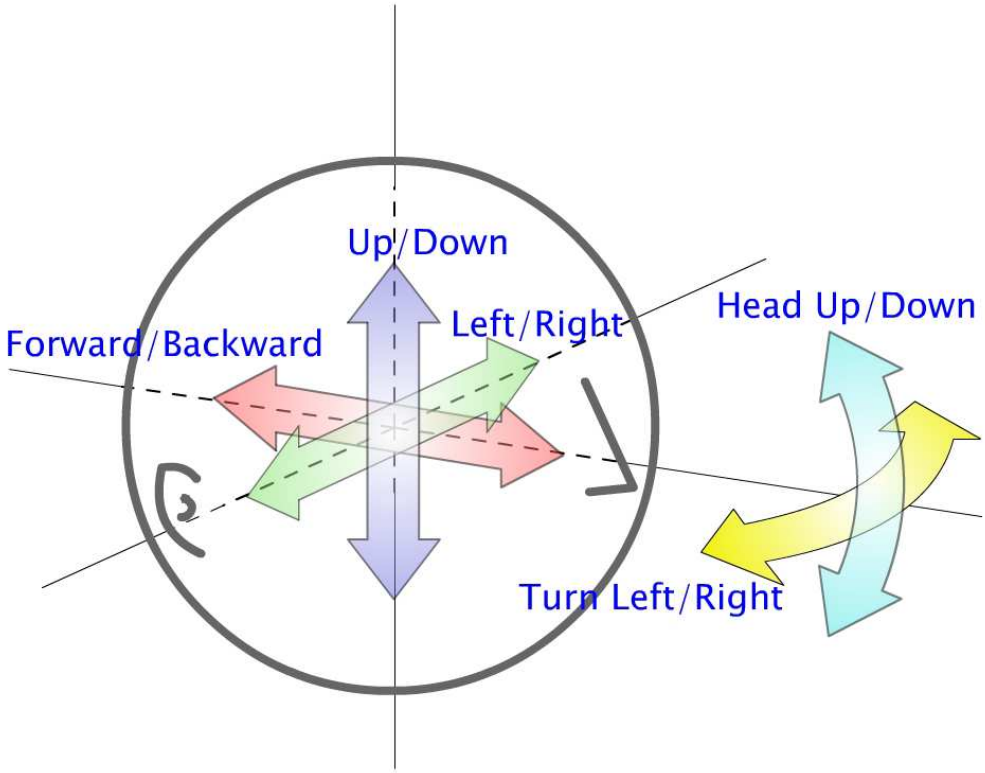


FIGURE 3.11: 5 degrees of freedom considered for the automated voice-feedback system. When the displacements from the reference position in terms of the 3 translational and 2 rotational degrees of freedom are more than the predefined tolerances, a voice instruction for the direction with the greatest deviation is played over loudspeaker to guide subject to reposition.

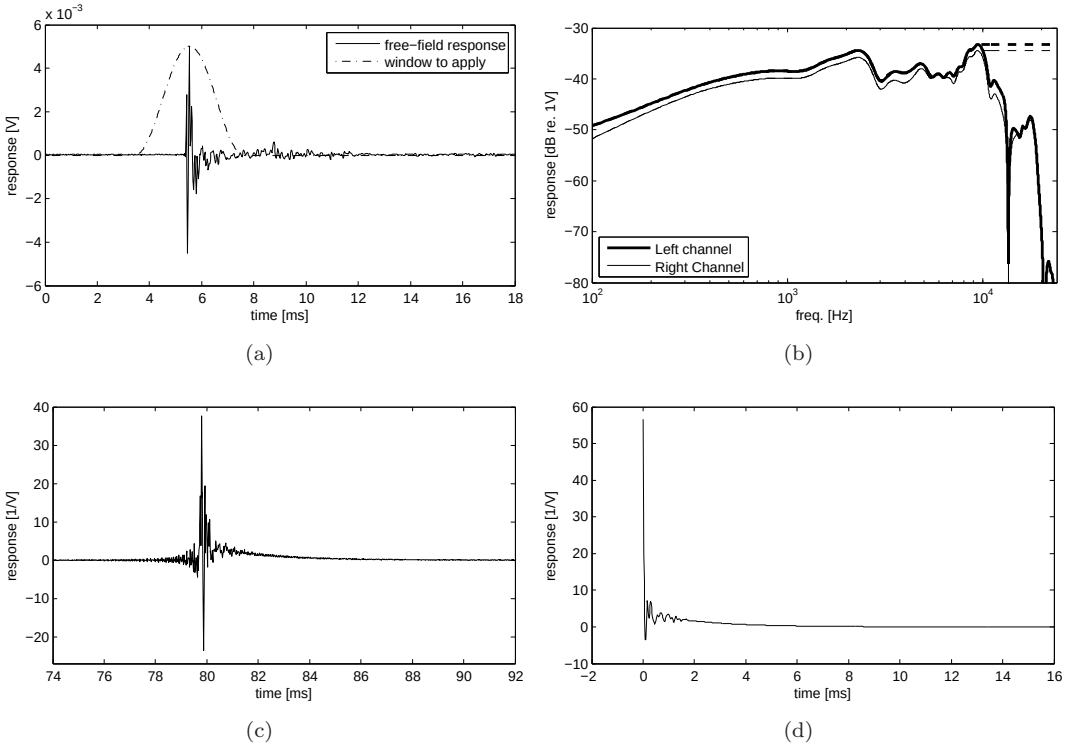


FIGURE 3.12: Inverse filter made for the distal-region HRTFs. (a) 200-point Hanning window has been applied to the free-field response at the peak. (b) Converted to the frequency domain, the magnitude response has been flattened from 9.5 kHz. (c) The inverse of the modified frequency response is converted back to the time-domain and FFT-shifted. (d) Finally, the real cepstra is taken to produce minimum phase inverse filter.

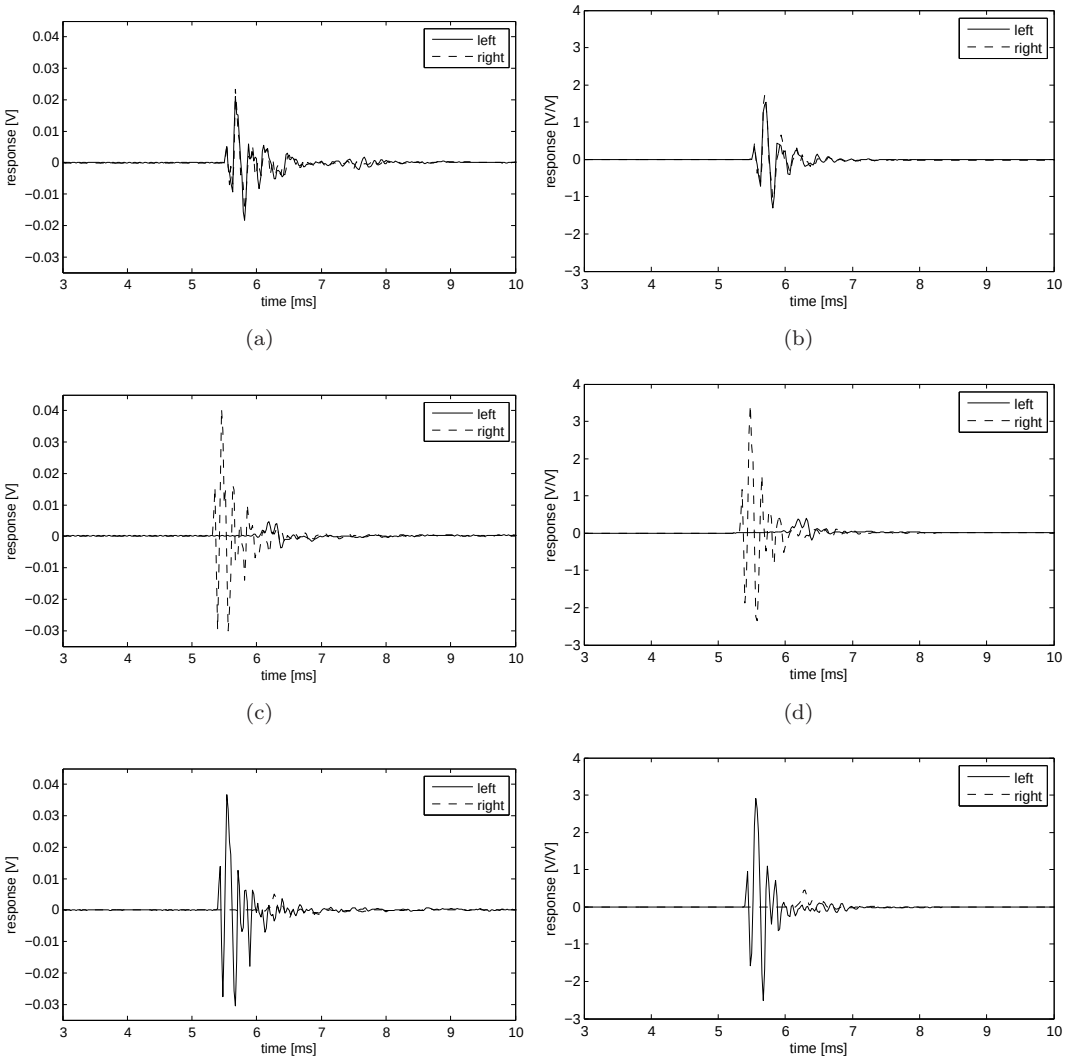


FIGURE 3.13: Distal-region HRIRs (subject SF) at representative azimuth angles at (a)(b) 0°, (c)(d) 90° and (e)(f) 270°. Panels (a), (c) and (e) show HRIRs before post-processing, while panels (b), (d) and (f) after post-processing.

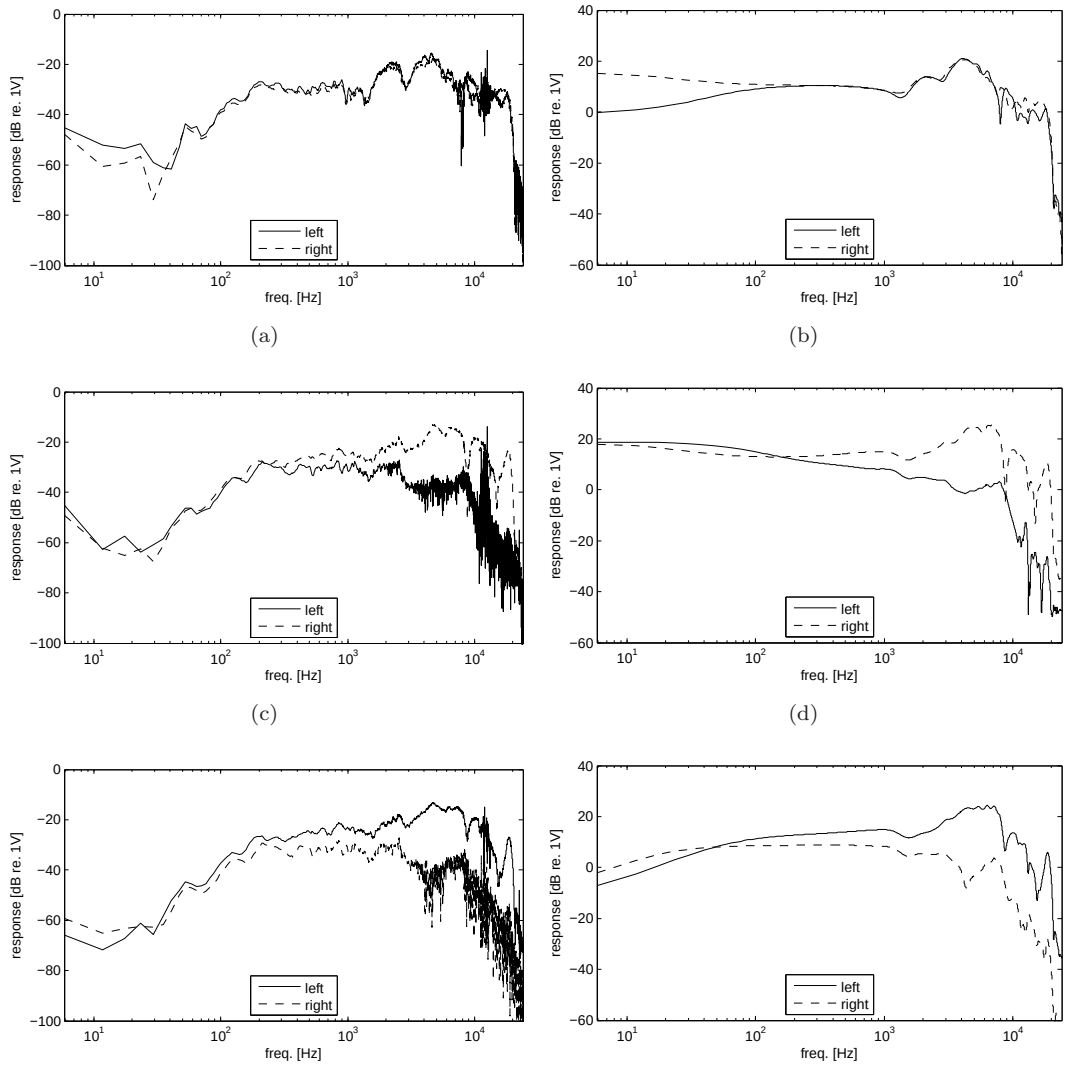


FIGURE 3.14: Distal-region HRTFs (subject SF) at representative azimuth angles at (a)(b) 0°, (c)(d) 90° and (e)(f) 270°. Panels (a), (c) and (e) show HRTFs before post-processing, while panels (b), (d) and (f) after post-processing.

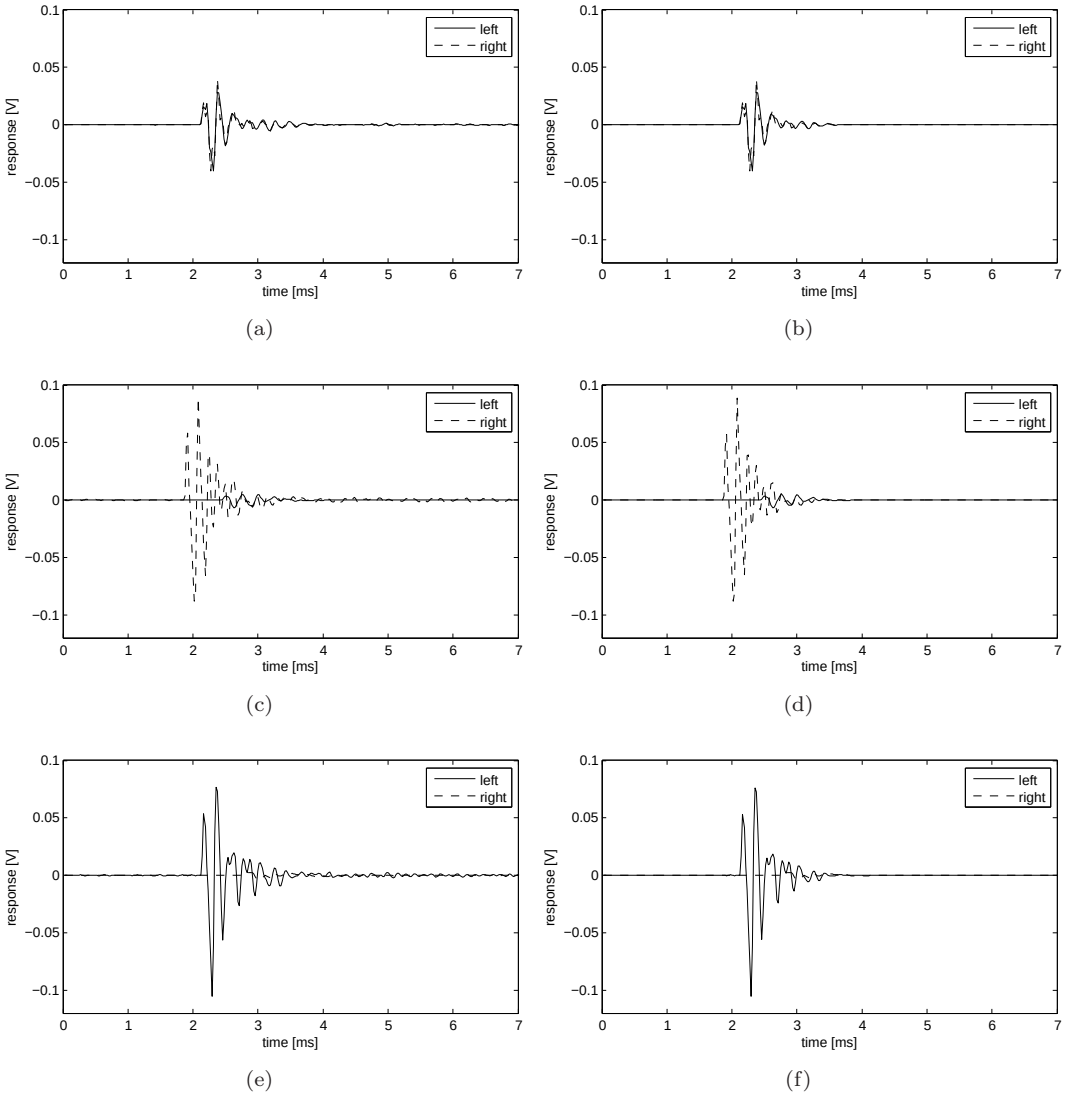


FIGURE 3.15: Proximal-region HRIRs (subject SF) at representative azimuth angles at (a)(b) 0°, (c)(d) 90° and (e)(f) 270°. Panels (a), (c) and (e) show HRIRs before post-processing, while panels (b), (d) and (f) after post-processing, which includes windowing and zeroing only.

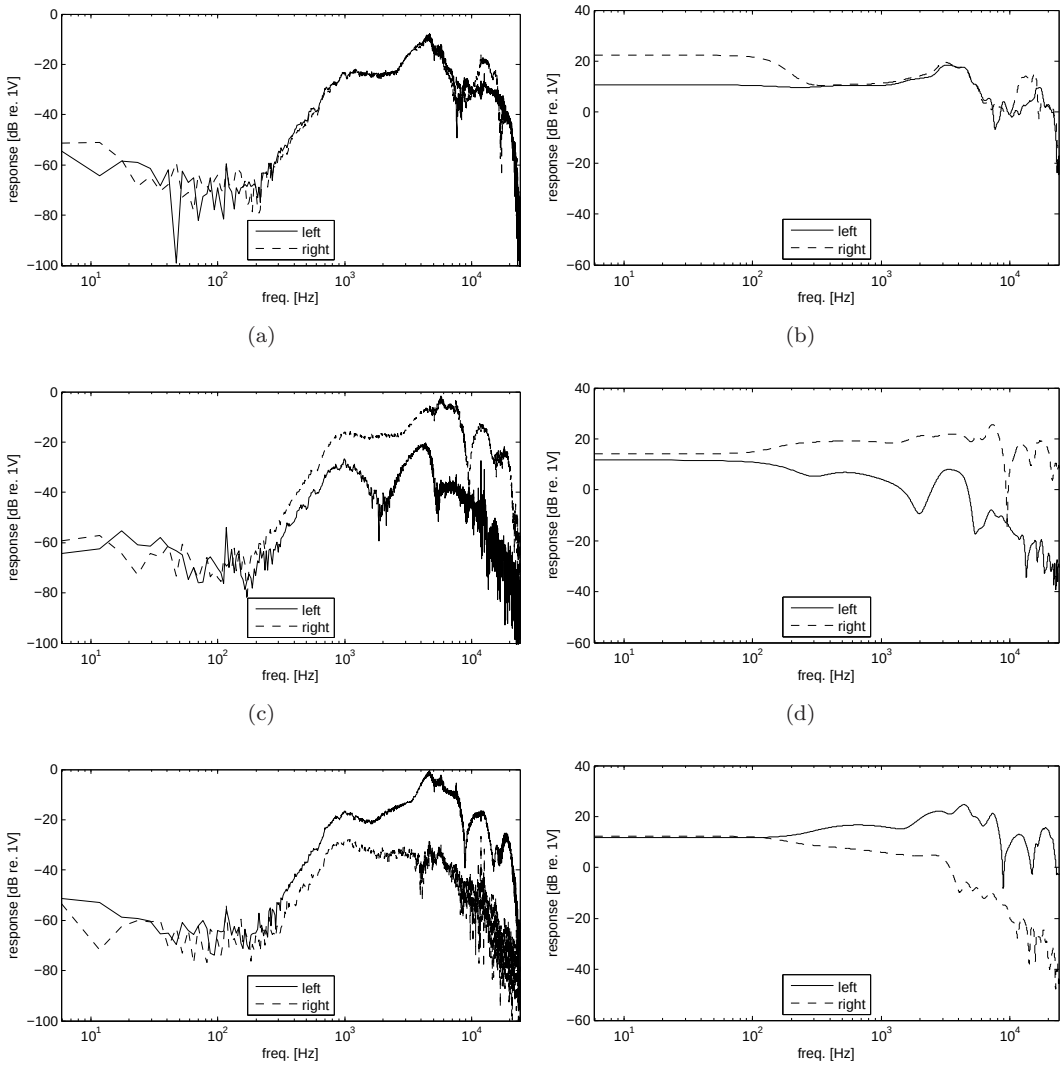


FIGURE 3.16: Proximal-region HRTFs (subject SF) at representative azimuth angles at (a)(b) 0°, (c)(d) 90° and (e)(f) 270°. Panels (a), (c) and (e) show HRTFs before post-processing, while panels (b), (d) and (f) after windowing in the time-domain and magnitude equalisation with respect to the free-field response.

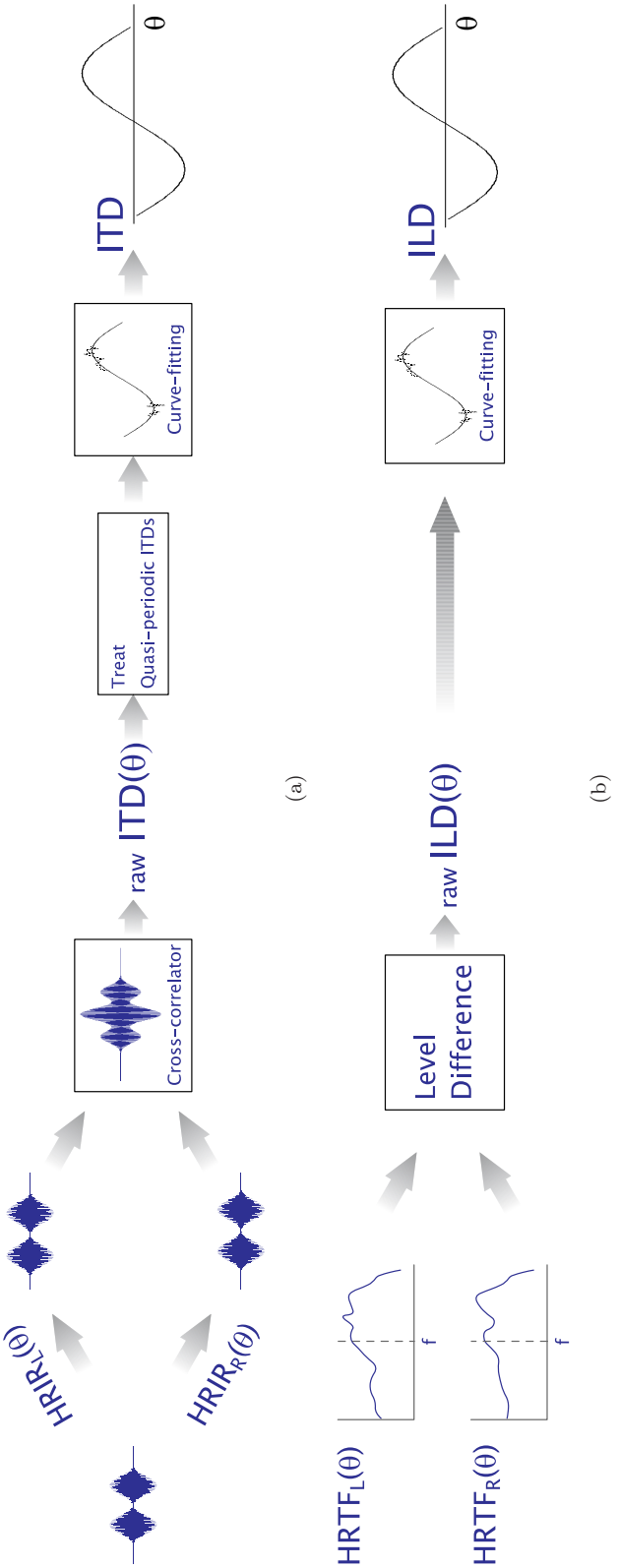


FIGURE 3.17: Procedures to obtain (a) ITDs and (b) ILDs from HRTFs are illustrated. ITDs are computed in the time domain using cross-correlation, while ILDs are acquired in the frequency domain by comparing magnitude responses.

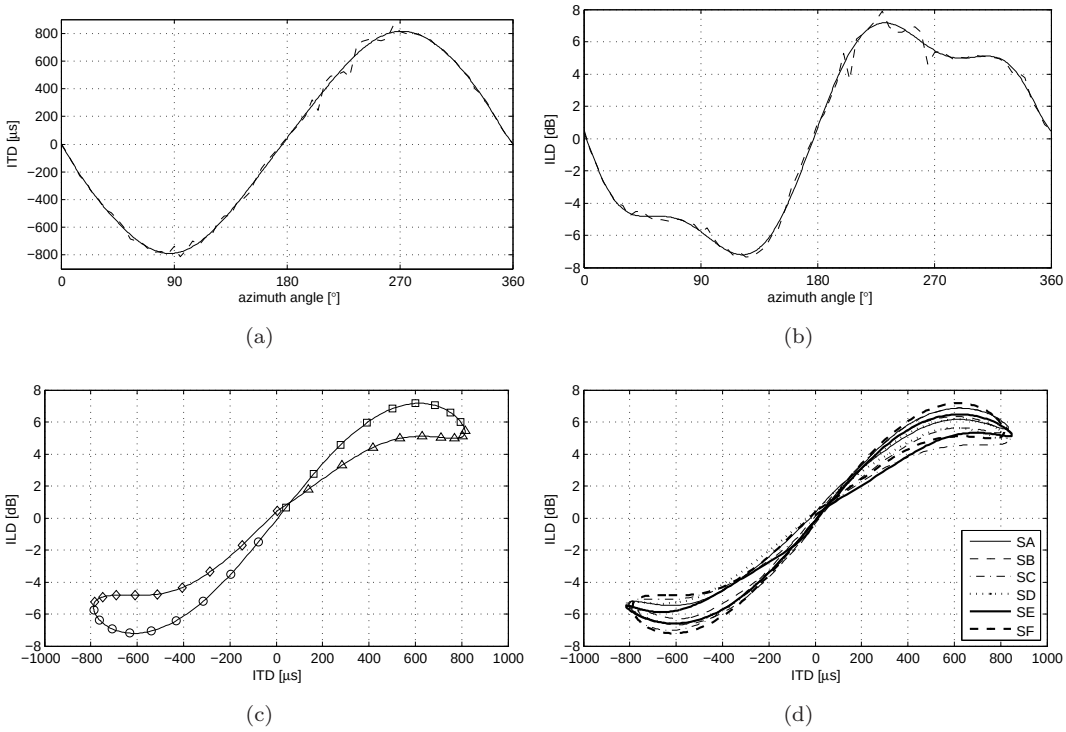


FIGURE 3.18: (a) ITDs and (b) ILDs obtained from the distal-region HRTFs (subject SF). The raw ITDs and ILDs before curve-fitting procedure are shown as dashed lines. (c) The distal-region characteristic curve at 600 Hz is shown for the subject SF, which has been marked at every 10° (◇: 0° ∼ 80°, ○: 90° ∼ 170°, □: 180° ∼ 260°, △: 270° ∼ 350°). (d) Distal-region characteristic curves at 600 Hz are shown for all subjects.

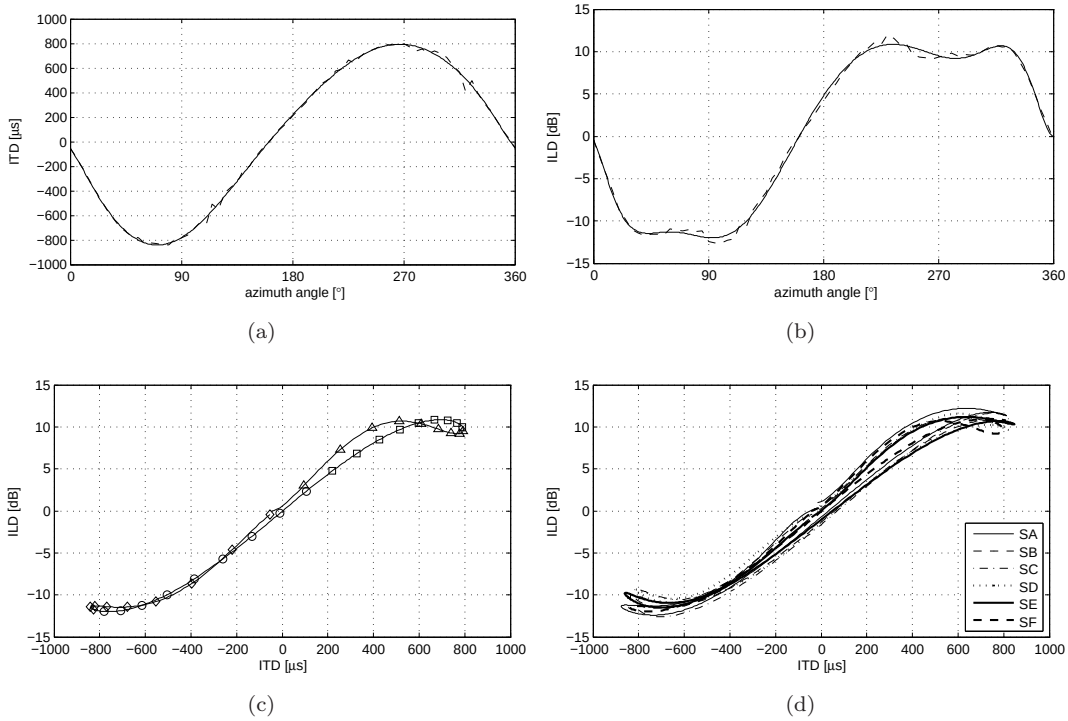


FIGURE 3.19: (a) ITDs and (b) ILDs obtained from the proximal-region HRTFs (subject SF). The raw ITDs and ILDs before curve-fitting procedure are shown as dashed lines. (c) The proximal-region characteristic curve at 600 Hz is shown for the subject SF, which has been marked at every 10° ($\diamondsuit$: $0° \sim 80°$, $\circ$: $90° \sim 170°$, $\square$: $180° \sim 260°$, $\triangle$: $270° \sim 350°$). (d) Proximal-region characteristic curves at 600 Hz are shown for all subjects.

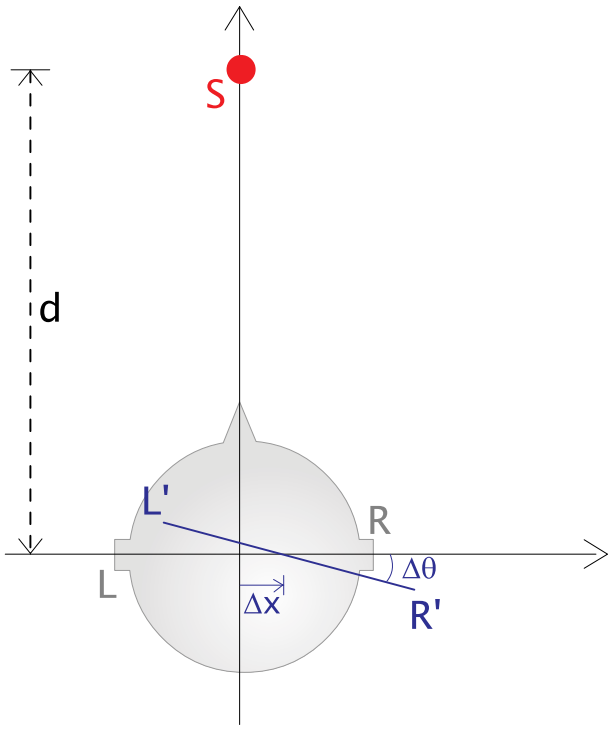
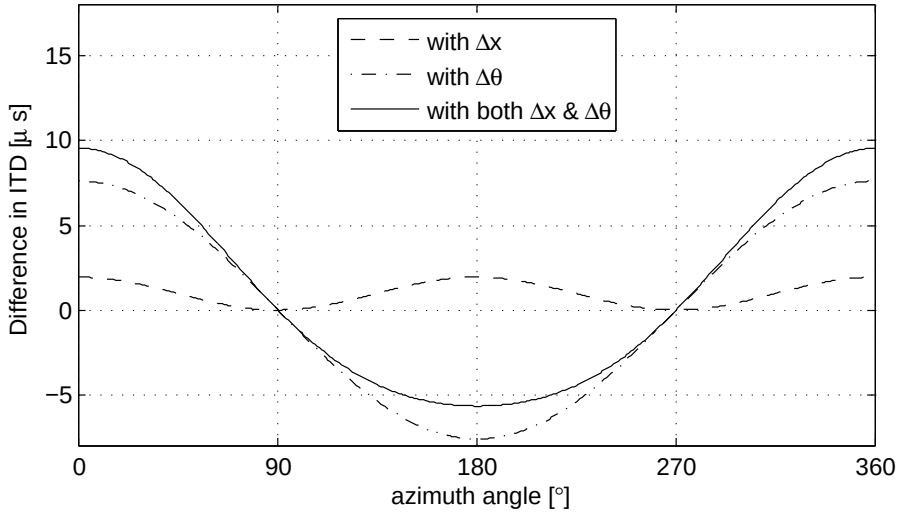
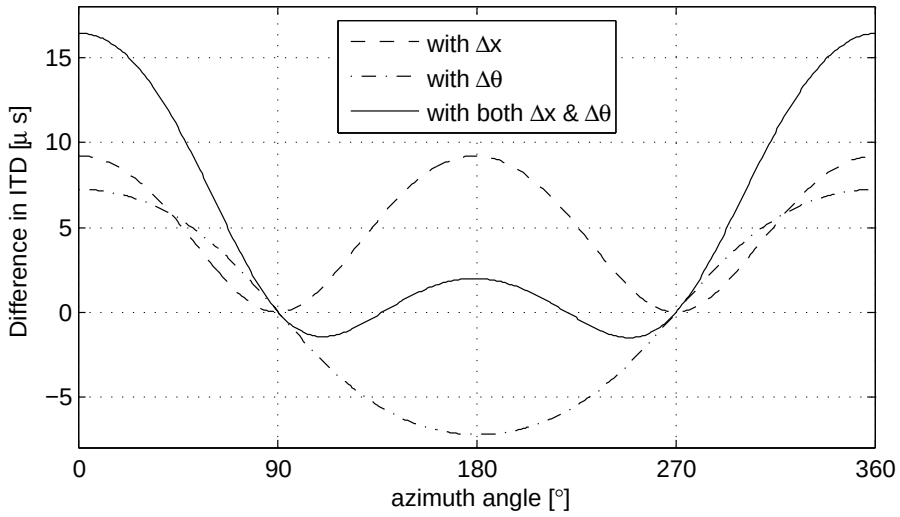


FIGURE 3.20: L and R indicate the ideal positions of subject's ears where S represents the source location. Path lengths from the source to the misplaced ear locations, L′ and R′ can be computed to give the degree of deviation in ITDs, when the subject is displaced in the lateral direction by Δx, and rotated in the azimuth-sense by $\Delta\theta$.



(a)



(b)

FIGURE 3.21: The deviations in ITDs for (a) the distal-region and (b) the proximal-region measurements are shown across azimuth angle, when $\Delta x = +0.5\,cm$ (misplacement to the right) and/or $\Delta\theta = -0.75^\circ$ (head turning to the right) have been assumed for the configuration shown in Fig. 3.20. It is shown that the proximal-region measurement is more vulnerable to random positioning errors within the tolerance of the voice-feedback guidance system.

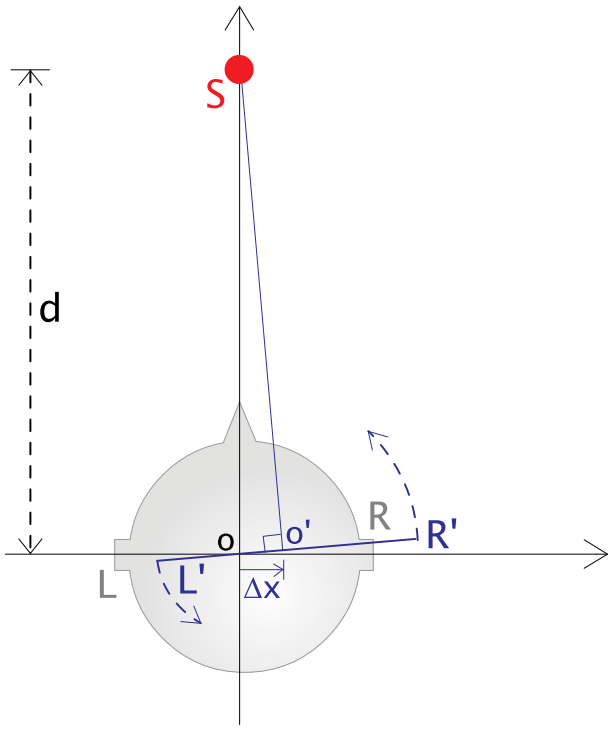


FIGURE 3.22: The consequence of the error in the initial positioning procedure is illustrated. Once the subject is slightly misplaced, say, to the right, the angle alignment procedure will turn subject's head to equalise the path lengths. Therefore, the misplaced ear positions L′ and R′ will draw two different circular trajectories when the seat is rotated with respect to the origin, O.

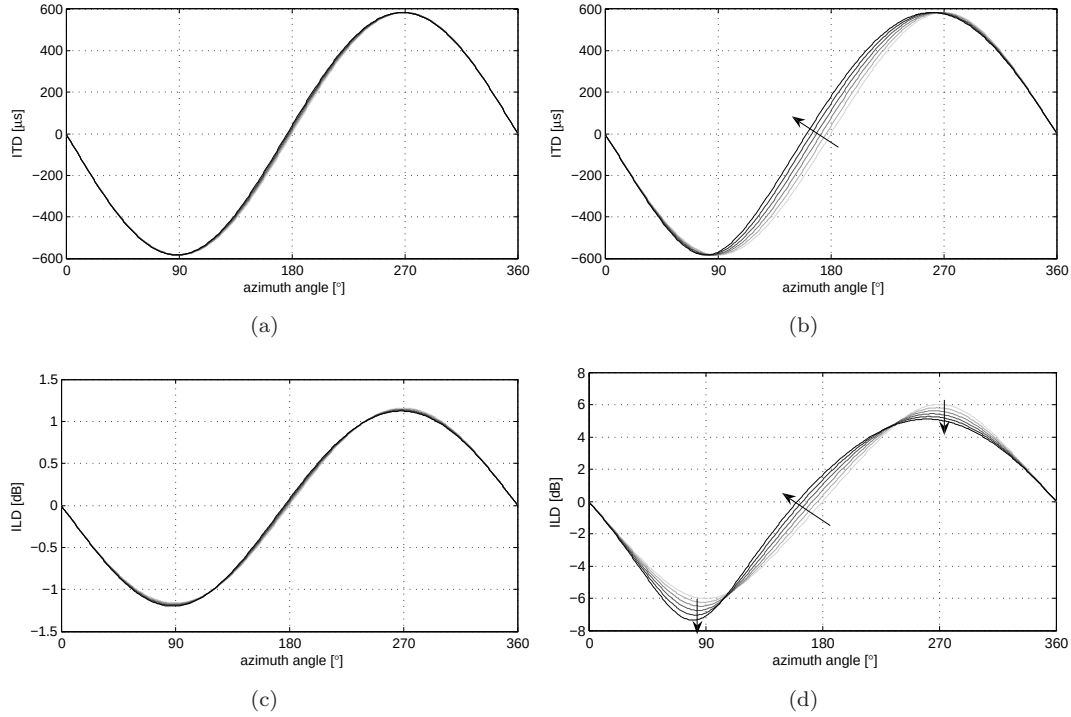


FIGURE 3.23: From the configuration shown in Fig. 3.22, ITDs and ILDs have been simulated for the distal- and the proximal-region measurements. Panels (a) and (c) show ITDs and ILDs for the distal-region, and panels (b) and (d) for the proximal-region, respectively. Δx has been assumed to vary from 0 to 5 cm where the darker line colour indicates the greater displacement. It is clear that the proximal-region measurement is more vulnerable to the initial mispositioning errors.

Chapter 4

Listening test I - lateralisation of dichotic pure tones

4.1 Introduction

While the localisation of various acoustic stimuli is very important for the successful human orientation in the living environment, the associated auditory images are perceived, with reasonable accuracy, almost always to be out in the space at the actual positions of the sound sources (e.g. see Hartmann and Wittenberg [32]). The spatial perception of such auditory scenes external to the head is termed as 'the outside-head localisation' (OHL) [17], or simply 'localisation.' In the literature, it has been also reported that 'inside-head localisation' (IHL) [17] is possible in human perception (e.g. see Blauert [4]), and it is sometimes referred to as 'lateralisation,' since the positions of the internal images are mostly reported only in terms of their lateral displacement from the head centre. In contrast to the localisation of external images, inside-head auditory images are usually created in dichotic listening environment where the two ears receive independent signals without any cross-talk [17]. For example, in an old but well-known experiment, two ends of a long tube have been placed at the two ears of a subject, who reported the positions of the perceived auditory images while the interaural time difference (ITD) was manipulated by the experimenter tapping the tube at different locations [4].

In addition to the ITD, it is well known that the interaural level difference (ILD) can also be manipulated to displace the intracranial auditory images [4], and in the recent listening tests regarding lateralisation, the time delay and the relative amplitude gain are digitally controlled to give target signals that are usually presented over headphones.

Apart from the methods used to generate the source signal, there have been many different ways employed in relevant listening tests to quantify the subjective judgements of the image positions. For example, a scale chart has been displayed in front of the listener during the test, who was instructed to make judgements on the given scale, or at least use the chart as a visual reference for the size of the head [31, 34, 35]. More recently, the acoustic pointer method has been widely used for lateralisation listening tests where the subject controls the perceptual position of a pointer signal by adjusting the ITD or the ILD [29, 36, 56, 57]. Using the former method with a visual chart, the perceived laterality for a target signal is directly quantified as a number, while the latter requires the listener to perform a matching task, reporting the image position as equivalent to the location of one of the acoustic pointers. It appears that the matching task possibly reduces the variability in the subjective judgements compared to the visual chart method, but the limitation of the pointer range is often regarded as an issue in association with the ambiguity of the phase difference (ITD pointer) and the excessive unilateral loudness (ILD) [4, 29].

The laterality judgements obtained in listening tests have been often compared with the predictions of relevant hearing models (e.g. see Domnitz and Colburn [29]), most of which are at least partly based on the coincidence model suggested by Jeffress [5] where the neural computation of the interaural cross-correlation is implemented over a delay-line structure. Readers are referred to section 2.1 for a summary of binaural hearing models.

The aim of the experimental study reported in this chapter is similar to the previous works described in the literature, such that the laterality judgements for dichotic pure tone signals will be obtained and compared with the predictions made by the decision-making model suggested in chapter 2. However, it has been additionally suggested that the current model operates on the basis of individual HRTFs, producing unique predictions for each subject's auditory perception. Therefore, using the individual characteristic curves given by the HRTF (see chapter 3), the particular relationship between a subject's judgement and the prediction from his own decision-making model will be investigated in comprehensive ranges of target ITD and ILD.

In terms of the test methodology, the acoustic pointer based on non-individual HRTFs [58] will be employed for the matching task in the current listening tests. Compared to the ITD- or the ILD-based acoustic pointers, the HRTF-based pointer is expected to be more naturally perceived by the listeners with spectral characteristics identical to the target signals. In addition, the subjective judgements can be represented as the azimuth angle of the matched HRTF, in the same unit as the model predictions, facilitating a comparative analysis.

The design of the current listening test will be first described in section 4.2 where the HRTF-based acoustic pointer and the employed software platform will be detailed. The result of the laterality test will be then presented in section 4.3, followed by the simulation results and the relevant discussion in section 4.4. Finally, in section 4.5, the two results from the subjective listening test and the model simulation will be compared, and conclusion is presented in section 4.6.

4.2 Test method

A small semi-anechoic room designed for the audiology clinic at the Institute of Sound and Vibration Research (ISVR), University of Southampton has been found to be sufficiently quiet for the current listening tests where stimulus signals will be presented only over headphones. Fig. 4.1 illustrates the test room in which a subject sits wearing headphones, with access to input devices and a display connected to a desktop PC, which is located in a separate control room. All the test procedures including the generation and the playback of the stimulus signals have been controlled by this PC via a graphic user interface (GUI) designed in Matlab 7.0. The details regarding this GUI will be dealt with during the description of the test procedure presented later in this section.

The test signal is composed of two parts, the target and the pointer as shown in Fig. 4.2. Two consecutive identical pulses of dichotic pure tones at frequency f are the target signals of which the ITD and the ILD across the left and the right channels are controlled according to the test design. On the other hand, the following two pulses are the acoustic pointer, the lateral position of which can be adjusted by the subject using the GUI. $10\,ms$ smooth rise and fall periods have been applied to all pulses preventing audible 'clicks,' where the duration of each part and the intervals are denoted in Fig. 4.2 in the unit of ms.

The acoustic pointer has been created by filtering the pure tone signal identical to the target signal (before being given the target ITD and/or ILD) with one of the KEMAR HRTFs [27] in the horizontal plane. Filtering a monaural signal with HRTFs is usually intended to give a full 3-dimensional illusion of an auditory scene including the perception of distance [4, 59], but it has been an important issue with the binaural technology that the acoustic images provided by non-individual HRTFs are mostly perceived inside the listener's head especially in case of the headphone playback [32]. Having lost the distance localisation cue, however, the binaural signals created by the HRTFs maintain a reasonably unique intracranial position, which is sufficient to be an acoustic pointer or anchor for the current listening tests. The maximum and the minimum lateralities perceived for dichotic tones may vary from subject to subject, but the angular range of the KEMAR HRTFs from -90° (subject's left) to +90° (subject's right) appears to be the best attempt. The resolution of the HRTFs has been increased from 5° (measurement) to 1° according to the interpolation scheme described in appendix A. After all, the subject can report the lateral position of a given target signal by adjusting the acoustic pointer from -90° to +90° with 1° resolution.

The current acoustic pointer based on the characteristics of non-individual HRTFs is expected to be very naturally perceived by the listener, in contrast to the other types

of acoustic pointer using either ITD or ILD. The laterality given by the ITD pointer becomes ambiguous when the given relative time delay is more than the half-period of the signal. The range of ILD pointer is also arbitrarily limited when the loudness on one side becomes unpleasantly excessive. Wider-band frequency contents may resolve the arbitrary limitation of the laterality range especially for the ITD pointer, but spectral characteristics different from the target signal might bias the judgement of the perceived lateral position [29]. In addition, the HRTF acoustic pointer will produce test data in the unit of azimuth angle, which is directly comparable to the model predictions discussed in chapter 2, hence relating the lateral position inside the head to the source position outside in the space. In this way, the relationship between the two important quantities in spatial hearing, lateralisation and localisation can be better understood, the link between which has not yet been fully revealed.

Test frequencies have been selected to be 600 Hz and 1200 Hz for the following reasons. First, there are listening test results for the 600-Hz laterality reported in the literature [31] which can be compared with the data given by the current test. Second, for this pair of test frequencies, the period of the laterality curve at 1200 Hz is expected to be half of that at 600 Hz, and the range of the laterality can be also easily compared between frequencies. This comparison across frequency is expected to be readily made only below 1500 Hz, above which the waveform ITD becomes less influential on the auditory image formation [26]. By using an artificial ear (B&K 4153 and 4134), the sound level of the test signals have been measured to be 73.2 dBA and 79.5 dBA at 600 Hz and 1200 Hz, respectively, when no interaural disparity has been given to the target signal with the acoustic pointer at the 0° position.

In accordance with the range of the test frequencies, pure tone audiometry for the 6 subjects (SA, SB, SC, SD, SE and SF) has been carried out at 4 frequencies from 500 Hz to 1500 Hz where all subjects showed acceptable hearing levels for both ears (less than 20 dB hearing level). It is noteworthy that only the subject SE had experience of listening tests, while the others are technically inexperienced subjects. All subjects have participated in the test at 600 Hz, while the 1200-Hz laterality has been measured only by the two subjects, SA and SF. ITDs from $-1000\,\mu s$ to $+1000\,\mu s$ at every $100\,\mu s$ have been combined with ILDs of -6, 0, and 12 dB to give a total of 63 (21×3) trials for the 600-Hz target signals. On the other hand, due to the shorter signal period, the range of the target ITD has been reduced for the 1200-Hz target signals to be between $\pm 700\,\mu s$ but at every $50\,\mu s$, giving a total of 87 (29×3) trials. Normally, one set of trials has been completed within an 1-hour session, and for the 600-Hz target, 5 sessions have been completed by each subject on different days. For the 1200-Hz target, however, only 3 sessions have been arranged for SA and SF. To summarise, each pair of ITD and

ILD has been tested for 5 times at 600 Hz and 3 times at 1200 Hz for each selection of subjects.

The graphic user interface used in the current listening test is shown in Fig. 4.3. Once a session starts, a pure tone target signal amended with the first pair of ITD and ILD is presented to the listener. Subjects have been instructed to listen to this target-only signal for a few times, then to choose to add the pointer sound by unticking the "Play only the target." By pressing the left or the right arrow keys (arrow heads on the GUI blink on pressing), the subject can move the acoustic pointer between -90° and +90° while the whole test signal with the identical target signal is being presented repeatedly. If the subject finds that the pointer signal is best matched to the target signal in terms of the lateral position, he may go on to the next stimuli for a new pair of ITD and ILD by pressing the "NEXT" button, which will also save his judgement into the data file. If a judgement has been made by mistake or the subject wants to make an amendment, he can press the "BACK" button to return to the previous stimuli.

There are two other tick boxes on the GUI for the case where the subject has difficulties in reporting the perceived image positions. In the literature, it has been reported that subjects in a similar listening test may perceive 'dual' images, especially when the interaural phase difference given by the target ITD is ambiguous [31]. For this reason, a subject could choose to report two locations by activating the "Report Dual Images" option. Meanwhile, it is also likely that subjects may not be able to report their perception at all, for which cases "Can't decide the position" option allowed subjects to skip the current trial. Subjects could take a break or even terminate the test at any time by pressing the "PAUSE" or the "STOP" buttons, but otherwise, a 5-minute break was given in every 10 minutes.

It is noteworthy that, during a session, the target ITD has been varied in an increasing order for each target ILD, and the position of the acoustic pointer has not been refreshed between trials. There can be some concerns about possible bias associated with the deterministic order of stimulus presentation, but, considering the limited time for each session, such an approach was inevitable. This issue will be dealt with in the discussion of the test results in section 4.3.2.

In a separate session, the performance in 'virtual localisation' has been tested at 600 Hz for all 6 participants, which has been arranged to investigate the subjects' accuracy in adjusting the acoustic pointer. For the left-right balance, the acoustic pointer signals corresponding to -50°, 60°, -70°, 80° and -90° have been employed as target signals. In other words, subjects were instructed to judge the intracranial position of these selected acoustic pointers with reference to the acoustic pointers themselves. Each target signal has been randomly presented for 5 times during a total of 25 (5×5) trials. The result

of this virtual localisation test as well as the main test data regarding lateralisation will be presented and discussed in the following section.

This experimental study was approved by the Safety and Ethics Committee of the Institute of Sound and Vibration Research (ISVR), University of Southampton (Approval number: 755).

4.3 Test results

4.3.1 Virtual localisation of acoustic pointers

The data acquired in the 'virtual localisation' test have been converted to be positive (remember that target positions have been distributed to balance the left and the right), and the averages and the 95% confidence intervals have been computed for each subject as shown in Fig. 4.4. As the dashed line indicates the perfect reference localisation, the responses made for the range of target angles are observed to vary from subject to subject, but it appears that the signals corresponding to the source locations from 50° to 70° are either well localised (SA and SB) or overestimated (SC and SF) whereas the angular locations of pointer signals for 80° and 90° have been mainly underestimated (SA, SB, SD and SF). In particular, the subject SD consistently reported that the target signal is located closer to the median plane than it actually is, while the subject SE mostly overestimated the target source position.

Having examined the individual performance in adjusting the position of the acoustic pointer, the overall responses from the virtual localisation test can be presented as a histogram and an error-bar plot as shown in Fig. 4.5. To draw this 3D histogram, subjective responses have been pooled to four bins per 10°, and the relative count in each bin has been indicated by the grey-level scale. At 50°, subjects have given perfectly matching or very close responses to the target angle. However, the virtual localisation task became more difficult for the target positions closer to 90°. Interestingly, it has been found that, over 60°, it seems that subjects tend to report with the maximum possible pointer location near 90°, regardless of the actual target positions. Accordingly, the average responses for the 60° and 70° target locations have been greatly increased, which becomes, however, less prominent as the target angle approaches to 90°.

As denoted by the error-bar plot in Fig. 4.5, the overall averages with the confidence intervals confirm the discussion presented above for the individual subjects that the target position is overestimated below 70°, and underestimated above 80°. In addition to the 'direction' of the localisation error with respect to the reference target locations, the amount of error also varies, increasing with the target angle, which is similar to the observations made in the real-source localisation test reported in the literature [4, 38, 39]. In fact, the less accuracy at more lateral positions is probably an issue not only with the HRTF-based acoustic pointer but also with the other types of pointers, and the test results presented above suggest that the current type of acoustic pointer can be effectively used to give an objective indication of the perceived image location to within reasonable accuracy. The errors and the variabilities of the subjective responses in

the following main test results can be analysed in relation to the virtual localisation performance.

4.3.2 Lateralisation of dichotic pure tones

First of all, it is noteworthy that during the main tests regarding the lateral position of the ITD/ILD-manipulated pure tone signals, there was no sign, in general, that subjects had difficulties in the matching task using the acoustic pointer provided. A few particular target signals have been reported to be relatively difficult to make judgements for, but it was probably not because the HRTF-based acoustic pointer was unreliable, but because the intracranial locations for those signals were ambiguous, and thus hard to define, which is also reported in the literature in relation to the dual images.

600-Hz target signals

For the 600-Hz tests, all data acquired during the 5 sessions have been collected including those corresponding to the dual images as shown in Fig. 4.6. Data have been marked differently for each session, where the sample means and the 95% confidence intervals are also denoted as error-bars. As blue, green and red colours indicate the target ILDs of -6, 0 and 12 dB, respectively (this colour coding scheme will be used consistently in this chapter), the general patterns of the lateral position judgements can be observed and compared with the results of the listening test reported by Sayers [31] (see Fig. 2.5 in section 2.3). When the ITD is zero, auditory images are perceived at those positions shifted according to the amount of the target ILD, and as the ITD decreases/increases from zero, the images migrate to the side favoured by the given ITD. When the ITD and the ILD conflict with each other, the image makes a sudden shift to the contralateral side at a certain **critical ITD** value. This sudden transition is observed approximately at $-800 \sim -600\,\mu s$ and $600 \sim 800\,\mu s$ for 0 dB, $+400 \sim +600\,\mu s$ for -6 dB and $-500 \sim -300\,\mu s$ for 12 dB, and, despite the inter-subject variability, the absolute value of the critical ITD appears to consistently decrease when the absolute ILD increases [29, 31], as reported in the literature.

In the ranges of the target ITD where the sudden shift takes place, it is also observed that the response angles have greater variabilities than for other target conditions. This increased uncertainty in the subjective judgement can be perhaps related to the lower accuracy of the HRTF-based acoustic pointer at lateral angles as mentioned in section 4.3.1. However, it is more likely that the inherent uncertainty of the actual image locations, the so-called dual images, resulted in the greater variability. Although subjects except for SE and SF have never reported that they have perceived two distinctive

images, a detailed inspection of the responses in each plot in Fig. 4.6 reveals that the responses in those transitional ranges of the target ITD are often divided roughly into two groups where the first group continues the monotonic increase/decrease of the laterality whilst the second indicates that the images shifted to the other side of the ear. Such a division is particularly prominent with the test data obtained with the subject SA where, in his data shown in Fig. 4.6(a), the critical ITD for the contralateral transition varies from session to session. Accordingly, the subjective responses, particularly those for the 12-dB ILD, can be grouped into two, which resulted in greater variabilities for the target ITDs nearby the critical values.

As hinted in the above paragraph, the subjects SE and SF did report the presence of the dual images a few times, and their responses in those cases are separately presented in Fig. 4.8 with the mean responses for each target ILD. The 'double' responses are interconnected and denoted by the same markers for each trial, so that the link between the two may be easily observed. It is obvious that, as an experienced subject, SE has made some judgements that clearly indicate the characteristics of the dual images discussed above. However, although the subject SF has reported two positions for certain target signals, it is unclear whether he actually perceived them, since the distance between each pair of images is very close as shown in Fig. 4.8(b), and he even produced two identical responses for 0 dB at $-600\,\mu s$ and for -6 dB at $1000\,\mu s$. Nevertheless, it is still possible to connect these multiple judgements made by SF to the increased diffuseness of the perceived auditory images.

Four out of six subjects have skipped some of the target signals without judgements, reporting that the auditory images created by the signals are too vague to estimate their loci. Fig. 4.9 illustrates those target conditions for which subjects felt too hard to make judgements. On top of the global sample means from all subjective responses, the initial of the subject who made 'no response' is denoted followed by the number indicating the frequency. For example, 'F2' marked on the red line at $-400\,\mu s$ means that the subject SF skipped the target condition of 12-dB ILD and -400-μs ITD twice during the 5 sessions. It is interesting to see that most of the no-response target conditions are found around where the dual images have been reported with greater variability. It is also remarkable that subjects skipped certain target signals consistently in the course of the sessions, as can be seen by the numbers greater than 1 in Fig. 4.9. Since there was no indication of the test progress displayed on the GUI, the consistent report of 'no response' is not related to the possibly biased judgement, but only to the nature of the auditory images. (Subject SD skipped all 5 trials for the first target condition of ITD $= -1000\,\mu$s and ILD $= -6$ dB, which is, however, possible to be associated with subjective bias.) To summarise, the arguments made so far regarding the greater

variability, the presence of the dual images and the skipped target signals are all related to the heavily diffused auditory images that are found nearby the critical ITDs.

Returning to the discussion on the individual laterality judgements shown in Fig. 4.6, it is further suggested that the given range of the acoustic pointer has been insufficient for the subjects SE and SF as shown in panels (e) and (f). In particular, their responses for some target conditions with 12-dB target ILD are converged at -90° when the absolute value of the target ITD is greater than $\sim 600\,\mu s$, where both subjects also verbally reported that the perceived locations of some target signals were out of the range covered by the pointer signal. Therefore, any exceptionally low variance, especially near the boundary of the pointer, should not be regarded as reflecting the true statistics in the test data shown in Fig. 4.6.

In addition to the issue of the limited range of the acoustic pointer, it should be recalled that the deterministic order of the stimulus presentation was also of concern in the test design. In each trial in a session, participants could first listen to the target signal and the pointer signal where the pointer signal indicated their own judgement made for the previous target signal which was sometimes only very slightly different from the new target. As can be seen in Fig. 4.6(d), the relevant bias effect is most prominently observed in the data acquired for the subject SD when ITD is less than $-200\,\mu s$ for 12-dB ILD. In this range of the target conditions, the subjective response is found to vary gradually, drawing identifiable trajectories connecting each set of unique markers indicating different sessions. In order to investigate how the randomisation of the stimulus order could have affected the result, it is necessary to carry out additional listening tests for comparison, which are, however, considered to be beyond the scope of the current study.

The overall sample averages and the 95% confidence intervals for the laterality test at 600 Hz are shown in Fig. 4.7. Features found and discussed above regarding the individual data are also confirmed by the global statistics.

The current study of the subjective perception of the dichotic tone is in particular aimed at the investigation of the judgement of the auditory image position, which will be later compared to the predictions given by the CC model introduced in chapter 2. Therefore, it is important, as a first step, to examine whether the individual data presented in Fig. 4.6 are really distinctive from each other. Although the data look significantly different from subject to subject, especially in terms of the critical ITDs and the range of laterality, it is inappropriate to draw any definitive conclusion only by visual inspection. However, the one-way unrelated **analysis of variance (ANOVA)** and the **t-test** can be employed to give a statistical statement as to whether those samples have originated from different populations [60]. Since both statistical tests are based on the assumption

that the samples under investigation are from normal distributions, the **Lilliefors test** [61] has been additionally implemented to check the normality of the acquired data in advance.

Fig. 4.10(a) shows the result of a series of statistical tests on the listening test data. Similar to the display scheme used for Fig. 4.9, each number on the colour-coded curves of the global sample means represents the number of subjects whose data were NOT rejected by the Lilliefors test at 5% significance level for their normality. For many target conditions across the ranges of ITD and ILD, these numbers are equal to or greater than 5, but for some other target conditions, e.g. ITD = 12 dB and ITD = $-1000\,\mu$s, the null hypothesis has been rejected more often. Such a greater rejection rate (denoted by 3 or 4 in Fig. 4.10(a)) is particularly common for the test conditions with the target ITD nearby the critical value, and as a matter of fact, it is expected that the null hypothesis of the data normality is rejected for more target conditions in this range of the target ITD if the subjective responses are truly indicative of the presence of the dual images which are often associated to a bimodal rather than a bell-shaped unimodal distribution. It is considered that the greater variability caused by the limited sample numbers in each target condition is perhaps responsible for the unexpectedly high rate to satisfy the Lilliefors test. (Note that four valid samples are the minimum requirement for the Lilliefors test in Matlab 7.0 whereas the number of the current test samples was only 5.)

The unrelated one-way analysis of variance has then been applied only to those data that passed the normality test. Assuming that the laterality responses for different combinations of the target disparities are distinguished from each other, only the subject was considered as the variable of the ANOVA. In other words, the ANOVA has been applied across subject for each target condition, for which individual participant made 5 to 10 responses (including dual-image responses). In Fig. 4.10(a), some target conditions have been marked by a circle, for which the null hypothesis of the ANOVA that the listening test data have no significant inter-subject difference is NOT rejected. In other words, for those unmarked target conditions, the individual test data show significant differences from subject to subject, between at least a pair of subjects. As a majority of the target conditions are not marked, the subjective laterality judgements can be considered to be unique for each person within the scope of the applied statistical test, which possibly confirms one of the hypotheses established in this study regarding the individuality of human spatial hearing.

Having found that the listening test data from different subjects can not be considered to share a common population mean and variance, it is now of further interest to see specifically which pairs of subjects are found to be different in their data. Accordingly,

the **multiple comparison procedure** [60] has been carried out for each target condition for the data shown to be from a normal distribution. The output of this analysis can be regarded as a list of subject pairs whose data have been tested to be statistically different from each other. Then, if the number of appearances in this list is counted for each subject across the test condition, a bar graph shown in Fig. 4.10(b) can be plotted, from which the degree of the uniqueness in each subjective data can be approximated. As the ordinate labelled as the 'distinction index' is the percent ratio of the number of appearances to the total counts, it is obvious that the responses made by the subject SA are mostly distinguished, while the other 5 subjects show relatively equivalent degrees of uniqueness in their data.

1200-Hz target signals

The result of the current listening tests at 1200 Hz is shown in Fig. 4.11 for the two selected subjects SA and SF. Similar to the result at 600 Hz shown in Fig. 4.6, there are some noticeable features in the data at 1200 Hz: the periodicity of the response angles with respect to the target ITD, the earlier transition to the contralateral side with the greater absolute ILD and the more variabilities around the critical ITDs. Compared to the 600-Hz case, the period of the laterality has been halved as expected, and the range of the response angles have been also reduced. As for the subject SA, for example, the mean responses at 600 Hz were between -20° and 75°, -50° and 75°, and -90° and 20° for -6, 0 and 12 dB, respectively, which are now at 1200 Hz only between -5° and 50°, -20° to 30°, and -70° and 0° for the same target ILDs.

From the comparison between Figs. 4.7 and 4.11(c) in terms of the overall statistics, it is interesting to note that the 95% confidence intervals for the 12-dB ILD have been significantly increased at 1200 Hz while those for the 0- and -6-dB ILDs have been little changed. Perhaps, the difference in the number of samples at each test frequency (5 for 600 Hz but 3 for 1200 Hz) could have affected the variabilities, but it is not clear why the 12-dB laterality judgements have been influenced more prominently.

Since there are only two participants in the test at 1200-Hz, the t-test has been simply applied instead of the ANOVA to test the null hypothesis stated as "the test data acquired from the different subjects originate from an identical population." In contrast to the result of the ANOVA at 600 Hz shown in Fig. 4.10(a), the listening test data at 1200 Hz have been found to be relatively similar between the two subjects as shown in Fig. 4.12, although the null hypothesis has been rejected for some test conditions which can be associated with the target ITDs close to the critical value. However, it is noteworthy that the normality test has been skipped for the 1200-Hz result due to the shortage of the data samples (remember that the Lilliefors test requires at least

4 samples), and therefore, the result of the t-test presented in Fig. 4.12 has to be appreciated only conservatively.

4.4 Results of model predictions

For the test variables employed for the listening tests presented in the previous section, numerical simulations have been implemented to obtain predictions from the decision-making model based on the characteristic curve introduced in section 2.2. As the (distal-region) HRTF database measured in chapter 3 has been used to establish the characteristic curve for each participant, the model parameters have been set identical to those employed in section 2.3: the scaling factors k_τ and k_α are 44 μs and 1 dB, respectively, while the standard deviations of the internal errors, σ_δ and σ_ε are 10 μs and 1 dB. The source signal used for the listening test has been also considered as the input to the model, but the interval between the target ITDs has been reduced for a better resolution from 100 μs (listening test) to 50 μs, and therefore, there are 41 and 29 target conditions for each ILD at 600 Hz and 1200 Hz, respectively. A total of 500 predictions (iterative model runs) have been made for each target condition while the internal errors were varying according to two independent zero-mean Gaussian distributions. This simulation has been coded and implemented in Matlab 7.0.

600-Hz target signals

Fig. 4.13 shows the simulation results at 600 Hz, where the contrast of each point indicates the relative count of the model prediction at a certain response angle (ordinate) and a target ITD (abscissa) following the colour-coding scheme used in the previous section (blue for −6 dB, green for 0 dB and red for 12 dB). On top of this vertical view of the 3D histograms, the sample averages and the 95% confidence intervals have been also displayed as error-bars.

Features discussed in section 2.3 can be found in the result of the current simulations, which include the periodic pattern of the laterality judgements, the dual images in the vicinity of the critical ITDs and the earlier shift to the contralateral side for a greater target ILD. From a visual inspection, the mean responses shown in Fig. 4.13 appear to be similar across subjects, and they also look similar to the simulation result presented in Fig. 2.5 in section 2.3 where the characteristic curve from the KEMAR HRTF has been used. The inter-subject similarity is mainly found for the target ITD between −200 μs and +200 μs where the target ITD is relatively away from the critical ITD, but beyond this range, the model predictions start to show some subtle differences between subjects. Particularly for 0-dB ILD, one of the bimodal responses corresponding to the dual images is dominant over the other, which, after averaging, results in the inter-subject difference in the mean response.

As was the case with the analysis of the listening test result, the ANOVA has been applied in order to investigate whether the model predictions are unique for each subject. The normality test has been first carried out, but this time, the **chi-square goodness-of-fit test** [60, 61] has been employed which is efficient in dealing with samples of frequency data (note that in the chi-square goodness-of-fit test, the minimum count in each bin is 5 as a rule-of-thumb, not suitable to handle the listening test data in section 4.3.2). Similar to the plotting scheme applied to Fig. 4.10(a), the number of subjects is marked in Fig. 4.14(a) along the global mean of the simulation data, for whom the null hypothesis of the data normality is not rejected at 5% significance level. Compared to the listening test data, it is observed that a smaller number of the simulated data are normally-distributed, which probably resulted from the asymmetry of the data with respect to the response angle [see Fig. 4.17(b)]. Such an asymmetry was hardly recognisable either visually or statistically for the listening test data. This was due to the shortage of the samples per each test condition, which, however, became prominent in the simulation result with a large number of samples. It is also noted that the failure rate of the normality test is relatively high when the target ITD is close to the critical ITD, as particularly shown by the result of the statistical test for the 0-dB ILD.

Having screened the simulation result with the normality test, the following ANOVA for the selected data showed that the model predictions for the participants cannot be regarded as originating from a common population, where the null hypothesis is rejected for all test conditions as indicated by the absence of circles in Fig. 4.14(a). This seeming paradox between visual and statistical observations for the model predictions presented in Fig. 4.13 is possibly attributed again to the large number of iterations which perhaps facilitated the statistical test to better differentiate the subtle difference in mean and variance.

Meanwhile, in the same approach taken for the analysis of the listening test data, the multiple comparison procedure has been carried out to perform a series of pairwise comparisons between individual model predictions, producing the bar graph of the distinction index as shown in Fig. 4.14(b). Similar to the result shown for the listening test data in Fig. 4.10(b), the distinction index for the subject SA stands out in the simulation result, where the other individual indices are also reasonably comparable. If the distinction index can be regarded as representing the uniqueness of the data as assumed in this study, its reasonable consistency found in the simulation result and the listening test data provides an indication that the decision-making model based on the characteristic curve successfully reflects the individual perceptual process of spatial hearing.

While a full comparison between the listening test data and the model simulation is postponed until section 4.5, a further statistical analysis is presented below. Since the samples are required to be from a normal distribution, the scope of the t-test, the ANOVA and the multiple comparison procedure have so far been limited. However, if there are a sufficient number of samples under investigation, the **chi-square statistic** [60, 61] can be a good measure to compare multiple groups of data regardless of the specific type of distribution. Considering that the minimum count required in each bin is 5 as a rule-of-thumb, the current simulation result is qualified for the use of the chi-square statistic where 500 repetitions have been made for each condition.

Fig. 4.15 shows the result of the chi-square statistic for the individual model predictions at 600 Hz, where the plotting scheme used in Fig. 4.9 has been also applied such that the two letters at each target condition indicate the pairs of the subjects' initials whose data have been found to be similar to each other at 5% significance level. For instance, at $-800\,\mu s$, the simulation data for SC & SD, SC & SE, SC & SF and SD & SF have been found to be similar pairwise for the target ILD of 12 dB (red). In general, it is obvious that the simulation results have been rarely found to be similar between subjects, and this observation can be regarded as reconfirming the inter-subject uniqueness of the current decision-making model which was only partially supported by the ANOVA result shown in Fig. 4.14. On the other hand, there are some test conditions, where the null hypothesis is not rejected for a tested pair, and it is interesting to note that those conditions are mainly found when the target ITD is relatively distant from the critical ITDs, where such a link between the degree of the data similarity and the critical ITD is observed throughout the target ILDs.

1200-Hz target signals

Similar to the procedures followed for the 600-Hz target signals, the model predictions for the perceived image locations have been simulated at 1200 Hz. The 1200-Hz characteristic curves have been obtained for the subjects SA and SF from their distal-region HRTFs measured in chapter 3, and the model has been prepared with the parameters identical to those employed for the 600-Hz simulation.

The individual model predictions at 1200 Hz are shown in Fig. 4.16 where blue, green and red colours have been again used to represent the data for the -6, 0 and 12 dB ILDs, respectively. The periodic nature of the laterality is easily noticed with a period of approximately half the value found for the 600-Hz simulation, and the ranges of the model responses are also observed to be reduced compared to the previous simulation at the lower frequency. In addition, the simulation data at 1200 Hz appear to be slightly

more spread than at 600 Hz, implying greater variabilities, as the contrast of each point indicates the relative frequency of the responses.

The simulation data have been further investigated using the statistical tests introduced for the analysis of the 600-Hz result. However, as implied by many zeros in Fig. 4.17(a), the normality of the model predictions has been rejected for most of the target conditions, and therefore, any following analysis using either the t-test or the ANOVA is not considered to be meaningful. As discussed for the simulation result at 600 Hz, the failure in the normality test can be attributed to the asymmetry in the model responses. For instance, Fig. 4.17(b) illustrates a histogram depicting the model predictions for one of the test conditions, where it appears to be almost bell-shaped by visual inspection, but fails the chi-square goodness-of-fit test probably due to the slight slant towards the left side.

The comparison between the individual model predictions at 1200 Hz has been finally made by the chi-square statistic which is able to operate regardless of the normality of the data, and it has been shown that the simulation results for the two participants can not be regarded as originating from a common population as the null hypothesis has been rejected for all test conditions.

To summarise the simulation results in this section, the model predictions both at 600 Hz and 1200 Hz have been statistically shown to be relatively unique for each subject for a majority of the test conditions, although some similarities could be found especially when the target ITD was distant from the critical value. In addition, some principal features found in the listening test data could be also observed in the model simulation, while the comparison between the two results will be made in detail in the next section.

4.5 Comparison between test results and model predictions

It should be recalled that the participants' task in the current listening test was to match the perceived image location of the pointer tone to the target tone where there were 181 pointers available corresponding to the angular range of the KEMAR HRTFs from -90° to +90° at every 1°. Therefore, a subjective judgement represented by one of the acoustic pointers at, say, θ_{kem} means no more than the fact that both target dichotic tone and the pointer tone have been perceived to be roughly at the same position, while θ_{kem} can not be directly associated with the subject's own HRTFs. On the other hand, considering that the result of the model simulation is given with respect to the subject's own characteristic curve, the model response, say, θ_{sbj} is indicative of the azimuth angle corresponding to his own HRTFs. Consequently, in order to make a sensible comparison between the listening test data and the model predictions, it is necessary to convert one of the two results, either mapping θ_{kem} to θ_{sbj} or vice versa. The function relating θ_{kem} to θ_{sbj} can be obtained by headphone listening tests where participants report the perceived angular position of the binaural signal convolved with the KEMAR HRTF, obviously with reference to their own auditory space (not to the acoustic pointer created by the KEMAR HRTF as implemented in section 4.3.1). However, such an empirical investigation was unavailable in this study.

Alternatively, the current decision-making model can be utilised to numerically estimate the function mapping θ_{kem} to θ_{sbj}. In Fig. 4.18(a), the 600-Hz characteristic curve (subject SF) is shown by the thin line with various markers, along with the KEMAR's characteristic curve between -90° and +90° (thick solid and dashed line). Considering each point on the KEMAR's characteristic curve as a series of target signals (corresponding to the azimuth angle θ_{kem}), the current matching scheme can find a nearest-neighbour on the subjective characteristic curve, and thus give the azimuth angle, θ_{sbj}. Figs. 4.18(b) and (c) show the mapping functions between the KEMAR and the subjective characteristic curves at 600 Hz and 1200 Hz, respectively. These results produced by the CC model predict that the location of the target signal given by the KEMAR HRTF will be underestimated by most of the subjects, especially when the target angle is greater than 50°. For example, the binaural pure tone signal at 600 Hz created by the KEMAR HRTF at 90° is possibly perceived by the subject SF to be incident from 70°. It is recalled that during the listening test some subjects have reported that the range of the acoustic pointer could not cover the spatial extent of the presented target signal, for which the mapping function depicted in Figs. 4.18(b) and (c) might be able to give a reasonable explanation.

600-Hz target signals

Only after the angular conversion discussed above, the results of the listening test and the simulation can be compared for each subject as shown in Fig. 4.19, where the mean responses and the 95% confidence intervals of the subjective judgements have been plotted as error-bars along with the mean of the model predictions. It is observed that the agreement between the simulation and the test results is especially good when the target ILD is 0 dB, where the range of the subjective response and the critical ITD values have been successfully predicted by the model. Also for other target ILDs, there are some test conditions that have been relatively predicted well by the model, and those conditions are mostly found when the target ITD is away from the critical ITD. For example, the right tails of the subjective responses for the 12-dB target ILD are reasonably matched to the model predictions, and the central parts of the laterality data for the -6-dB ILD are also found to be relatively consistent between the two results. On the other hand, most of the discrepancies between the model and the subjective judgements can be found around where the sudden image shift takes place. Particularly for the nonzero ILDs, the critical ITDs predicted by the model are, in absolute value, much less than those suggested by the listening test data, and the period of the transition is relatively short in the model responses, shaping sharp edges of the curves. It is obvious that these differences in the transitional phase resulted in the significant disagreement between the subjective judgements and the model predictions as shown in Fig. 4.19.

Considering the large number of the simulation data far exceeding the number of the subjective judgements, it is reasonable to assume the averages of the model predictions as the population means, based on which the t-test can be implemented for a further comparison. As the means of the simulation data have been marked by either $\circ$ (not rejected) or $\times$ (rejected) in Fig. 4.19, the t-test has been applied to examine whether, for each target condition, the mean of the model predictions is found within the 95% confidence interval given by the subjective judgements. As a result, the 'success rate' has been plotted in Fig. 4.20 which shows the relative count of the successful model predictions for each target ILD, where blue, green, red and black colours have been coded for the three target ILDs and the overall average. As expected from the visual inspection discussed above, the agreement between the model predictions and the listening test data is very encouraging when the target ILD is zero. The usual range of the success rate has been found to be between 30% and 60%, while the success rate for the subject SA appears to be below the average. It is obvious that this comparative analysis provides an insight to the predictive scope of the current decision-making model, but the arguments made above should not be regarded as being conclusive, since the results of the normality tests should have been considered before the implementation of the t-test.

Accordingly, the result of the comparison has been redrawn in Fig. 4.21 only for those test conditions where both model predictions and subjective test data have been found to be from a normal distribution. In Fig. 4.21, the thick error-bars indicate the statistics of the subjective judgements for those selected test conditions, while the means of the model predictions are marked by either ∘ (not rejected) or × (rejected). Since one or both of the two results were rejected for their normality, most of the test conditions for -6 dB were disqualified for the comparison, and there are also only a few data points remaining for the other target ILDs, depending on the subject. The success rates for the 'qualified' test conditions have been recalculated as presented in Fig. 4.22 where the agreement between the model predictions and the subjective test data appears to be slightly improved, which mainly resulted from the reduced number of available samples. For example, there are only two qualified test conditions for -6 dB in the data for the subject SC [see Fig. 4.21(c)], and in this case, the corresponding success rate is found to be 100% in Fig. 4.22.

1200-Hz target signals

A similar comparison has been made for the results of the listening test and the model simulation at 1200 Hz as presented in Fig. 4.23. It should first be recalled that, both in the subjective test and the model simulation, there were no 'qualified' test conditions at 1200 Hz in terms of the data normality, and accordingly, the result of the comparison shown in Fig. 4.23 can be only a rough indication of the performance of the model prediction, similar to the argument made for Fig. 4.19.

The agreement between the subjective judgements and the model predictions is noticeable at 1200 Hz when the target ILD is 0 dB and 12 dB, and the predictive scope of the current model seems to be better at this higher frequency compared to the result at 600 Hz shown in Figs. 4.19 and 4.21. As the agreement for the 12 dB target ILD has been particularly improved, this observation is also confirmed by the success rate plotted in Fig. 4.23(c) where the image positions for up to 75% of the test conditions have been predicted well by the model for the subject SF at 12 dB. Despite the very encouraging result of the comparative analysis, such an improvement at 1200 Hz has to be carefully interpreted only after considering the greater variabilities of the subjective judgements at 1200 Hz when compared to the the lower frequency case, which widened the confidence interval significantly, thus, giving better chances for the model predictions to be found therein.

There are also test conditions at 1200 Hz where the two results have been found inconsistent, and as was the case at 600 Hz, the target ITDs for these conditions were close to the critical values. As can be seen by the × markers far off from the 'main stream' data

in Figs. 4.23(a) and (b), the model predictions appear to be significantly misleading for the target ITDs from $-50\,\mu s$ to $300\,\mu s$ with -6-dB ILD and for those from $-150\,\mu s$ to $100\,\mu s$ with 12 dB, and undoubtedly, these ranges of the target ITDs associated with the poor success rates are repeated according to the signal period.

Cross-comparison

Having examined the agreement between the listening test data and the simulation result using the CC model customised for each participant, a further 'cross-comparison' can be carried out in order to investigate whether the prediction of the individual model is truly unique for the subject. Accordingly, the judgements made by a subject are not only compared to the predictions of his own model, but also to those given by the other individual models, and the average success rate provided by the t-test can be obtained in each case. In Fig. 4.24, the result of the cross-comparison is presented as a group of graphs for each subject, where each grey-scaled bar indicates the success rate of the t-test with reference to one of the six models. If the current decision-making model reflects the individual difference in the subjective perception, one of the six bars that corresponds to his own model is expected to stand out among the others. The result for the subjects SA, SB and SC at 600 Hz are encouraging as shown in Fig. 4.24(a) in which the subjects' own models gave predictions better than, or at least equivalent to the others. However, as the results for the other subjects do not meet the expectation, it is also arguable whether the difference between the highest bar to the second highest is statistically significant to distinguish one model from the others. Similarly, it is not certain whether the cross-comparison made for SA and SF at 1200 Hz [see Fig. 4.24(b)] can be indicative of the unique link between the subject and his own hearing model.

It is recalled that in sections 4.3.2 and 4.4, the inter-subject similarities in both listening test data and simulation results have been found, if they existed, mostly when the target ITD was relatively far away from the critical ITDs, in which range the data samples were relatively stable with only small variances. Therefore, if it is to be argued that the model is uniquely established after each subject's individual auditory space, it should be demonstrated probably for the test conditions near the critical ITDs where the inter-subject difference is assumed to be maximal. However, due to the greater diffuseness of the auditory images in that range of the target ITD, it is expected to be difficult to estimate the exact distribution of the subjective judgements even with the increased number of samples. It therefore seems unlikely that the unique relation between the model and the subjective perception can be easily established.

General discussion

To summarise the results of the comparison presented so far, first, it is uncertain whether the model predictions are unique for each subject as illustrated by the analysis of the cross-comparison. Nevertheless, the decision-making model based on the characteristic curve has made reasonable predictions of many of the test conditions examined in the current listening test both at 600 Hz and 1200 Hz. The agreement between the two results was especially remarkable for the 0-dB target ILD, and for the rest of the test conditions, the global patterns of the model predictions were also relatively consistent with those of the subjective test data. However, the point-to-point comparison of both visual and statistical inspections showed that the model can give a misleading indication of the subjective judgements, particularly for those test conditions with target ITDs around the critical values.

The relatively poor performance of the CC model around the critical ITDs can be understood in relation to the characteristics of the nearest-neighbour matching process. For example, the primary and the secondary characteristic curves at 600 Hz are shown in Fig. 4.25, where the target conditions and their matched positions on the characteristic curves have been marked as circles and triangles, respectively. As the target ITD increases from $-1000\,\mu s$, the triangle moves from one point to the other on the characteristic curve continuously, and at the critical ITD that depends on the target ILD, it is 'transferred' to the next characteristic curve. On the contrary, the search for the nearest-neighbour at 1200 Hz never requires the whole range of the characteristic curve as shown in Fig. 4.26. This is particularly due to the shorter interval between the critical ITDs that is equivalent to the signal period, and this is the very reason for the reduced ranges of the laterality compared to the lower frequency result. In addition, depending on the target ILD, some parts of the characteristic curves have been systematically skipped even before the model response is transferred to the next characteristic curve. For instance, the model predictions for the -6-dB target ILD shown in Fig. 4.26(a) have been made mainly around the two regions of each characteristic curve leaving the area in-between 'unused,' whereas at 0 dB and 12 dB [panels (b) and (c)] one continuous part of the characteristic curve has been 'scanned' for the matching process. Such a discontinuous use of the characteristic curve at 1200 Hz is associated with the hook-shaped end which resulted from the multiple number of the local peaks and troughs in the ILD function as depicted in Fig. 4.27(a). If the two turning points of the 1200-Hz characteristic curve had been shaped differently, for example, as straight as that for 600 Hz, the overestimation around the critical ITDs particularly for non-zero ILDs [see Figs. 4.23(a) and (b)] could have been avoided (remember that the turning points of the characteristic curves roughly correspond to $\pm 90°$). This would improve the degree of agreement between the model predictions and the subjective test data.

Since the characteristic curve is defined in the ITD–ILD space that is assumed by the model to reflect the individual auditory space, reshaping the curve to improve the model's predictive scope inevitably requires redefining the whole ITD–ILD space. Therefore, within the framework of the current model, employing new scaling factors k_τ and k_α may be the first appropriate attempt to alter the shape of the characteristic curve. Although it is difficult to empirically define the scaling factors, k_τ and k_α, they can be tentatively assumed to increase with the absolute value of the ITD and the ILD, indicating the dependence of the neural sensitivity in a similar way the function $p(\tau)$ has been established by Stern and Colburn [11]. Fig. 4.27 shows an example of how the ILD-dependent scaling factor can deform the characteristic curve. In panel (a), the ILD function at 1200 Hz (subject SA) is displayed, where two distinctive local minima and maxima can be found for the right and the left hemispheres, respectively. If k_α is now substituted with a function increasing with the absolute ILD [see Fig. 4.27(b); refer to the caption for the details of the temporary function $k_\alpha(\alpha)$], then those peaks become less prominent as shown in Fig. 4.27(c), and therefore the 'bent ends' of the characteristic curve are unfolded, which might give model responses better predicting the subjective judgements.

As a matter of fact, the diagram shown in Fig. 4.28 reflects the authour's hypothetical picture regarding the neural selectivity in the ITD–ILD space where the smaller and more densely populated 'cells' indicate finer neural resolution. In addition to the neural sensitivity decreasing with increasing absolute ITD and ILD, the importance of the natural combination of the ITD and the ILD can be found as the size of the cell becomes smaller for a pair of ITD and ILD closer to the characteristic curve, which is consistent with the discussion made in section 2.5.2 regarding the implication of the model to the image diffuseness.

The above arguments regarding the tentative use of an ILD-dependent scaling factor and the authour's hypothesis regarding the neural sensitivity are only to illustrate the flexibility of the current model and its predictions in accordance with the future findings in the relevant neurophysiological/psychoacoustical studies, and should not be regarded as suggesting a certain form of neural structure in a systematic approach.

Consequently, it is considered that a further improvement of the current model, especially the modification of the characteristic curves in a newly transformed ITD–ILD space is beyond the scope of the current investigation, but may be dealt with in future work. Nevertheless, the above result of the comparison between the model predictions and the subjective data is remarkable, considering that the current model, based on a simple assumption and procedure, can give reasonable predictions for the intracranial

position of dichotic pure tones consistently at two different frequencies with an identical set of model parameters.

4.6 Conclusion

In this chapter, the result of the listening test and the simulation have been presented and discussed regarding the perception of the laterality created by dichotic pure tones. Being very naturally appreciated by listeners, the acoustic pointer provided by the non-individual HRTF has been found to be effective as a reference tone, and it is suggested that the current paradigm of the position matching task can be further investigated in an attempt to reveal the link between the intracranial and the extracranial auditory images.

The result of the listening test has been found to be qualitatively consistent with the experimental studies reported by Sayers [31] and Domnitz and Colburn [29], where the variability of the subjective judgements increases around the critical ITDs that correspond to the sudden shift of the image position to the contralateral side, often related to the dual images. In addition, the listening test data for each different subject have been found to be close to unique for many target conditions, and the relevant statistical tests have been also applied to the simulation result to confirm the uniqueness of the individual models associated with the HRTFs.

In the comparative analysis, the agreement between the subjective judgements and the model predictions has been found to be reasonable for many target conditions, whereas the discrepancy could be observed mostly for the target ITDs close to the critical values. Assuming the data from the simulation as individual populations, the t-test provided an indication that the model predictions have been successful for up to 67% and 76% of the test conditions depending on the target ILD at 600 Hz and 1200 Hz, respectively. However, the following cross-comparison demonstrated that it is uncertain whether the judgements of the image laterality can be better predicted by the subject's own model, mainly due to the large variance in the listening test data, especially around the critical ITDs.

From the analysis of the actual nearest-neighbour matching procedure, it has been suggested that the current model is possibly improved by adopting new scaling factors, $k_\tau(\tau, \alpha)$ and $k_\alpha(\tau, \alpha)$. Depending on both ITD and ILD, the new scaling factors can transform the entire ITD–ILD space to reshape the characteristic curves in a certain auditory frequency band, which has been reasonably demonstrated by a tentative function of $k_\alpha(\alpha)$.

While the simplicity and the flexibility of the current model are notable, the reasonable agreement between the listening test data and the simulation result is remarkable considering that such an agreement has been observed across subjects at the two test frequencies with frequency-independent model parameters.

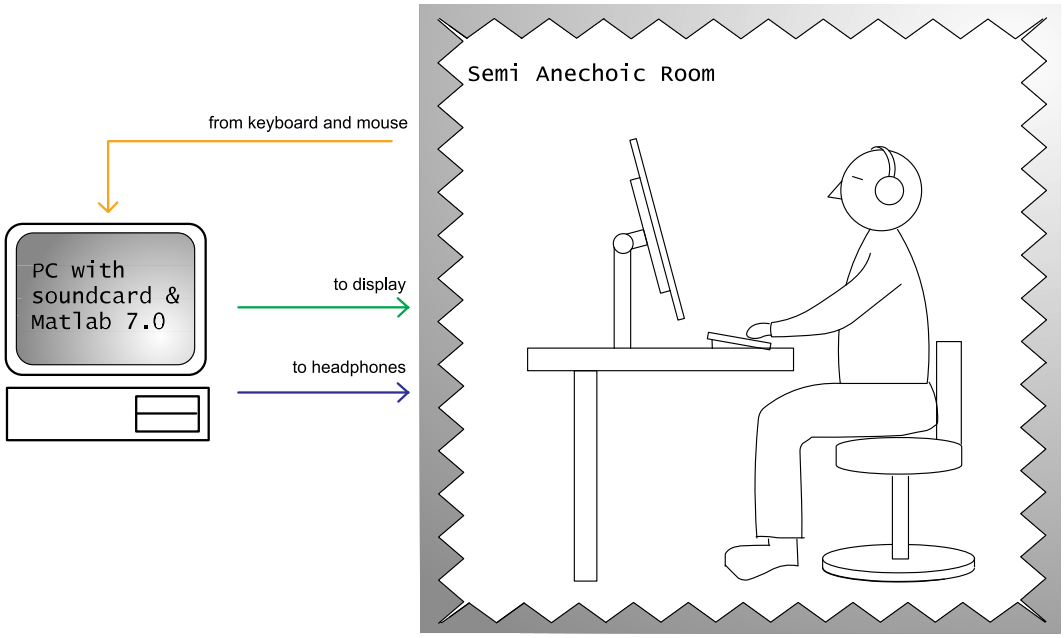


FIGURE 4.1: The current listening test has been carried out in a semi-anechoic room. A desktop PC with a soundcard has been used to generate test signals where subject performed the matching task using the graphic user interface shown on the display.

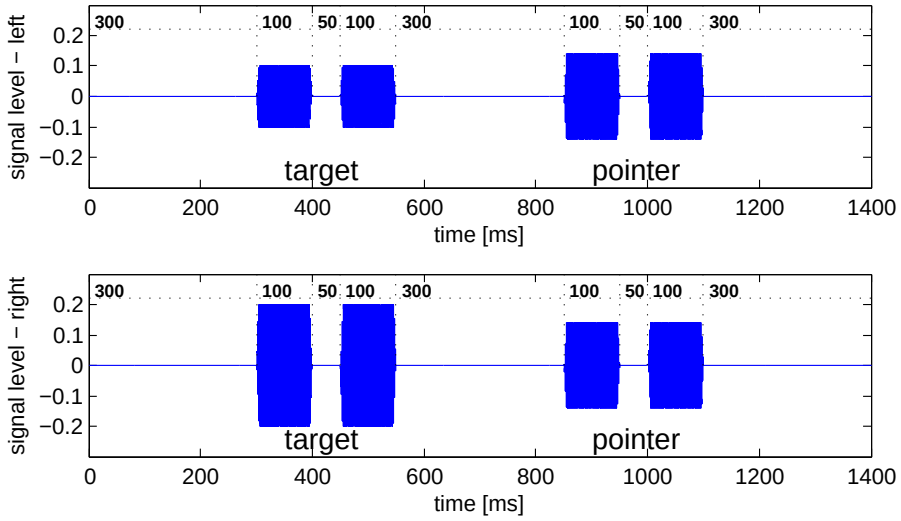


FIGURE 4.2: Test signals are shown for each channel. The first two pulses are the target signals, while the last two are the acoustic pointer signals. Numbers above the dotted line indicate the duration in *ms*.

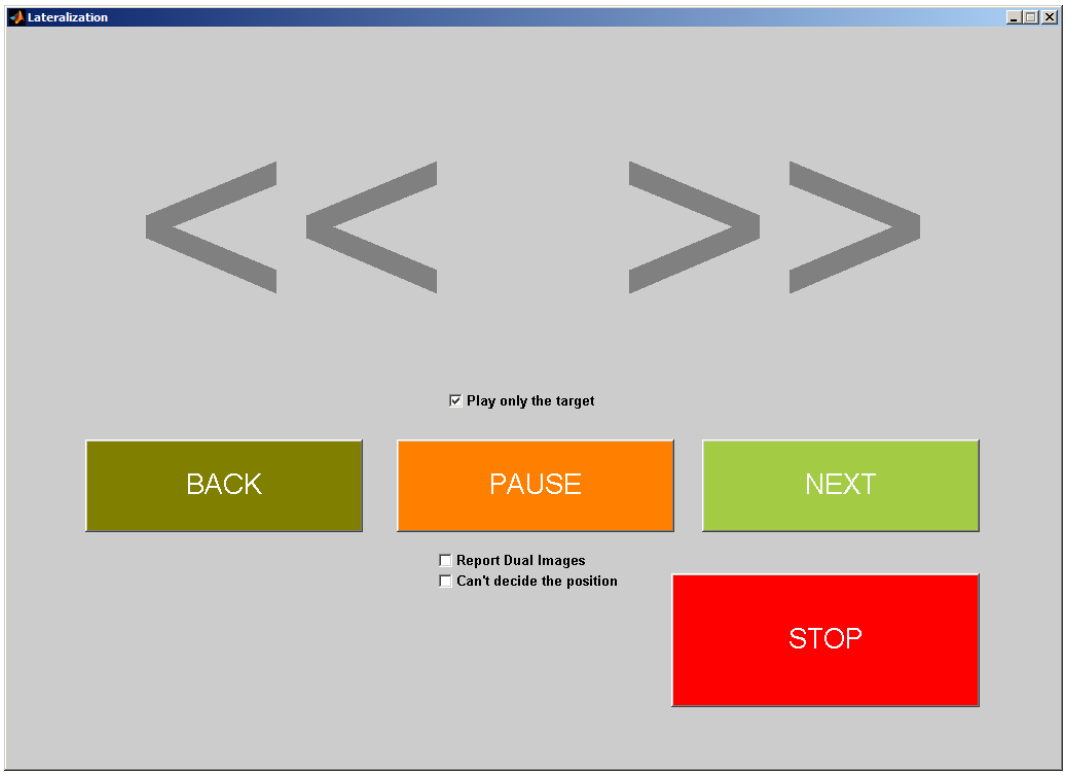


FIGURE 4.3: The graphic user interface used in the listening test is shown. Subject has a full control of the test where he can move from trial to trial, pause and even stop the session. The position of the acoustic pointer is controlled by the left/right arrow keys, and on each press, the large arrow heads shown on the GUI blink. Also, three options are available for the subject: "Play only the target," "Report Dual Images," and "Can't decide the position."

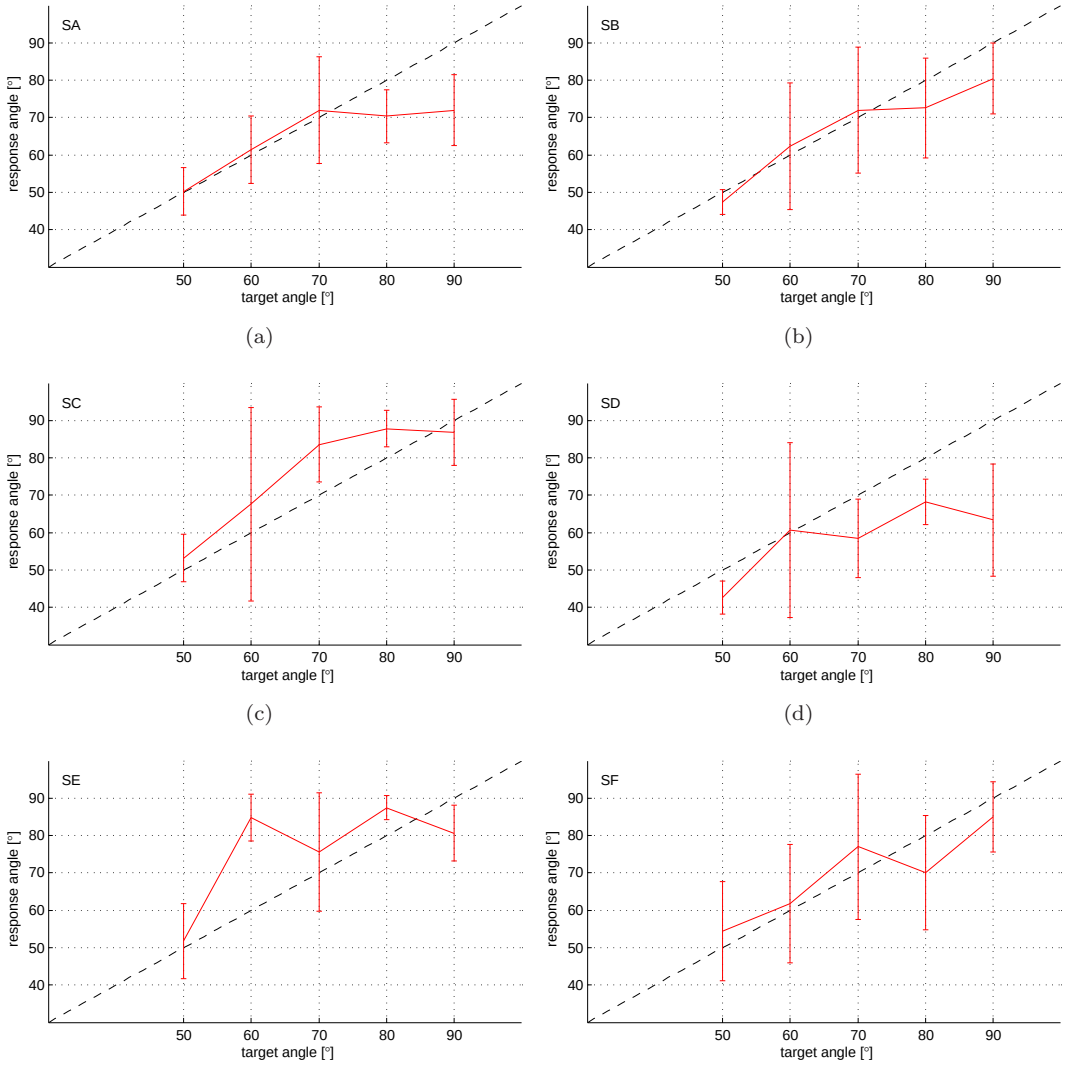


FIGURE 4.4: The results of the virtual localisation test are shown for the participants of the lateralisation listening tests. Subject's initials are shown on the top left corner of the plots, where the error-bars indicate the average responses and the 95% confidence intervals.

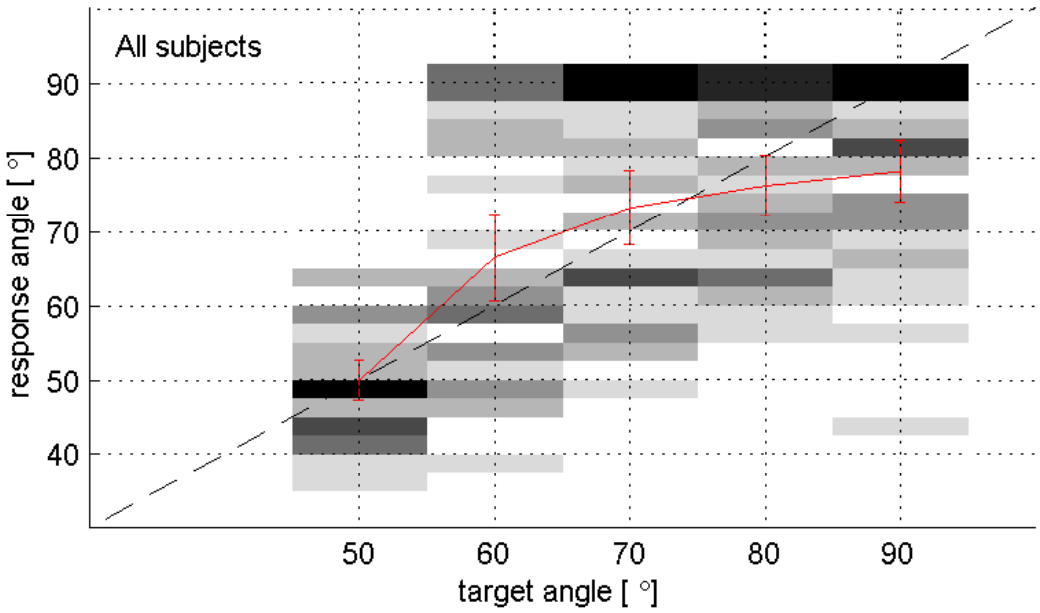


FIGURE 4.5: All responses in the virtual localisation test are shown where the grey-scale level indicates the relative count of the responses in each bin (the darker, the more counts), and the error-bars represent the average responses and the 95% confidence intervals.

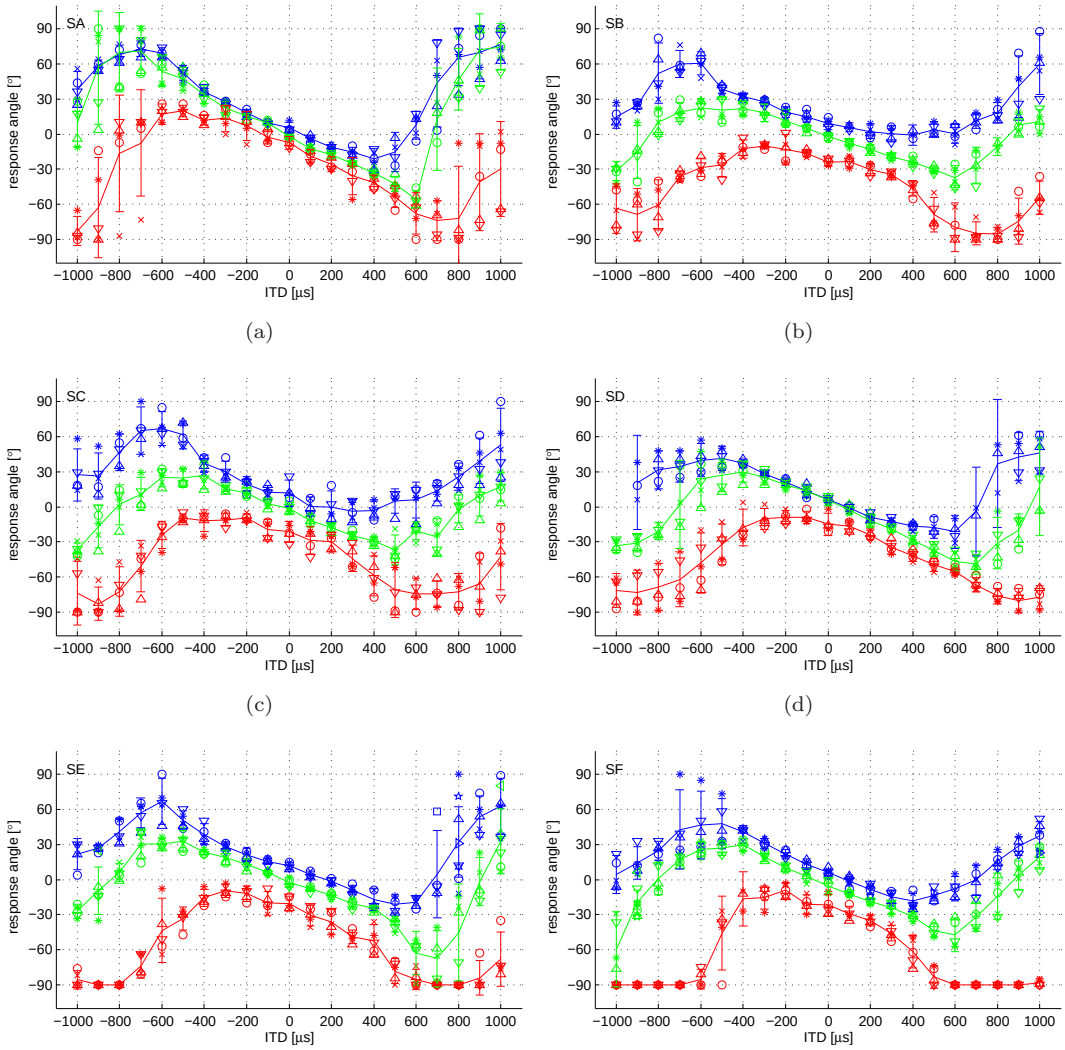


FIGURE 4.6: The results of the laterality test at 600 Hz. Subject's initials are shown on the top left corner of the plots, and blue, green and red colours indicate -6, 0 and 12-dB target ILDs. Markers are used uniquely for different sessions, where the error-bars represent the mean responses and the 95% confidence intervals.

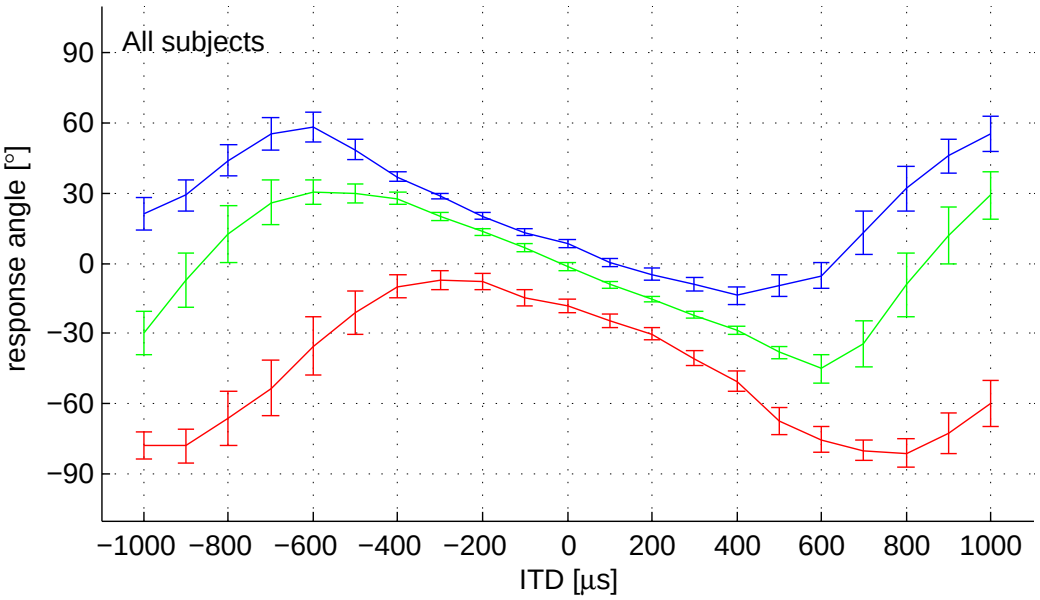
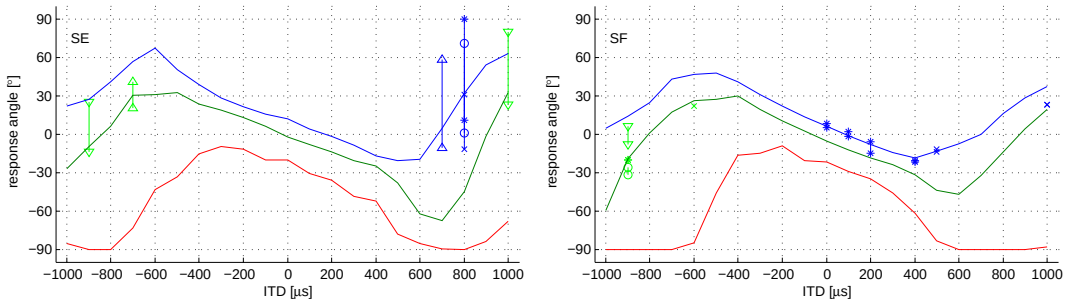


FIGURE 4.7: Overall statistics of the subjective judgements shown in Fig. 4.6. Blue, green and red colours indicate -6, 0 and 12-dB target ILDs, where the error-bars represent the mean response and the 95% confidence interval.



(a) (b)

FIGURE 4.8: Dual images reported by (a) SE and (b) SF are shown on top of the average responses (blue, green and red for -6, 0 and 12-dB target ILDs, respectively). The two responses corresponding to the dual images are connected by line, where markers have been used uniquely for different sessions.

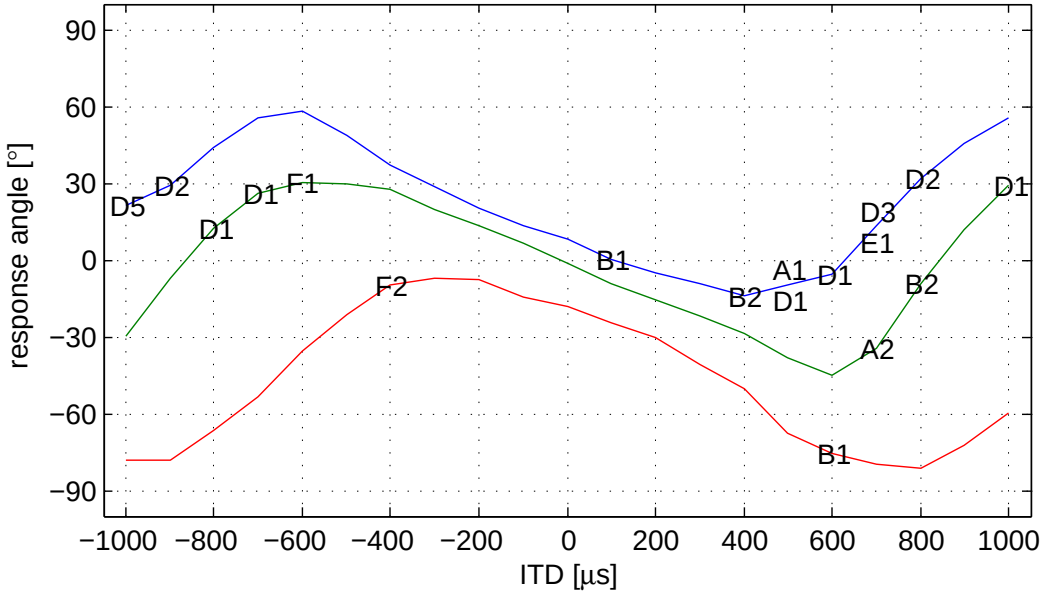
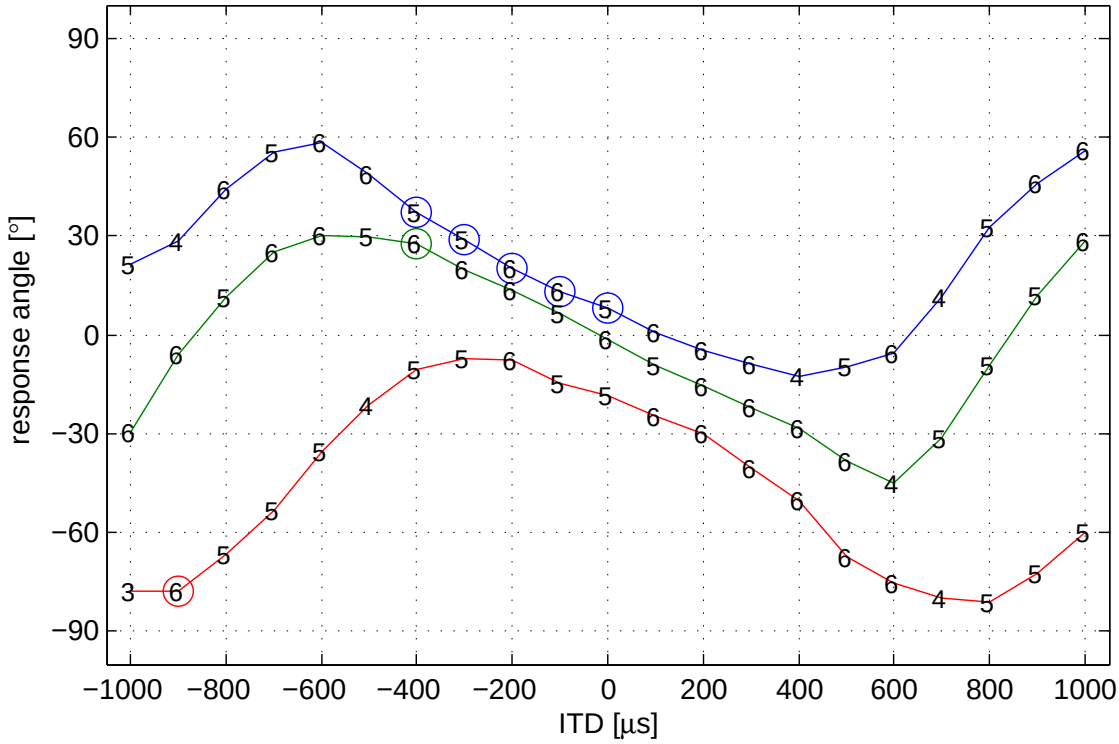
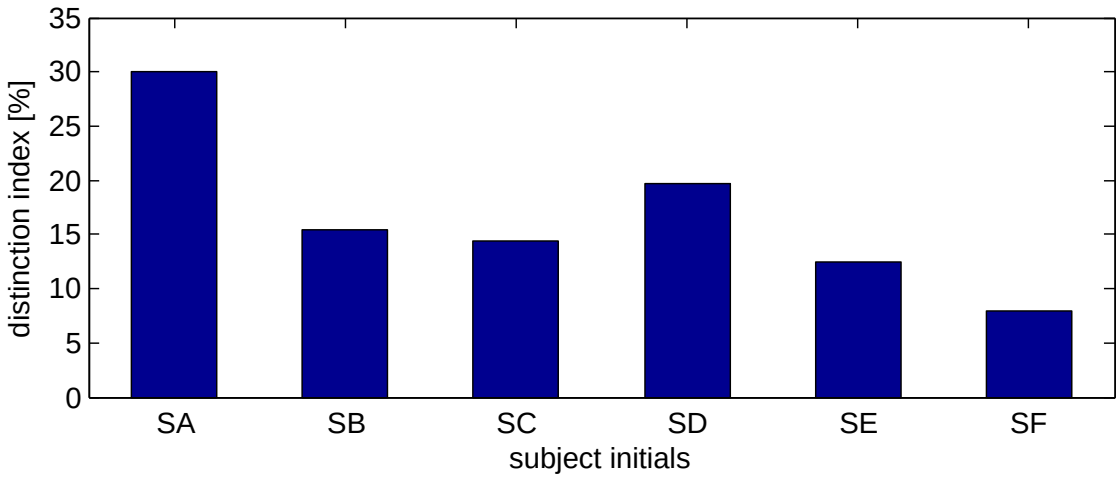


FIGURE 4.9: Test conditions where subject has made no judgement are shown. For example, the subject SF skipped the test condition of -400-μs ITD and 12-dB ILD twice during the 5 sessions.

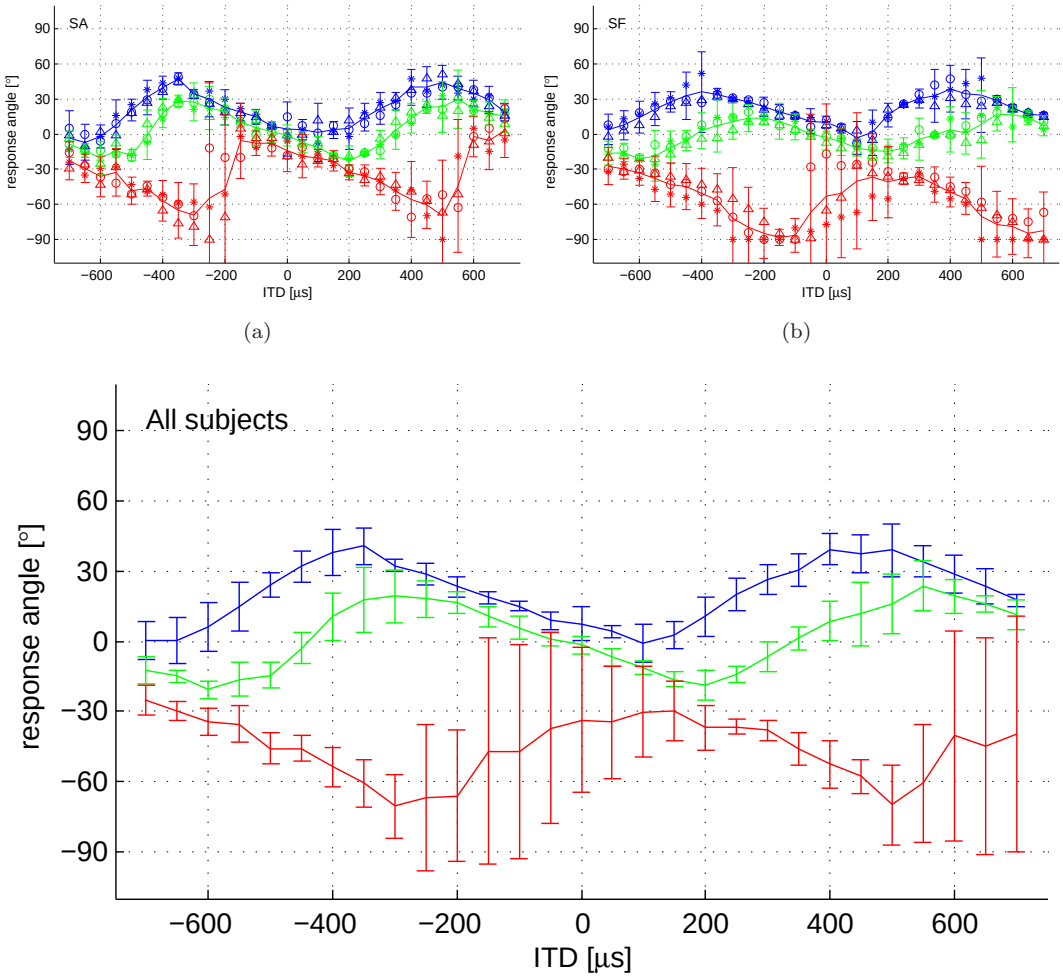


(a)



(b)

FIGURE 4.10: (a) On top of the average responses, the number of subjects is shown for each target condition, whose data have NOT been rejected by the normality test. The circles for certain test conditions indicate where the listening test data have been found by the ANOVA to be statistically similar between subjects. (b) The result of the multiple comparison procedure is represented as the distinction index. (Refer to the text for the definition of the index.)



(a) (b)

(c)

FIGURE 4.11: The results of the laterality test at 1200 Hz. Subject's initials are shown on the top left corner of the plots, and blue, green and red colours indicate -6, 0 and 12-dB target ILDs. Markers are used uniquely for different sessions, where the error-bars represent the mean responses and the 95% confidence intervals. Panel (c) shows the global statistics.

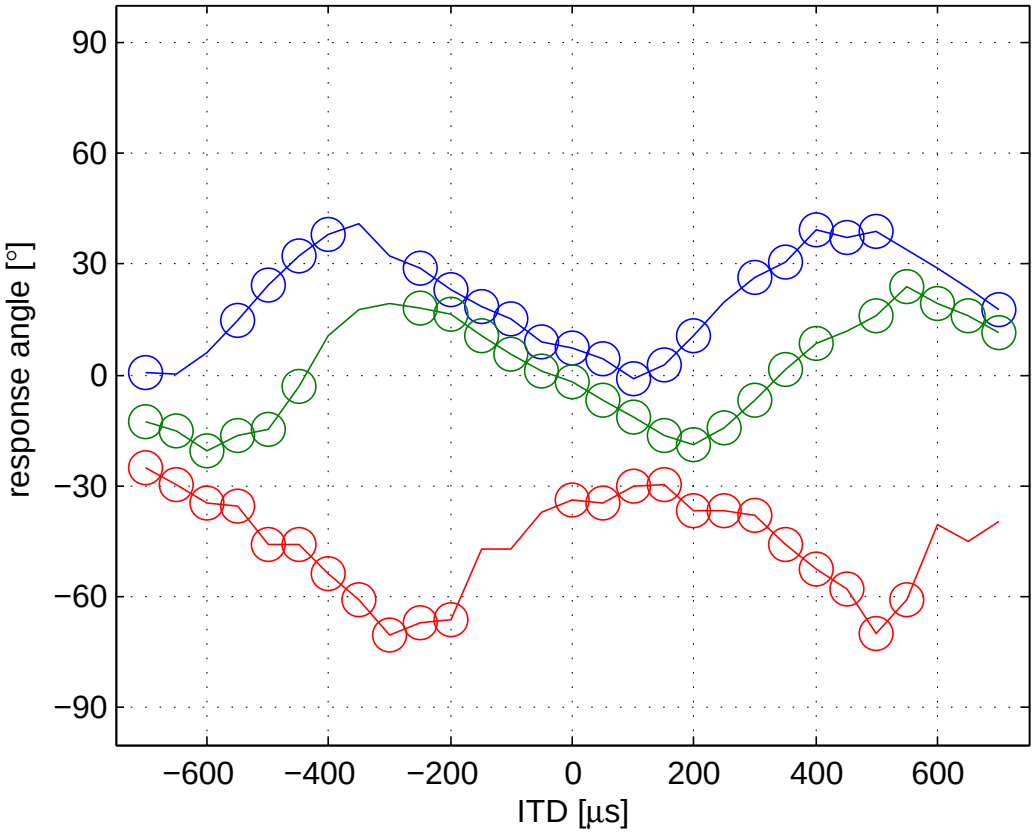


FIGURE 4.12: The result of the t-test is shown. The circles for some test conditions indicate where the listening test data of the two subjects, SA and SF have been found to be similar by the t-test.

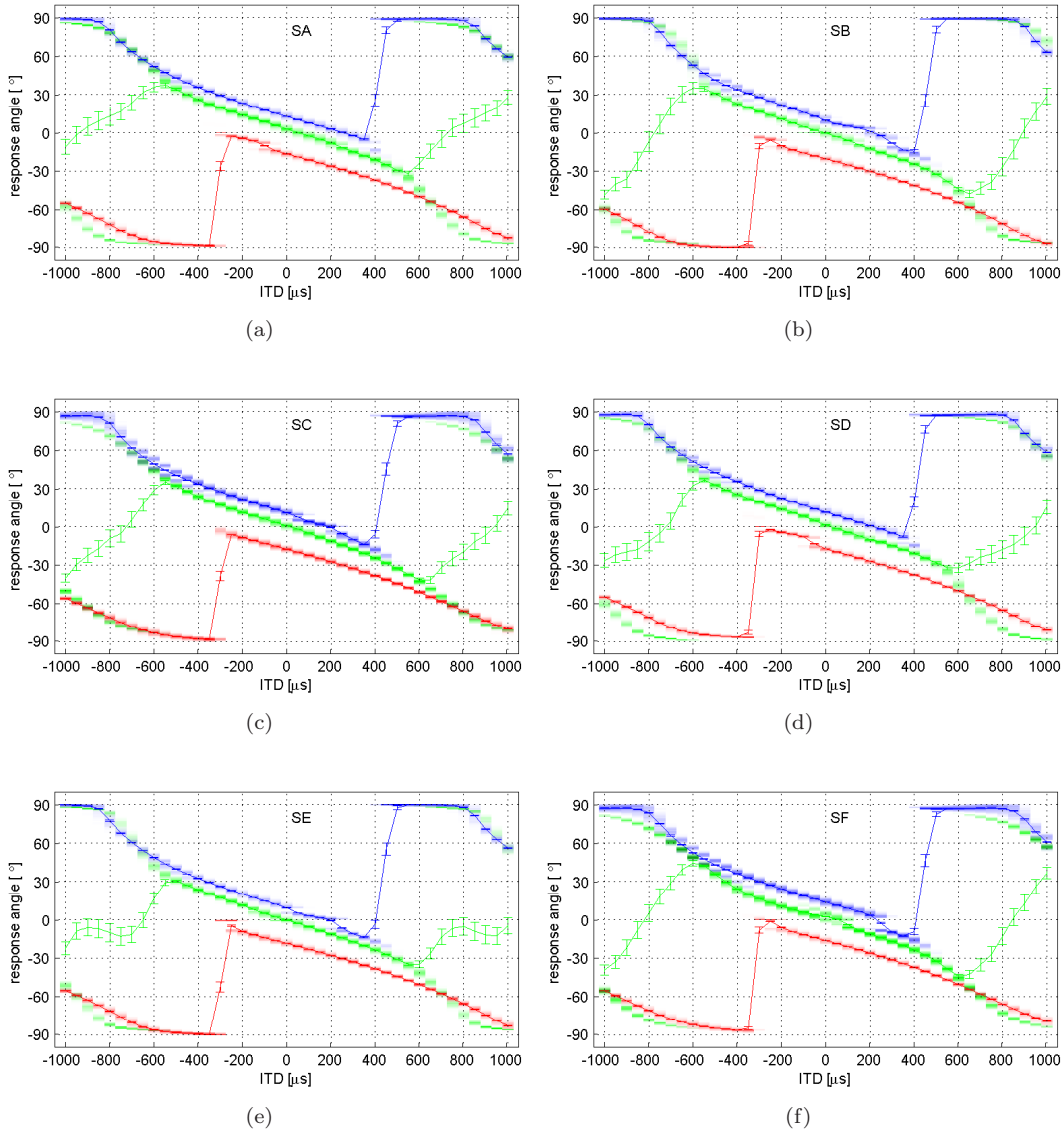
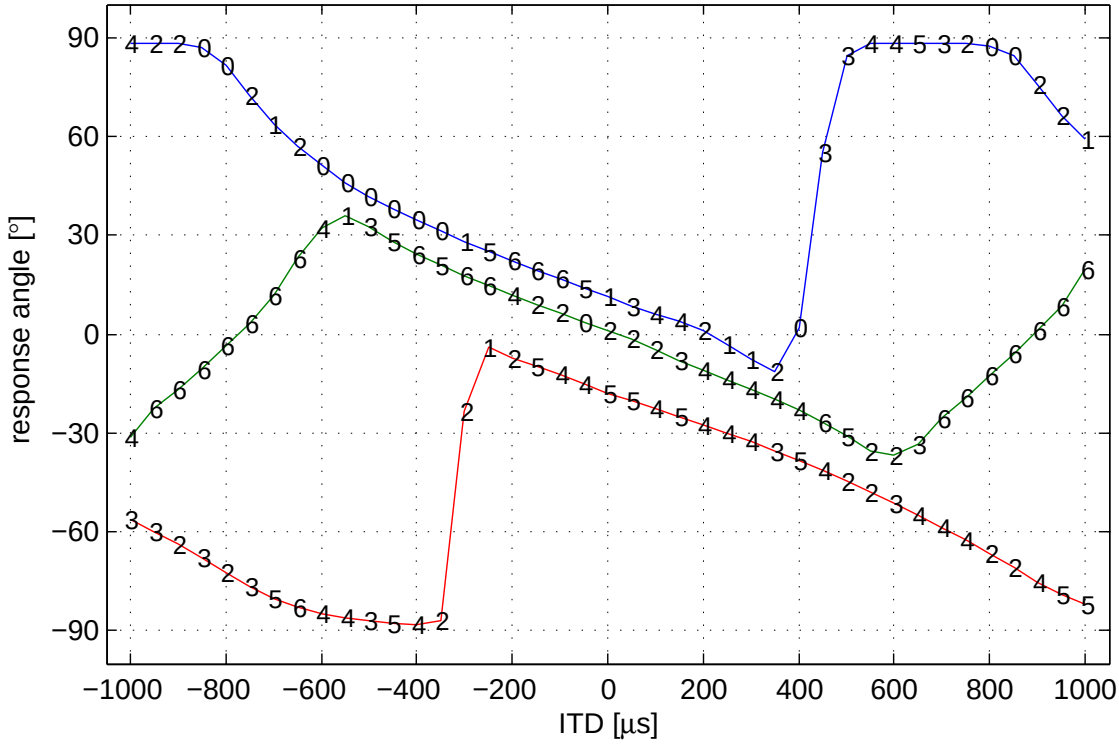
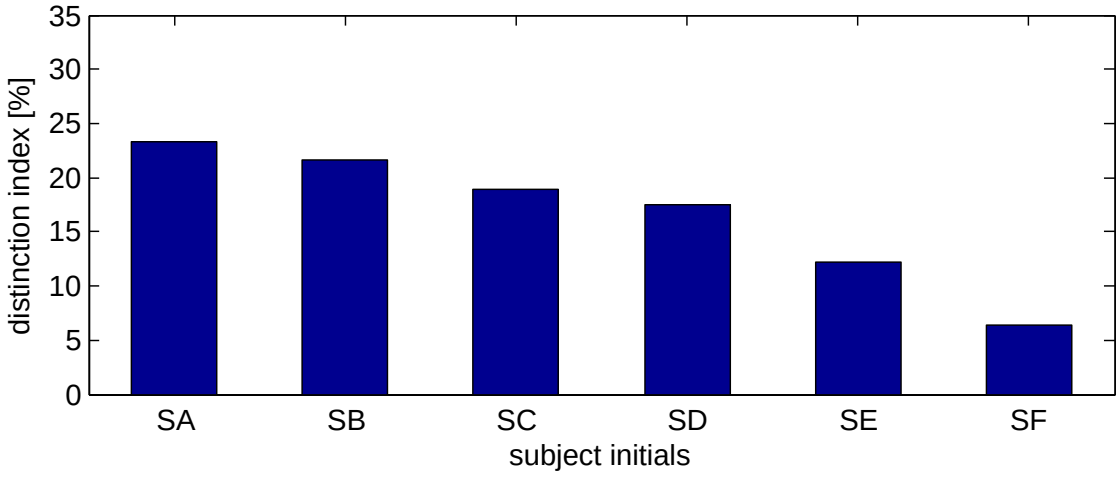


FIGURE 4.13: The predictions of the individual CC models are shown at 600 Hz. Subject's initials are shown on the top centre of the plots, and blue, green and red colours indicate -6, 0 and 12-dB target ILDs. The contrast of each colour indicates the relative count of the model responses in each bin at every 1°, where the error-bars represent the averages and the 95% confidence intervals.



(a)



(b)

FIGURE 4.14: (a) On top of the averages of the model predictions at 600 Hz, the number of subjects is shown for each target condition, whose simulation data have NOT been rejected by the normality test (blue, green and red for -6, 0 and 12-dB target ILDs, respectively). No statistical similarity has been found by the ANOVA between model predictions. (b) The result of the multiple comparison procedure for the model predictions is represented as the distinction index. (Refer to the text for the definition of the index.)

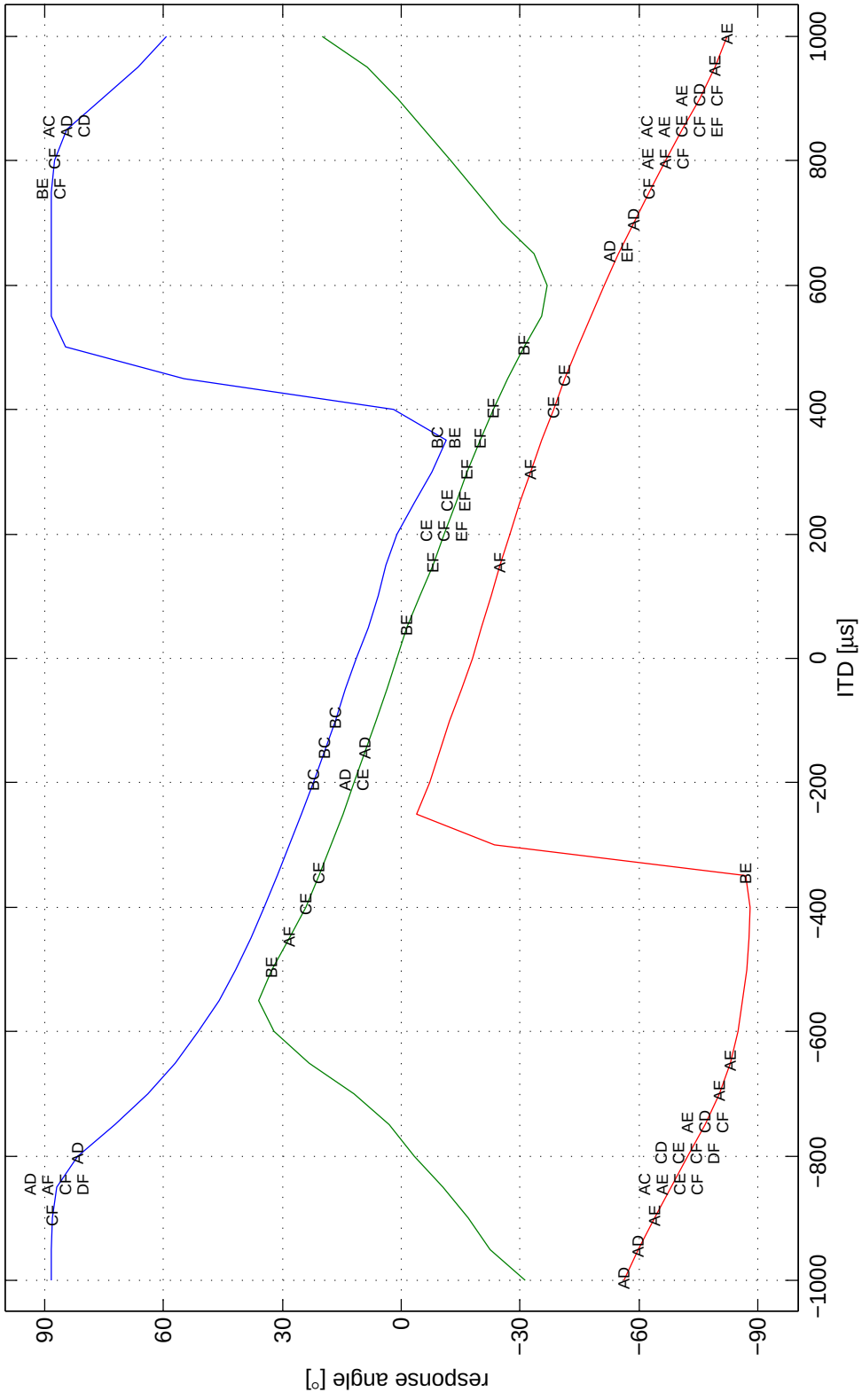
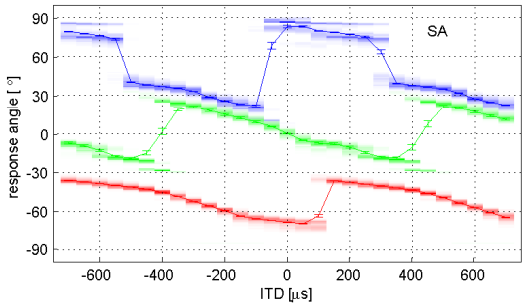
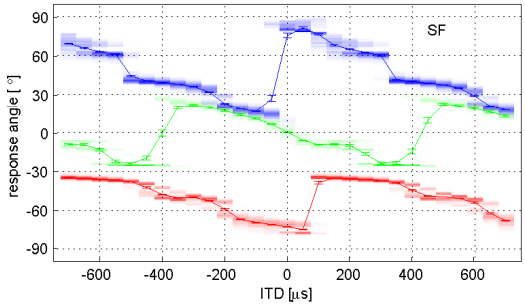


FIGURE 4.15: The chi-square statistic further shows that the model predictions are mostly unique, where for some conditions, similarities between pairs of the subjects have been observed (blue, green and red for -6, 0 and 12-dB target ILDs, respectively). For example, the model predictions between SC and SD, SC and SE, SC and SF, and SD and SF have been found similar for -800-μs ITD and 12-dB ILD.

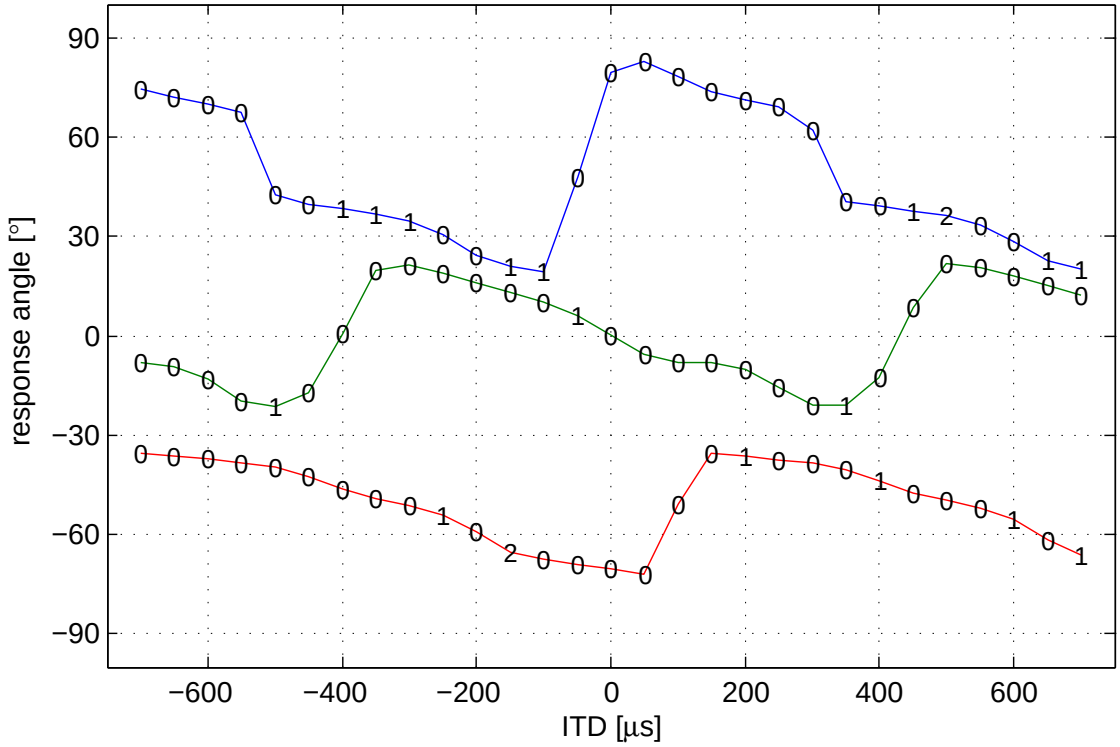


(a)

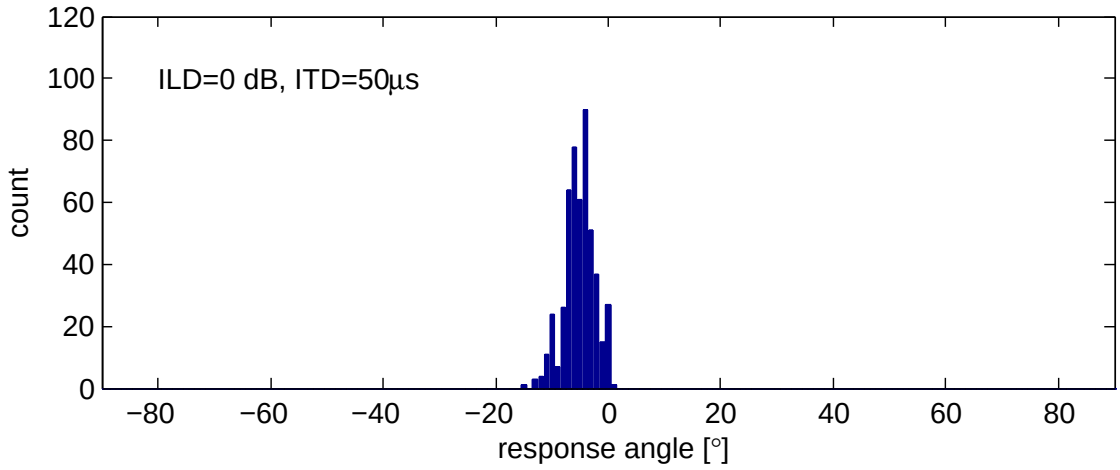


(b)

FIGURE 4.16: The predictions of the individual decision-making models are shown at 1200 Hz. Subject's initials are shown on the top right of the plots, and blue, green and red colours indicate -6, 0 and 12-dB target ILDs. The contrast of each colour indicates the relative count of the model responses in each bin at every 1°, where the error-bars represent the averages and the 95% confidence intervals.

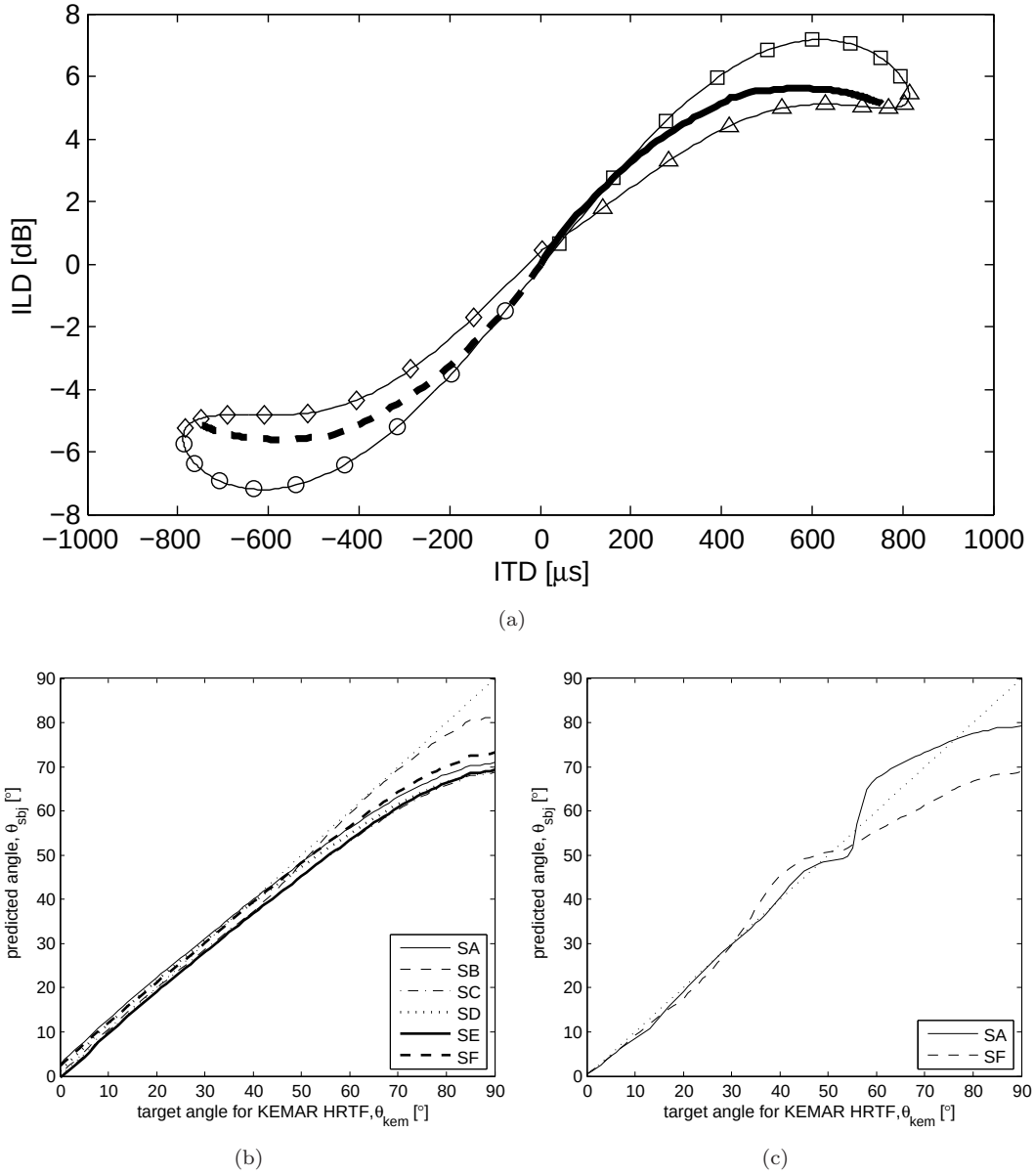


(a)



(b)

FIGURE 4.17: (a) On top of the averages of the model predictions at 1200 Hz, the number of subjects is shown for each target condition, whose simulation data have NOT been rejected by the normality test (blue, green and red for -6, 0 and 12-dB target ILDs, respectively). No statistical similarity has been found by the t-test between model predictions. (b) A histogram of the model predictions for 50-μs ITD and 0-dB ILD is shown as an example. The asymmetry of the model predictions appears to be visually insignificant, where the chi-square goodness-of-fit test rejected the data normality.



(a)

(b)

(c)

FIGURE 4.18: (a) The 600-Hz characteristic curve for the subject SF is marked at every 10° (◇: 0° ∼ 80°, ○: 90° ∼ 170°, □: 180° ∼ 260°, △: 270° ∼ 350°). The characteristic curve obtained from the KEMAR HRTF is also shown in the range of the azimuth angle between -90° and +90° (the dashed line for the positive angles, while the solid for the negative). The mapping functions relating the azimuth angle for the KEMAR HRTF to that corresponding to the participants' HRTFs are shown at (b) 600 Hz and (c) 1200 Hz. These functions have been obtained from the model predictions. (Refer to the text for the details.)

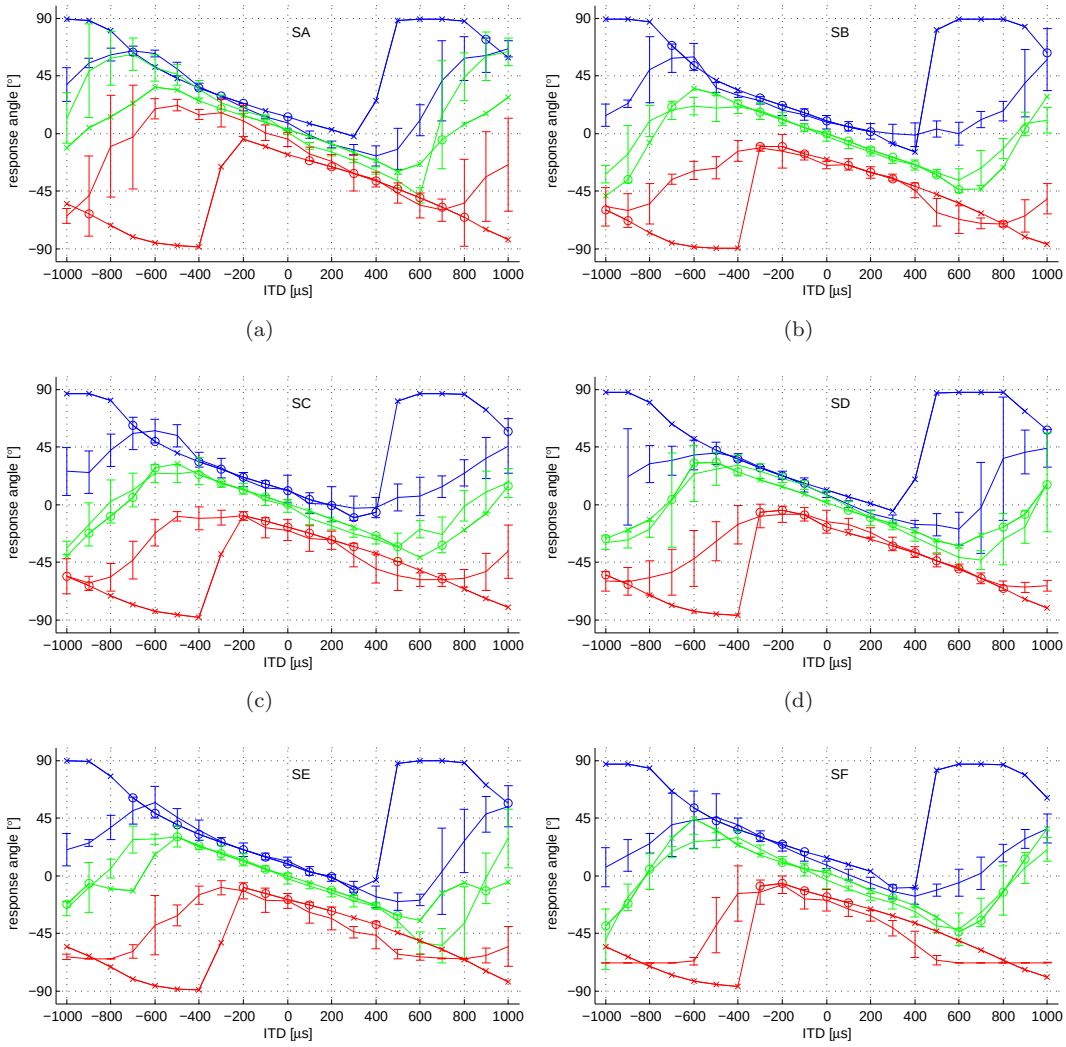


FIGURE 4.19: The result of the comparison between the listening test data and the model predictions at 600 Hz. Subject's initials are shown on the top centre of the plots, where blue, green and red colours indicate -6, 0 and 12-dB target ILDs. Subjective judgements are shown as error-bars indicating the mean responses and the 95% confidence intervals, while the averages of the model predictions are marked as ∘ (not rejected) and × (rejected) to indicate whether or not the null hypothesis in the t-test is rejected. If the average of the model predictions is within the confidence interval, then the null-hypothesis is NOT rejected, and the model is regarded as predicting well the subjective judgements.

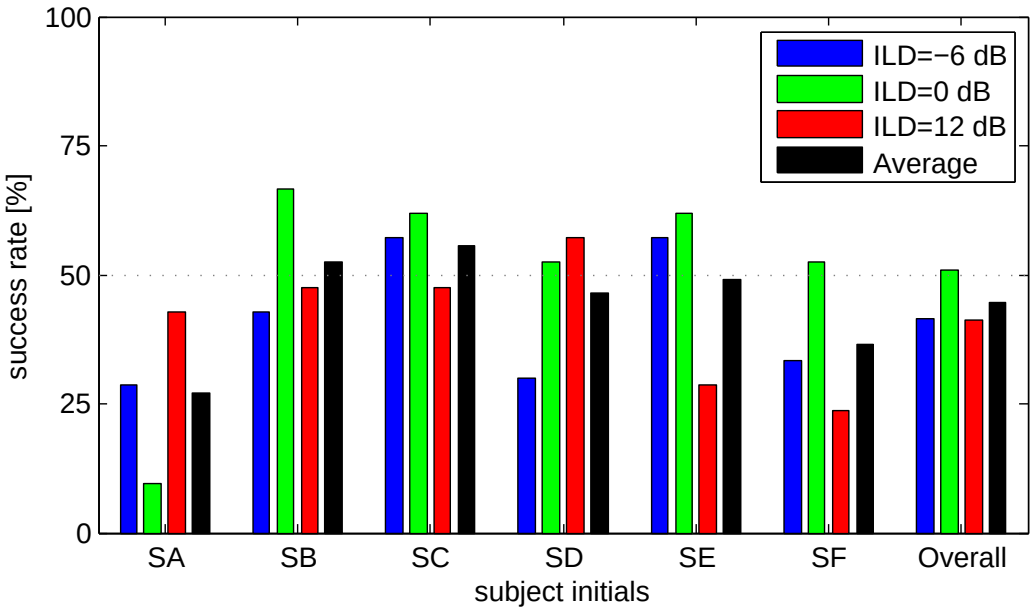


FIGURE 4.20: The success rates at 600 Hz based on the results of the t-test shown in Fig. 4.19.

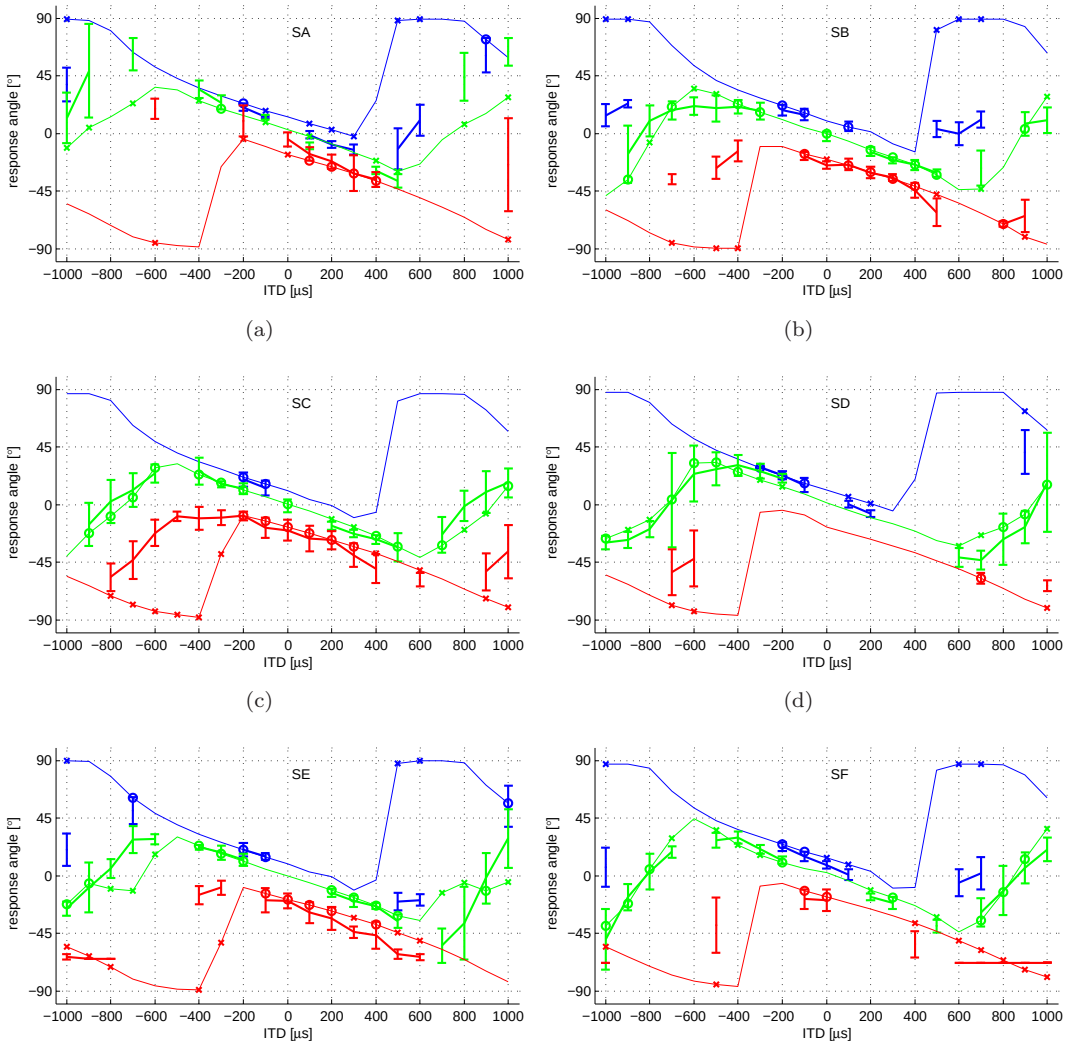


FIGURE 4.21: Fig. 4.19 has been redrawn showing only the target conditions where the normality of both listening test data and subjective judgements has NOT been rejected. Whereas the thin lines represent the averages of the model predictions, thick error-bars represent the subjective judgements for the 'qualified' target conditions. (Refer to the caption in Fig. 4.19 for other conventions in the graphs.)

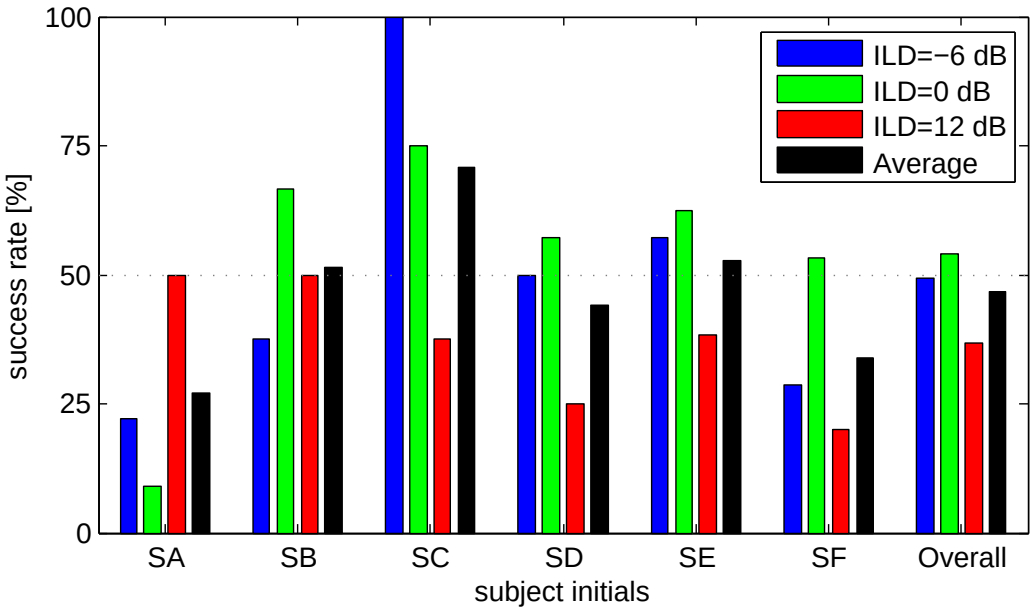
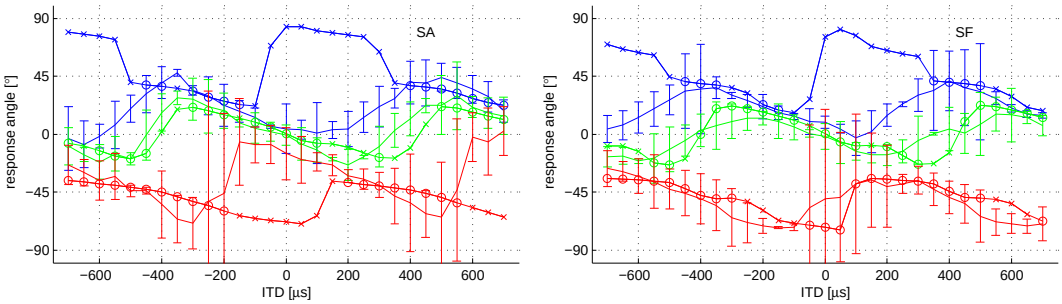
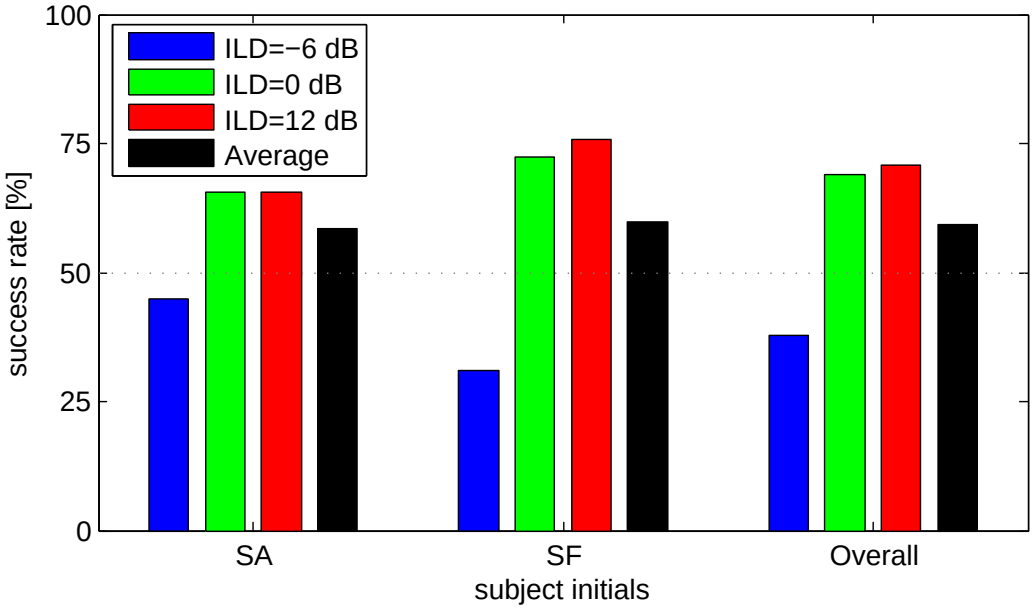


FIGURE 4.22: The recalculated success rates at 600 Hz based on the results of the t-test shown in Fig. 4.21.



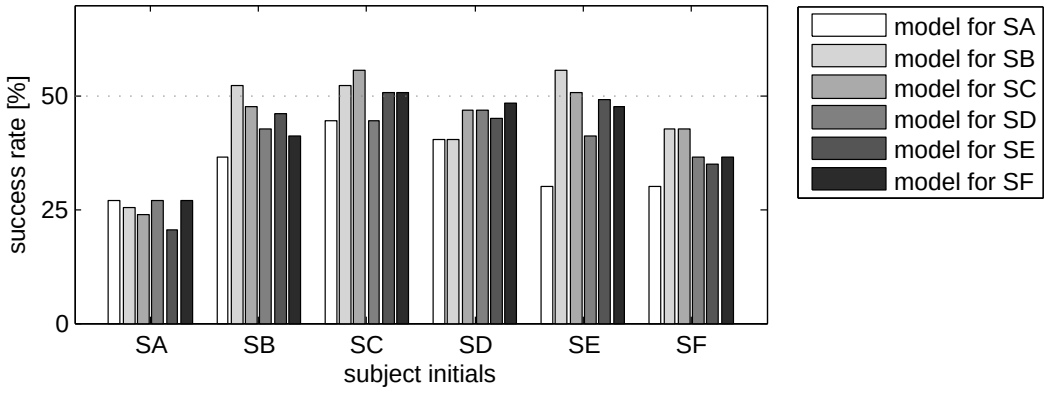
(a)

(b)

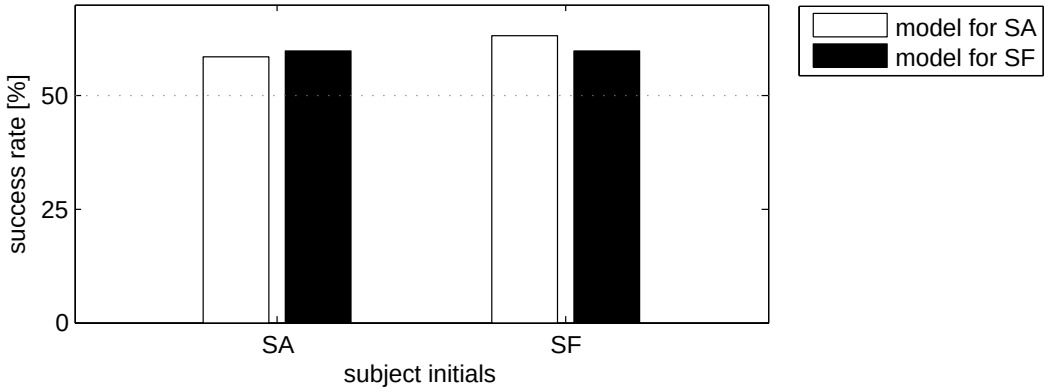


(c)

FIGURE 4.23: The result of the comparison between the listening test data and the model predictions at 1200 Hz. (a)&(b) Subject's initials are shown on the top right of the plots, where blue, green and red colours indicate -6, 0 and 12-dB target ILDs. Subjective judgements are shown as error-bars indicating the mean responses and the 95% confidence intervals, while the averages of the model predictions are marked as ∘ (not rejected) and × (rejected) to indicate whether or not the null hypothesis in the t-test is rejected. If the average of the model predictions is within the confidence interval, then the null-hypothesis is NOT rejected, and the model is regarded as predicting well the subjective judgements. (g) The success rates based on the results of the t-test are shown for each target ILD and subject.

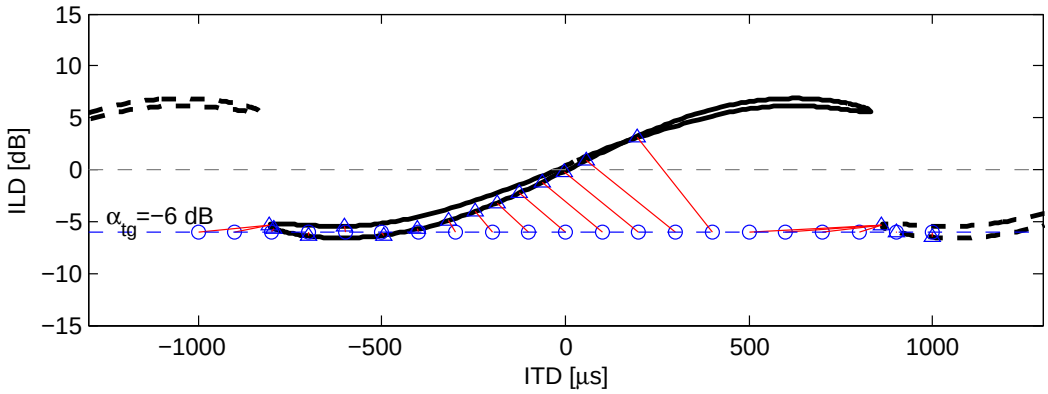


(a)

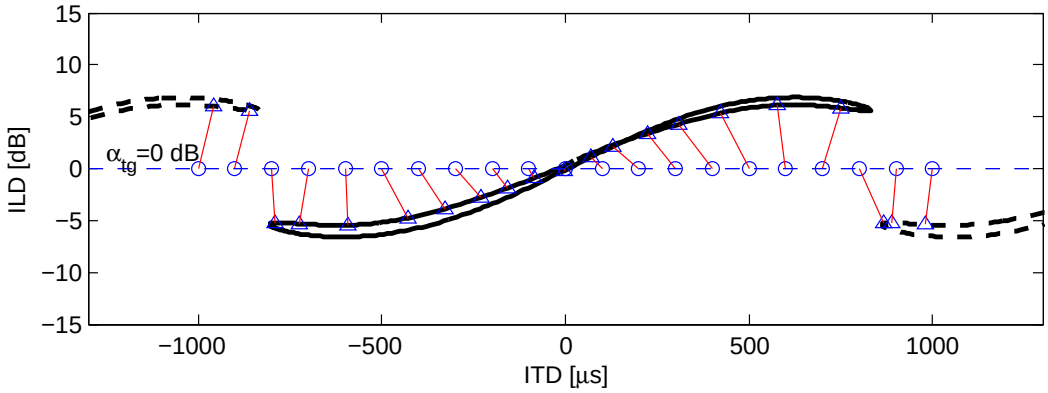


(b)

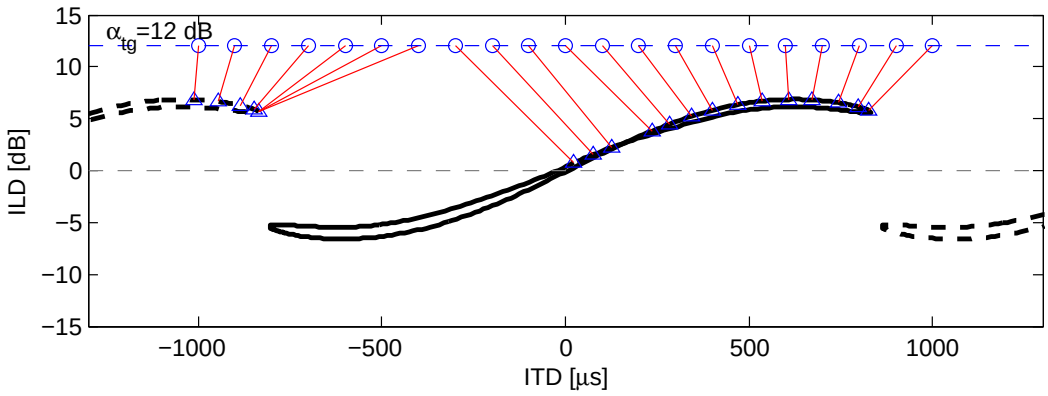
FIGURE 4.24: The result of the cross-comparison between the model predictions and the subjective judgements at (a) 600 Hz and (b) 1200 Hz. Each bar represents the success rate of the t-test, where each subjective listening test data has been compared with the predictions by all individual models.



(a)

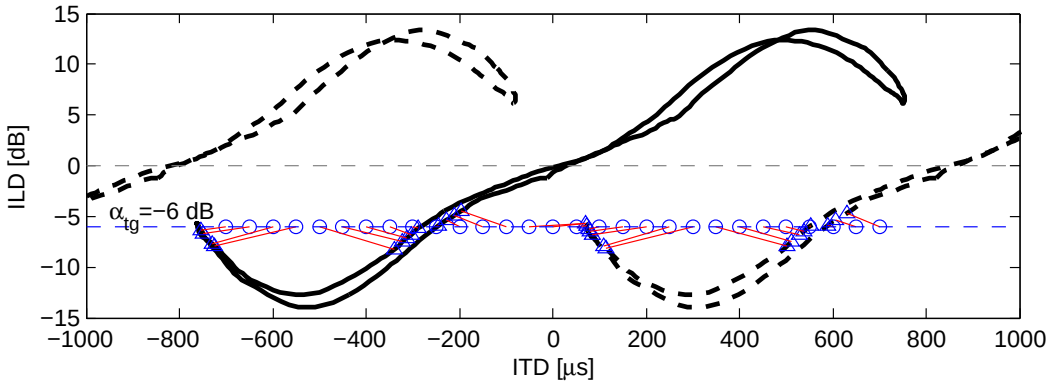


(b)

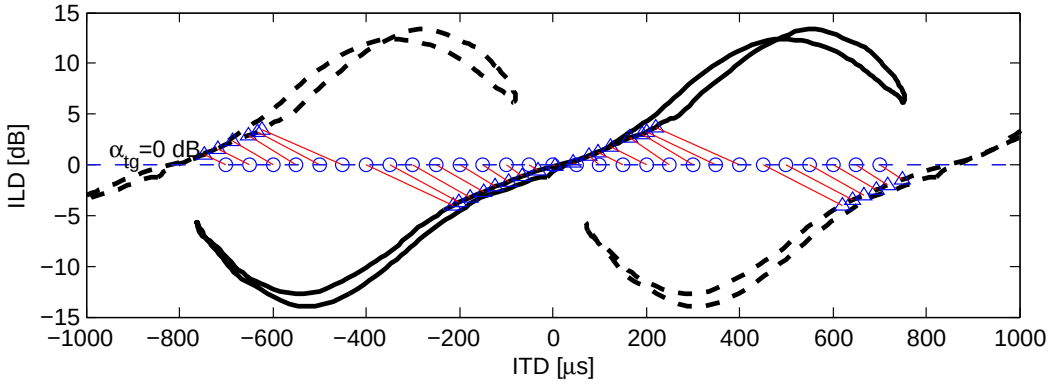


(c)

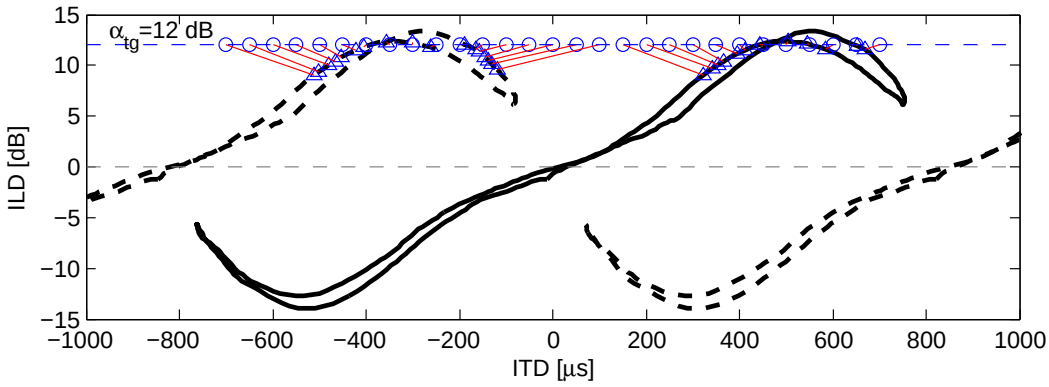
FIGURE 4.25: Diagrams showing the actual matching process at 600 Hz to find the nearest-neighbours for the target ILDs of (a) -6 dB, (b) 0 dB and (c) 12 dB. Circles indicate the target conditions while triangles represent associated model predictions on the characteristic curves. (Characteristic curves for the subject SA. the thick solid and dashed lines for the primary and the secondary curves, respectively.)



(a)



(b)



(c)

FIGURE 4.26: Diagrams showing the actual matching process at 1200 Hz to find the nearest-neighbours for the target ILDs of (a) -6 dB, (b) 0 dB and (c) 12 dB. Circles indicate the target conditions while triangles represent associated model predictions on the characteristic curves. (Characteristic curves for the subject SA. the thick solid and dashed lines for the primary and the secondary curves, respectively.)

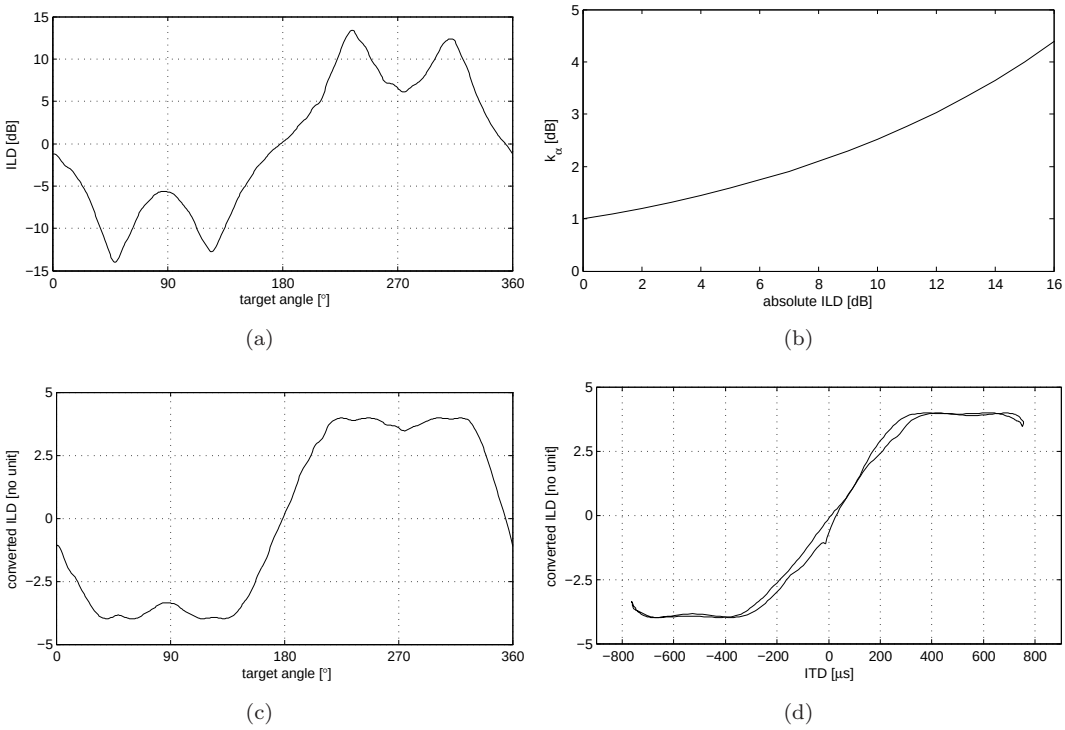


FIGURE 4.27: (a) The ILD function at 1200 Hz for the subject SA. The multiple local maxima and minima resulted in the hook-shaped ends of the characteristic curves shown in Fig. 4.26. (b) A temporary function for the new scaling factor k_α, which exponentially increases from 1 dB to 4 dB as the absolute ILD increases from 0 dB to 15 dB. (c) The ILD function at 1200 Hz after the conversion using the new scaling factor shown in panel (b). (d) The characteristic curve in the transformed ITD–ILD space, where only the ILD axis is scaled by the function shown in panel (b).

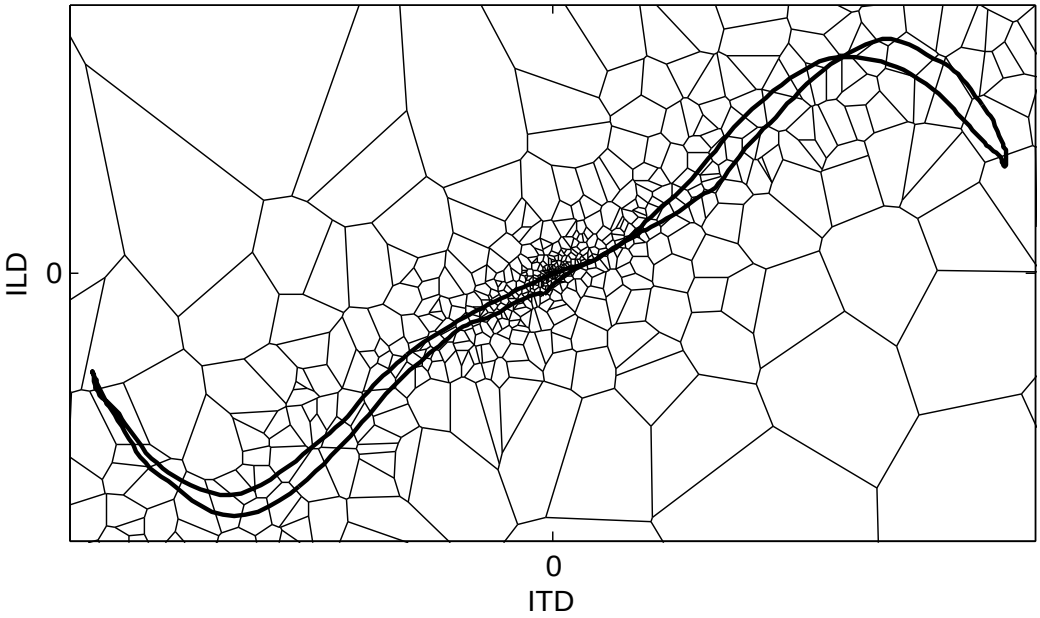


FIGURE 4.28: A schematic diagram showing the author's hypothesis regarding the neural selectivity in the ITD–ILD space. Finer neural resolutions is represented by smaller 'cell,' which is particularly observed for the ITD and the ILD closer to 0 in absolute values, and for those ITD–ILD pairs closer to the natural combinations, that is, the characteristic curve.

Chapter 5

A pattern-matching model of sound lateralisation and localisation

5.1 Introduction

Human auditory processing models have been developed for decades in the name of binaural signal processing [4]. They are computational but reflect psychoacoustic and physiological findings associated with the human auditory system, and therefore, once structured, it is expected that they operate as artificial listeners.

Apart from the models concerning the subjective perception of sound quality [62–64], there have been many models focused on how humans obtain the spatial information associated with sound sources [5, 6, 8, 12], and in chapter 2, one such model has been described. In particular, the characteristic-curve model was focused on the central decision-making process, where the estimate of source location could be obtained only at a single frequency. It has been suggested that, in order to handle wider band sound signals, there has to be a weighting scheme to integrate the estimates computed in each auditory frequency band, which necessarily has to be assumed in view of the lack of evidence from neuroscientific findings. In addition, establishing the characteristic curves in the high frequency range appears to be difficult, since the frequency boundary between the regions of the waveform and the envelope ITDs is hard to define.

The hearing model that will be established in this chapter is also designed to predict the location of a sound source in the horizontal plane, but possibly with a wider range of frequency content. The model employs a binaural processor recently suggested by

Breebaart et al. [1], which can be considered to be an extension of Jeffress' coincidence detector model [5]. Incorporating an additional transfer line with attenuation taps, this hypothetical binaural processor generates so-called 'EI-cell activity pattern' ('EI' abbreviates 'excitation-inhibition;' hereinafter referred to as an 'EI-pattern.'), the minimum of which indicates the probable ITD and ILD of the binaural input signals. Together with an adequate model of peripheral processing that reflects the mechanism of neural transduction in the inner hair cells, the suggested binaural processor can resolve the issue of the ambiguous frequency boundary between the waveform and the envelope ITDs, where the transition between the two regions is made in the EI-patterns without discontinuity.

Regarding the frequency weighting, many models of spatial hearing suggest that the influence of auditory frequency bands with a low signal level has to be discounted or omitted from the final model prediction, while those with greater signal energy should receive higher weighting [7, 9, 65]. In a similar manner to these previous models, so-called 'power-weighting' will be considered in the current model where the sum of the left- and the right-channel signal energy will be assumed to influence the local estimate in each frequency band on the final model prediction.

Each of the peripheral, binaural and the central processes of the current model will be detailed in section 5.2, where, in particular, the characteristics of the EI-patterns will be discussed. In section 5.3, the implications of the model for various conditions of human spatial hearing will be explored, specifically for the lateralisation of dichotic pure tones and the localisation of broadband signals both in real- and virtual-field conditions. Finally, the conclusion for this chapter is given in section 5.4.

5.2 Description of model

The current model based on the pattern-matching procedure is computational, consisting of three main modules (see Fig. 5.1): 1) the peripheral processor for the transfer characteristics of outer, middle and inner ears and the neural firing mechanism of the inner hair cells in the cochlea, 2) the binaural processor where the EC (equalisation and cancellation) process [1, 33, 66] is implemented to obtain EI-patterns across auditory frequency bands, and 3) the central processor or the decision-making device giving a final judgement of the source location based on the localisation cues obtained in the binaural processor.

5.2.1 Peripheral processor

Compared to the binaural processing and the decision-making stages in the central nervous system, studies regarding the mechanical or electrophysiological aspects of the ears are relatively well established, and so the computational models concerning the transfer characteristics from the outer ear to the cochlea are more or less consistent throughout the literature [1, 8, 12–14, 67]. In designing the current model, excessive detail less relevant to the goal of this work has been discarded whilst appropriate processes have been selected and modified from the established models. The signal flow across the peripheral processor is illustrated in Fig. 5.1.

Whilst the input signals to the peripheral processor are the signals recorded near the listener's ear drums, for the simulations presented later in this chapter, a monaural source signal has been convolved with HRTFs corresponding to a specific azimuth angle, giving synthesised binaural input signals to the model. It is noteworthy that all the signal processing tasks described in the following sections have been performed at a sampling frequency of 48 kHz in accordance with that of the HRTF database (see section 3.2.1).

Given the input signals, the transfer characteristics from the ear drums to the oval window of the cochlea have been accounted for by a bandpass filter with roll-off of 6 dB/oct below 1 kHz and -6 dB/oct above 4 kHz [1].

The frequency selectivity of the basilar membrane has been investigated and modelled in many ways, and in recent studies, this has been realised in the form of a filterbank. The gammatone filterbank [68] followed by the gammachirp filterbank [69] has been extensively used in similar modelling work, and there are a few associated software modules open to the public such as the **Auditory Image Model** (AIM) [70]. These software modules process an input signal through several filters that imitate the bandwidths and the shapes of the auditory filters on the basilar membrane, producing as many channels

of output as the filter number. These multi-channel outputs are assumed to be handled separately in the following processes [4, 26]. In an effort to reduce computational load and to take a more intuitive approach, a stand-alone module coded by Slaney [71] has been used, which implements a fourth-order gammatone filterbank [see Fig. 5.2(a) for the frequency response]. There is no agreement regarding the density of frequency channels in the auditory filter, and therefore models in previous work use a different number of filters in different frequency ranges, as partly summarised by Jin et al [14]. The current model has been designed to have 60 channels from 300 Hz to 12 kHz where one half of each equivalent rectangular bandwidth (ERB) [26] overlaps with the nearby filters [see Fig. 5.2(b)], which appears to be reasonable in comparison to similar models.

Inner hair cells in the organ of Corti convert mechanical movement of the basilar membrane to neural activity. Since the neural excitation occurs in relation to the basilar membrane movement relative to the tectorial membrane, the first process taking place in the inner hair cells can be modelled as a half-wave rectifier, as has been the case in most of the relevant models. In addition, the loss of the phase-locking in neural firing is taken into account by a low-pass filter so that only an envelope remains at high frequencies. Among filters of different characteristics used in previous studies, the current model employs a fifth-order butterworth low-pass filter cut-off at 770 Hz which is identical to that used by Breebaart et al [1].

Beside the signal-processing modules described above, some additional processes can be found in the literature such as the amplitude compression and the adaptation loops [1, 13] to reflect the nonlinearity of the basilar membrane input-output function and the forward/backward masking effect. The former was approximated by a square-root compression [71], however, the latter was omitted in the current model where sound localisation is sought only for a relatively stationary sound source.

Fig. 5.3 shows an example of the transformation of input signal in the peripheral processor where the source signal has been assumed to be a broadband Gaussian noise.

5.2.2 Binaural processor

The binaural processor delays and attenuates the 60-channel signals from the peripheral processor on the predefined 'mesh-grid' taps of τ (characteristic ITD) and α (characteristic ILD) (see Fig. 5.4). The neural inputs, fed into the top and the bottom transfer lines from the left and the right ear, respectively, undergo a time-delay by $\Delta\tau$ at each triangular tap, and are then carried on to the vertical transfer lines, this time to be attenuated by $\Delta\alpha$ at each rectangular tap. If digitally implemented, the discrete signal is fed into the transfer lines sample by sample, and the time-interval between each ITD

tap is determined to be $1/f_s$ where f_s is the sampling frequency. Finally, at the circled tap labelled as EI, the signals from the two channels having a characteristic ITD and ILD are subtracted giving an EI-cell activity value. Mathematically, this process can be described by the following equations. First, the EI-cell activity at a time instant is represented by [1]

$$EI(i,t,\tau,\alpha) = (10^{(\alpha/40)}L_i(t+\tau/2) - 10^{(-\alpha/40)}R_i(t-\tau/2))^2 \tag{5.1}$$

where $L_i(t)$ and $R_i(t)$ represent the input signals from the left and the right peripheral processors for the i-th channel. From Eq. (5.1) and the fact that there are nonlinear processes in the preprocessor, it is clear that the characteristic ILD, α is not equal to the interaural level difference between the binaural input signals in the beginning.

The instantaneous representation of the EI-cell activity is integrated with a double-sided exponential time window $w(t)$ which takes into account a finite binaural temporal resolution [1]:

$$EI'(i,t,\tau,\alpha) = \int_{-\infty}^{\infty} EI(i,t+t_{int},\tau,\alpha)w(t_{int})dt_{int}, \tag{5.2}$$

where

$$w(t) = \frac{exp(-|t|/c)}{2c} \quad , \qquad (c = 30\,ms) \tag{5.3}$$

This time-averaged EI-cell activity is normalised by the energy of the input signals, e_L and e_R to regularise the EI values regardless of the amplitude and the duration of the signal (see Appendix B):

$$EI''(i,t,\tau,\alpha) = \frac{EI'(i,t,\tau,\alpha)}{\sqrt{2e_Le_R}} + n(i,t,\tau,\alpha) \tag{5.4}$$

where the internal noise $n(i,t,\tau,\alpha)$ has been introduced to take into account the imperfect equalisation and cancellation process in human hearing. The noise mask $n(i,t,\tau,\alpha)$ in $\tau-\alpha$ space has been assumed to be a zero-mean Gaussian noise specified by the standard deviation σ_n, which is the model parameter to be controlled to adjust the statistics of the predictions in accordance with the result of subjective listening tests. While the influence of the model parameter will be discussed in section 5.3.2 in relation to sound localisation, Fig. 5.5 shows an example of EI-pattern before and after the addition of the noise mask given by $\sigma_n = 0.12$.

It is noteworthy that the postprocess of the EI-pattern denoted by Eq. (5.4) is different from that in Breebaart et al. [1] shown in Fig. 5.4, where the logarithmic compression of the EI-patterns have been regarded as less relevant in the current model.

At 48 kHz sampling frequency, 38 ITD taps and 20 ILD taps have been incorporated, where the resolution of the network was designed to be $\sim 42\,\mu s$ and 1 dB, respectively,

giving approximately $\pm 800\,\mu s$ of ITD coverage and $\pm 10\,\mathrm{dB}$ dynamic range in ILD. The final output of the binaural processes is a group of EI-patterns across frequency, some of which are, for example, shown in Fig. 5.6 (without the internal noise added). At relatively low frequencies [panel (a)], EI-patterns preserve the periodicity in the τ direction. However, as frequency increases [panels (b) and (c)], the space between nearby minima becomes narrower, while each minimum in the pattern becomes more ambiguous. (Minima of EI-patterns are indicated by $*$.) This is due to the loss of phase-locking implemented by the low-pass filter in the peripheral processor, and the periodicity of the EI-patterns is no more observable above about 1.5 kHz, where the phase information is completely lost. However, it is also apparent that the EI-pattern still retains the information of envelope ITD at higher frequencies as illustrated by the moderate shift of the pattern in τ direction from $0\,\mu s$ to $\sim -450\,\mu s$ [panel (c)].

As mentioned before, the minimum position of the pattern indicates the most probable ITD and ILD between the binaural input signals, while the whole pattern is regarded as being unique for source location and frequency. Fig. 5.7 illustrates an example of the cross-correlation between EI-patterns corresponding to the azimuth angles from 0° to 359° at every 1° where it is obvious that the similarity between a pair of EI-patterns decreases as source locations become further apart from each other. It is also remarkable that the source locations mirror-imaged with respect to the frontal plane have very similar EI-patterns, which can be associated with the front-back confusion or the cone of confusion [26]. However, the cross-correlation between EI-patterns across azimuth angle is also a function of frequency, and at certain frequencies it becomes irregular, implying that the EI-patterns can be less distinguishable.

In order to further investigate the uniqueness of the EI-patterns in terms of the frequency, the patterns for 0° source location can be compared across frequency, where the EI-patterns have their local minima aligned at $(\tau, \alpha) = (0, 0)$. Fig. 5.8 shows the cross-correlation between pairs of EI-patterns across frequency, and it is clear that at frequencies higher than about 1.5 kHz, EI-patterns lose most of their unique features, which is attributed to the loss of time information in the neural transduction [26]. On the other hand, the high correlation between the low-frequency EI-patterns possibly results from the compact population of gammatone filters within the narrow range of frequency [see Fig. 5.2(b)].

5.2.3 Central processor

A decision-making device in computational models of human perception is a processor mapping the intermediate output to a final judgement, preferably designed to reflect

the cerebral structure and mechanism. Since, unfortunately, relevant information has yet to be fully understood, most of the binaural hearing models employ a decision-making device based mainly on assumptions that are consensually accepted. In the cross-correlation model, the peak position or the centroid of the cross-correlation function has traditionally been chosen as an indicator giving spatial location information of sound sources [8]. In the meantime, the development of artificial neural networks provided a more sophisticated non-linear decision device to combine all available information regarding the spatial extent of sound sources such as ITD, ILD and spectral cues [12, 72–74].

In the current model, given the uniqueness of the EI-patterns in accordance with the known characteristics of auditory signal processing, a pattern-matching process has been assumed to take place in the central decision-making stage. First, a white Gaussian noise is filtered through one of the KEMAR HRTFs [27] that have been interpolated from 5-degree to 1-degree resolution (see Appendix A for the HRTF interpolation). If this synthesised binaural signal is considered as the input signal to the peripheral and the binaural processes of the current model, the ultimate collection of the 60×360 EI-patterns corresponding to 60 auditory frequency bands and 360 azimuthal directions can be obtained to form a memory, or a template in a computational terms, of sound localisation, as each of these patterns is close to unique for corresponding direction of source in each auditory frequency band as discussed above.

Having established the template, a simple pattern matching procedure is employed to find the best match for a new target signal. Based on the cross-correlation between the target EI-pattern and the template, the pattern-matching procedure is represented by

$$\chi(\theta, f) = \frac{\sum_{\tau,\alpha} EI''_{tg}(\tau, \alpha, f) \cdot EI''_{T}(\tau, \alpha, \theta, f)}{\sqrt{\sum_{\tau,\alpha} EI''^{2}_{tg} \sum_{\tau,\alpha} EI''^{2}_{T}}} \tag{5.5}$$

$$\theta_p(f) = \arg\max_{\theta} \chi(\theta, f) \tag{5.6}$$

where EI''_{tg} and EI''_{T} are the EI-patterns from the target and the template, respectively, and χ indicates the normalised cross-correlation between the patterns.

It is expected that this pattern-matching process works in a similar way to finding the nearest neighbour in the characteristic-curve model described in chapter 2, where, presumably, the conversion factors k_τ and k_α in Eq. (2.1) are equivalent to the neural resolution determined by the amount of delay and attenuation in each tap of the 2D network shown in Fig. 5.4. Nevertheless, in order to show the equivalence between the two decision-making processes, it is essential to prove that the outcome of Eqs. (5.5)

and (5.6) is equal to that of Eqs. (2.2) and (2.3) at a single frequency, which is difficult since the EI-patterns are computed for individual HRTFs, and are not analytically represented. Assuming that the left and the right channels of the binaural input signals are related only by time delay and amplitude difference, the analytical form of the EI-patterns have been approximated in appendix B, and this can be further investigated in future work to clarify the link between the two models.

Figure 5.9 shows an example of the function $\chi(\theta, f)$ for a source at 45°, where circles indicate $\theta_p(f)$ in each of 60 frequency bands. It is obvious that greater similarity is found between the target EI-patterns and the template when the response angle is in the vicinity of the actual target location. In addition, it is noteworthy that this pattern-matching procedure can give mirror-imaged errors associated with the front-back confusion, which are indicated by the local estimates found around 135°. This is true even without the introduction of internal error, if a running, instead of frozen, noise source is used as an input signal.

It is essential to further combine the model predictions in each frequency band in order to produce a final global prediction. Working with the cross-correlation model, Stern et al. [75] and Shackleton et al. [37] previously dealt with this issue by making use of a frequency weighting of binaural stimuli. For instance, the latter has shown that the simple weighted addition of the cross-correlation functions across the auditory channel can represent a global cross-correlation function.

Similarly, the current model applies a weighting scheme to collect all the 'local' predictions to establish a 'global' probability function $D(\theta)$, where the weighting function has been obtained from the energy spectral density multiplied by the salience factor of binaural stimuli suggested by Raatgever [37]. The latter reflects the empirical dominance of binaural stimuli at low frequencies, while the former assumes that a signal band of greater energy has more influence on the final decision. Fig. 5.10 shows examples of the weighting functions depending on the spectral characteristics of source signals.

Mathematically, this 'power-weighting' scheme can be represented as

$$D(\theta) = \frac{\sum_f \delta_{\theta\theta_p(f)} \times W(f)}{\sum_f W(f)} \tag{5.7}$$

where $W(f)$ is the frequency weighting function, and δ is the Kronecker delta. (It should be recalled that the function $D(\theta)$ is defined only for integer numbers between 0° and 359°, limited by the resolution of the interpolated HRTF.)

An example of the probability function $D(\theta)$ is shown in Fig. 5.9(b), which resulted from Fig. 5.9(a). There are two prominent peaks in the histogram. The peak at about 45° is

the estimate of the true source position, whilst the other at 135° indicates the possibility of front-back confusion as already implied in Figs. 5.7 and 5.9(a). Since the pattern-matching procedure first produces estimates for each frequency band, it is possible to have many distinctive peaks in the plot of $D(\theta)$, arising from a multiple number of sound sources that are separated in the frequency domain or at least have non-overlapping spectral components. Similar to the case discussed in section 2.3 regarding the dual images created by ambiguous interaural phase differences, the listener's attention is assumed to play an important role when multiple peaks are observed in the probability function $D(\theta)$. Here, it is assumed that the model selects the estimate corresponding to the highest peak.

5.3 Implication of the model

Having established the procedure for obtaining a single estimate for target binaural input signals, the current model can be now investigated in terms of its predictions for the lateralisation and the localisation of acoustic stimuli. As mentioned before, the matching process of the current model is, arguably, similar to the nearest-neighbour finding process of the characteristic-curve model, and the implications of both models to various listening conditions are expected also to be similar. On the other hand, the improvement made for the current model is that broadband stimuli can be dealt with by incorporating the tentative frequency weighting scheme. Therefore, in addition to the lateralisation of dichotic pure tones, the model predictions can be compared with the results of the subjective listening tests reported in the literature regarding the localisation of broadband sound sources.

For the simulation results presented in the following sections, the current pattern-matching model has been implemented in Matlab 7.0 using the signals shown in Fig. 5.11 as an initial monaural input. A total of 500 repetitions have been made for each target condition while the internal noise $n(i, t, \tau, \alpha)$ being a random variable as suggested in section 5.2.2. Front-back confusion has been resolved so that the predictions made for the lateralisation may be found in the frontal hemisphere between -90° and +90°, while those made for the localisation be located only in the hemisphere corresponding to the target location. The lateral target locations at 90° and 270° have been exempted in this post-processing.

5.3.1 Lateralisation of dichotic pure tone

As was the case for the model based on the characteristic curves described in chapter 2, it is of primary interest to investigate the implication of the current pattern-matching model to the lateralisation of dichotic pure tones. (For details of the lateralisation, readers are referred to chapter 4.) Among the previous experimental studies introduced in chapters 2 and 4, the listening test results published by Sayers [31], and Toole and Sayers [76] contain some fundamental properties of the auditory process involved in lateralization, and these studies are regarded as a good starting point to investigate the capability of the current model.

Sayers [31] and Toole and Sayers [76] presented interaural disparities in pure tones and an impulse train, respectively, to subjects who were asked to indicate the lateral displacement of the test sound on a visual scale chart. From these experiments, Sayers [31] found that for low-frequency pure tones below 1500 Hz, the lateral displacement

presented by the interaural differences is periodic with the period of the test tones. In addition, it has been reported that there is a transition zone around the interaural phase difference of π where the image-position judgement moves to the contralateral side. It is also known that in this transition zone, listeners often report multiple images at each far side lateral position as well as the averaged position at the centre. Figs. 5.12 and 5.13 show the simulation results for the lateralization tasks under similar conditions as in Sayers [31], where the current model describes well the periodicity of the lateral displacement by ITD (Fig. 5.12) and the existence of the multiple images around the ambiguous phase differences (Fig. 5.13). These predictions by the current model resulted from the characteristic of the EI-pattern that its minimum position shifts according to the ITD (and the ILD) that becomes ambiguous near the critical values for the transition.

Although it has not been explicitly shown with experimental data, Sayers [31] also reported that the maximum laterality of perceived image position decreased as frequency increases while the slope of the position judgements against the given time delay appeared to be independent of the frequency. This feature is also successfully predicted by the current model as shown in Fig. 5.12, which is associated with the periodicity of the EI-pattern, thus the shorter intervals between the critical ITDs at higher frequencies.

Finally, Sayers [31] reported an interesting feature of human sound lateralisation where both ITD and ILD have been controlled. In his measurement data shown in Fig. 5.14(a), it is found that both interaural disparities can affect the lateral position of a sound image, and they can be cancelled or strengthened by each other to some degree. In addition, the asymmetry with respect to ITD is found to increase with ILD, and the transition to the contralateral side is shown to take place at smaller values of ITD for a larger ILD (both in an absolute sense). All these features are reasonably predicted by the current model as depicted in Fig. 5.14(b), where the increased asymmetry and the transition to the contralateral side are clearly shown. However, it is also clear that the transition to the contralateral side for non-zero ILDs takes place slightly earlier in the model giving a misleading indication of the subjective test results. Finally, it is noteworthy that the comparison between the listening test results and the model predictions shown in Fig. 5.14 is only qualitative since the two results have been given in different units.

5.3.2 Localisation of real broadband source

Before presenting the results of the model simulations, it is worth first discussing the results of subjective listening tests reported in the literature. There have been a large number of studies to measure human performance in locating a sound source, and some of them carried out more than several hundreds of trials, analysing them to attain the

associated statistics. Since the conditions of those experiments are diverse (as well as the analysis methods), a direct comparison is not easy to make between the localisation performances described by different studies. However, it is still meaningful to examine those subjective experiments that used similar stimuli in relatively consistent circumstances. Among such comparative studies, Blauert [4] summarised a couple of previous experiments in the 1960's and 70's, and suggested that listeners can locate a sound source in the front and back more accurately than at lateral positions, which are rather classical data, but still agree well with those of recent experiments.

In Table 5.1, a few subjective experiments considered to be relevant to the current study have been listed in terms of their methodologies and source signal specifications. A main improvement in recent subjective experiments is the way listeners indicate the source location: in the studies summarised by Blauert [4], listeners were asked to move a loudspeaker to the positions which they believe are 0°, 90°, 180° and 270°. However, in the majority of experiments performed in recent years, listeners wear an electromagnetic device that automatically reads the angular position to which their heads are directed. Undoubtedly, the latter method is fast and accurate in some ways, but there are concerns about systematic errors associated with, for example, spontaneous eye movement [39] and less mobility in indicating the rear source.

	Stimuli	**Duration**	**Direction**	**Response protocol**
Blauert [4]	White noise pulse	100 ms	Horizontal	Alignment of a loudspeaker
Makous and Middlebrooks [38]	Noise of random-phase flat-amplitude spectrum, 40∼50 dBSL	150 ms	Horizontal Vertical	Head movement monitored by electromagnetic device
Carlile et al. [39]	Broadband white noise, 70 dB	150 ms	Horizontal Vertical	
Recanzone et al. [77]	Gaussian noise, 30±2 dBSL	200 ms	Horizontal	
Current model	Gaussian noise, 60∼80 dB at ear entrance	100 ms	Horizontal	n/a

TABLE 5.1: Conditions of previous subjective experiments of sound localisation.

In Fig. 5.15, the listening test data from Blauert [4], Carlile et al. [39] and Makous and Middlebrooks [38] have been reproduced. For the purpose of comparison, some modifications have been made so that a positive localisation error may represent the centroid of response angles greater than the target angle throughout the range of 0° ∼ 360°. In addition, horizontal data were unavailable in Makous and Middlebrooks [38], and so the experimental results for a source at +5° elevation have been taken.

For the frontal positions, the mean responses in the listening test data seem to agree with each other, to some extent, up to about 130° [see Fig. 5.15(a)]. Then, some discrepancies start to emerge and grow, and at 180°, the data in Blauert [4] seem to diverge from the other available data reported by Carlile et al. [39] and Makous and Middlebrooks [38]. The latter two data sets show a similar tendency such that the localisation error turns

rapidly from positive to negative for the target locations around 90°, where a similar observation can be made for the responses around 270° reported by Carlile et al. [39].

It is remarkable that the localisation performance at 180° in Blauert [4] is much better than that in Carlile et al. [39] and in Makous and Middlebrooks [38] (if some extrapolation of data is allowed), which is demonstrated not only by Fig. 5.15(a) but also by Fig. 5.15(b) in terms of the standard deviation. Considering that their experiments have been carried out with fairly similar source signals, such a significant discrepancy for a sound source at the rear position can be probably ascribed to the method used to report the source location. In author's opinion, it is uncertain which listening test data reflect the true statistics of human performance in sound localisation, since there are insufficient target locations in the report by Blauert [4], while the results in Carlile et al. [39] and Makous and Middlebrooks [38] could be vulnerable to the measurement error associated with, for example, listener's use of eye movement.

It is also noteworthy that, despite some partial agreements noticed and discussed above, no pair of the three subjective experiments showed a satisfactory match with each other. This seems to imply that it is difficult to obtain any collective and conclusive statistics regarding human sound localisation ability by means of subjective experiments.

Having reviewed the results of the subjective listening tests reported in the literature, the current pattern-matching model can be adjusted to reflect the performance of human listeners in the task of sound localisation, in terms of the errors and the variances. For a source signal identical to that employed to establish the template EI''_T, the pattern-matching process without the internal error $n(i, t, \tau, \alpha)$ gives a perfect localisation of target sound sources, which is, however, undesirable. If a running noise is considered to be the source signal instead of the frozen noise used for the template, some errors can be incurred by the current model in the localisation task, but the range of the judgement errors and variances were found to be much less than those shown by the statistics for human subjects. Accordingly, the influence of the noise mask has been investigated by controlling the parameter σ_n in Eq. (5.4) in an attempt to adjust the accuracy of the current model.

Fig. 5.16 shows statistics of the current model with different values of σ_n for the target locations only in the right hemisphere. Both mean error and standard deviation of the model predictions increase in absolute value as more noise is added to the EI-patterns in each auditory frequency band independently. From Fig. 5.16(a), it is shown that the mean error increases until a certain target position depending on the value of σ_n, then starts to decrease as the source position approaches 90°. Interestingly, the sign of the mean error switches from negative to positive around 90°, which implies that the model predictions around this region are found closer to the median plane than is the

actual target position. In addition, there are two prominent maxima in the standard deviation, where the first is located between 30° and 45°, again, depending on the value of σ_n, while the second is always found at 90°. The greater variance for the target locations near 90° can be understood in relation to the higher correlation observed for pairs of EI-patterns in that region as shown in Fig. 5.7. It is also noticeable that the standard deviation at 90° is particularly high compared to the adjacent target locations, which is possibly attributed to the side-effect of resolving the front-back confusion for all other target angles except for 90°. Nevertheless, it is not clear how the first peak of the standard deviation in the frontal hemisphere can be linked to the aspects of the pattern-matching procedure used in the current model.

Superimposed on the subjective experimental results reported in the literature, the line-connected dots in Fig. 5.17(a) represent the mean responses of 500 model predictions, where the internal noise parameter $\sigma_n = 0.12$ has been found to give statistics most similar to those of subjective test results. Although the agreement between the simulation results and the published listening test data is not perfect, it is interesting to see that the current model gives predictions which are at least qualitatively consistent with the nature of the localisation tasks performed by human subjects. For example, the mostly positive errors for the target locations in the right frontal hemisphere have been reasonably simulated where the sudden sign change near 90°, in case of the test data reported by Carlile et al. [39] and Makous and Middlebrooks [38], has been also predicted well.

In terms of the standard deviation of the sound source judgements, the agreement between the model predictions and the results of the listening tests is noticeable for the target location at 90° and 270° and those in the frontal area. For other target locations, however, the simulation results appear to be misleading for the estimation of the variance in actual listening tests, where the variances for the rear target positions are particularly low. Finally, the raw predictions of the current model for $\sigma_n = 0.12$ are shown in Fig. 5.18 where the front-back confusion and the greater variabilities around lateral target positions are clearly observed.

To summarise, the localisation of broadband noise sources has been compared between the model simulation and the published listening test data. Some degree of agreement has been found between the two results especially in terms of the mean errors of the judgements, but there were also inconsistencies that are particularly prominent in the standard deviation. A further adjustment of the current model can be attempted by, for example, considering a non-uniform noise mask $n(i, t, \tau, \alpha)$ that reflects the probable signal-to-noise ratio in neural process depending on the ITD and/or the ILD. However, it is not clear that any improvement made by such a manipulation would be confirmed

by subjective test results, which tend to involve many psychological factors other than actual hearing process, often giving inconsistent results from experiment to experiment, as partly shown by the previous studies discussed above.

5.3.3 Localisation of virtual broadband source

Although there have been a few predecessors, stereophony is regarded as the first virtual acoustic imaging system capable of producing a reliable sound image. Invented by Blumlein [78] in early 1930's, it is a system that converts the phase difference of the signals recorded by a pair of microphones to the amplitude difference of in-phase input signals to two loudspeakers. It has been shown that this sound field can deliver an appropriate phase difference between listener's two ears (interaural phase difference) at low frequencies when free-field sound propagation is assumed [79].

The mathematical expression for the conversion from interchannel loudness difference to the interaural phase difference, thus the link between the amplitude ratio and the position of a virtual acoustic image can be given by 'the *sine* law' which is stated as (see appendix C)

$$\frac{\sin\theta_a}{\sin\psi} = \frac{\mathbf{L}-\mathbf{R}}{\mathbf{L}+\mathbf{R}} \qquad (f \lesssim 1000\,\mathrm{Hz}) \tag{5.8}$$

Here, ψ and θ_a represent the half aperture angle between loudspeakers and the azimuthal location of the phantom image, respectively, where $\mathbf{L}$ and $\mathbf{R}$ indicate amplitude gains given to the left and the right channels (see Fig. 5.19). In the conventional configuration of stereophony, ψ is usually set to be 30°, positioning the two loudspeakers and the listener on an equilateral triangle.

The sound field created by the simple stereophony described above has been considered in this section for an initial application of the current pattern-matching model to virtual acoustic imaging systems. Similar to the simulations presented in the previous section, a frozen white Gaussian noise of 150-ms duration has been used as a source signal [see Fig. 5.11(b)]. According to Eq. (5.8), this monaural signal has been then given a relative gain to create the loudspeaker input signals, which corresponds to the target image positions, θ_a from 0° to 30° at every 5°. The sound propagation from the transducers to the listener's two ears has been accounted for by the KEMAR HRTFs [27], which finally provided the binaural input signals to the model.

Fig. 5.20 shows the results of the simulation together with some subjective listening test data reported in the literature, where 500 predictions have been made by the current model with $\sigma_n = 0.12$ and averaged after resolving the front-back confusion. From the figure, it is shown that the current model gives predictions such that the stereophonic

sound images created by the *sine* law are perceived at azimuth angles greater than the design values, which is more prominent for the phantom images at intermediate target angles. In addition, the standard deviation of the model predictions (not shown in the figure) has been observed to gradually increase with the target location.

The listening test data cited in Rumsey [80] have been reproduced and superimposed to the model predictions in Fig. 5.20, where the agreement between the simulation results and the subjective test results is considered to be reasonable at least qualitatively in terms of the overestimation of the target position. The model predicts slightly less mean responses than the listening test data, but it is noteworthy that those subjective experiments employed speech signals or lowpass-filtered noise as source signals, which might give different results from the current simulation using a broadband noise signal.

In addition to the predictions produced by the current model, two other estimates of perceived image positions have been obtained and compared to the listening test results reported in the literature. The interaural time difference and the interaural level difference of the binaural input signals have been obtained at 600 Hz, which were then compared to the ITD and the ILD functions at the same frequency given by the KEMAR HRTF (see, for example, Fig. 3.18). Similar to the approach taken by Pulkki et al. [10], this mapping scheme gives estimates of the virtual image positions separately for the ITD and the ILD, which have been shown as the dashed and the dash-dotted lines, respectively in Fig. 5.20.

Comparing the three estimates plotted in thick lines in Fig. 5.20, it is shown that the predictions made by the current model are positioned, for most of the target positions, between the estimates given by the ITD and the ILD mapping schemes. If the extent of the overestimation is regarded as the criteria for successful predictions, the ILD mapping scheme has to be considered to give best estimates for the perceived locations of the virtual images. However, it should be recalled that the two estimates by the mapping schemes have been obtained only at a single frequency, where the level difference at low frequencies are not normally considered to be a reliable localisation cue by itself. On the other hand, it is noticeable that the prediction by the current model is reasonably consistent with the subjective test data, which has been given in a collective analysis across auditory frequency band.

5.3.4 Localisation in the reverberant environment

The well-known precedence effect states that human listeners consider the reflected sound waves that arrive within a certain time window as reinforcing the direct sound wave, hence enabling the localisation of sound sources even in a reverberant field. The performance of the PM model in the reverberant environment may be investigated in a simple configuration where binaural input signal is composed only of the direct sound from θ_D and the first reflection from θ_R, the delayed and attenuated version of the direct sound. As it is assumed that, in general, the duration of the signal is longer than the time delay, Δt between the direct and reflected sound, both sound signals are partly superimposed to upon one another. Fig. 5.21(a) schematically shows the model predictions which could have been made with reference to the 'instantaneous' EI patterns [see Eq. (5.1)], where '$\theta_D + \theta_R$' indicates the loci of the perceived image for the superimposed signals. As the auditory image has been created by two distinctive acoustic images at θ_D and θ_R, its position may be equivalent to that of a stereophonic image based on the delay- AND amplitude-panning method. Accordingly, it is obvious that the PM model working with the instantaneous EI patterns is not capable of predicting the precedence effect.

In the current model, however, the instantaneous EI patterns are further integrated according to Eq. (5.2), and with the time window, w [Eq. (5.3)] aligned with the onset of the direct sound, EI patterns corresponding to the reverberant part of the binaural signal will be made less influential to the final EI pattern, hence possibly simulating the precedence effect [see Fig. 5.21(b)]. Nevertheless, the slope of the window w is relatively slow [recall that the time constant in Eq. (5.3) is 30 ms], which might not be sufficient to discount the EI patterns representing the reflected sound waves. It should also be noted that Dau et al. [13] and Breebaart et al. [1] employed an adaptation loop in the peripheral processor which is designed to simulate the forward masking effect by greatly reducing the signal level shortly after the signal onset. Including this adaptation loop in the current model may result in even smaller EI-cell activities for the reflected sound signals, improving the model to better cope with reverberant sound fields. However, this aspect of the model has not been pursued in the current work in order to maintain the simplicity of the model, which has here been applied only to steady state sound stimuli.

5.4 Conclusion

In this chapter, a hearing model based on a pattern-matching technique has been suggested for sound lateralisation and localisation. Equipped with relevant peripheral processes, the current model employs 2D transfer lines by Breebaart et al. [1] as a binaural processor, where the left- and the right-channel signals are subtracted from each other according to the equalisation and cancellation procedure. The output of this binaural processor is the EI-cell activity patterns across frequency that contain the ITD and the ILD information of the input signals, while it has been found that these patterns are close to unique in each auditory frequency band for each different source location. In the following central processes, the target EI-patterns are compared to the template, a collection of the EI-patterns for all azimuth angles and frequency bands under consideration, and the local predictions made in each auditory frequency band are weighted according to a tentative frequency weighting scheme, finally giving an estimate of the source location on the horizontal plane.

Results of relevant listening tests have been simulated by the current model for the lateralisation of dichotic pure tones at low frequencies, the localisation of a single broadband sound source and the localisation of virtual acoustic images created by the *sine* law [79].

The lateralities of the inside-head images have been reasonably predicted by the current pattern-matching model, whereas the critical ITDs that are smaller, in an absolute sense, than those reported in experiments were found to be one of the issues, similar to the case of the characteristic-curve model as described in section 4.5.

The model parameter has been adjusted to reflect the published statistics associated with human localisation of a broadband noise source, from which $\sigma_n = 0.12$ has been found to be optimal. At this level of the internal noise, the qualitative agreement between the simulation results and the subjective test data reported in the literature has been found to be reasonable, particularly in terms of the mean error estimation. However, the variances of the subjective judgements were mainly underestimated by the current model except for a few target positions.

The application of the model has been relatively successful for the evaluation of the virtual images created by a stereophony system based on the *sine* law. The overestimation of the target positions has been predicted well by the model simulation, where estimates given by the ITD and the ILD mapping schemes have been also investigated to confirm the reliability of the predictions made by the current model.

Similar to the characteristic-curve model described in chapter 2, the current pattern-matching model is based on the two main assumptions in association with human auditory and cognitive processes: 1) a sound source on horizontal plane is localised by means of two interaural cues, the ITD and the ILD, and 2) the central decision-making process in the brain works with previous memories of sound localisation and corresponding feedback, performing its task by matching a new stimulus thereto. In fact, these assumptions quite probably oversimplify the overall complexity of human cognition. However, considering that they are also very commonly accepted hypotheses in the related field of neuroscientific research, the current model is regarded as worth investigating for its predictive scope in various conditions of human spatial hearing.

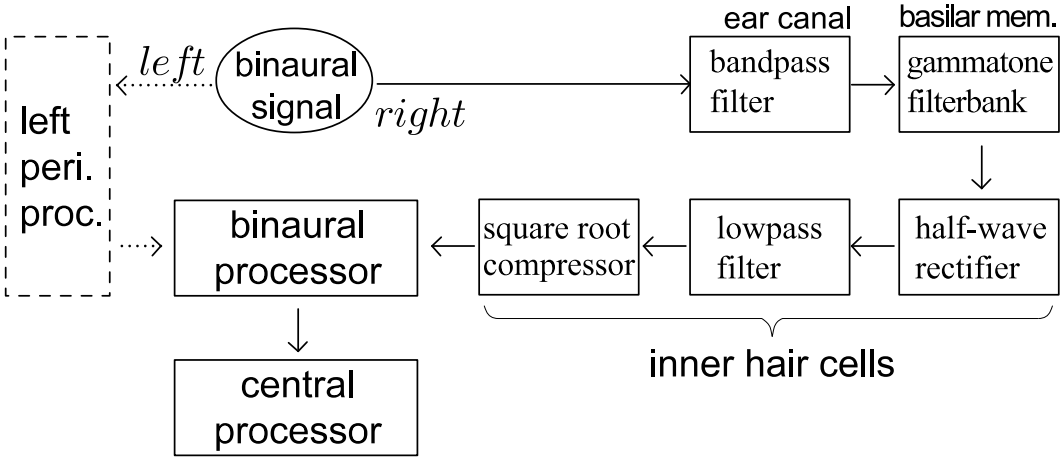
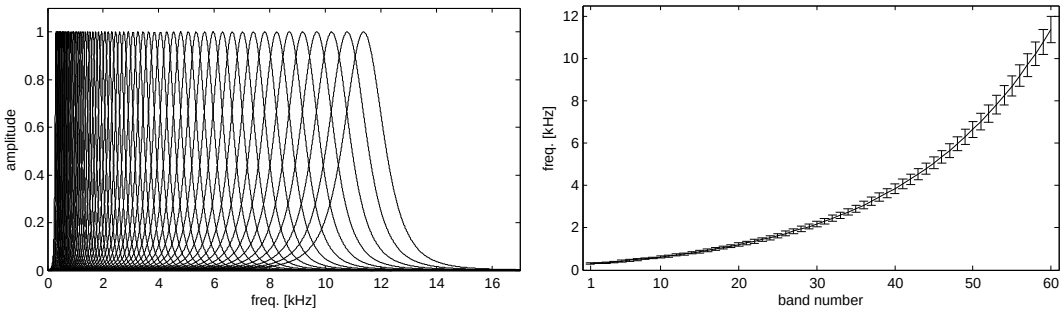


FIGURE 5.1: Signal flow in the current model is shown from the peripheral processor to the binaural and the central processor, where signal processing modules simulating the peripheral processor are detailed.



(a) (b)

FIGURE 5.2: 60 gammatone filters from 300 Hz to 12 kHz have been employed in the current model to account for the signal transformation in the basilar membrane. (a) Amplitude response and (b) equivalent rectangular bandwidth (ERB) [26] across band centre frequency, where about a half of each band is overlapped.

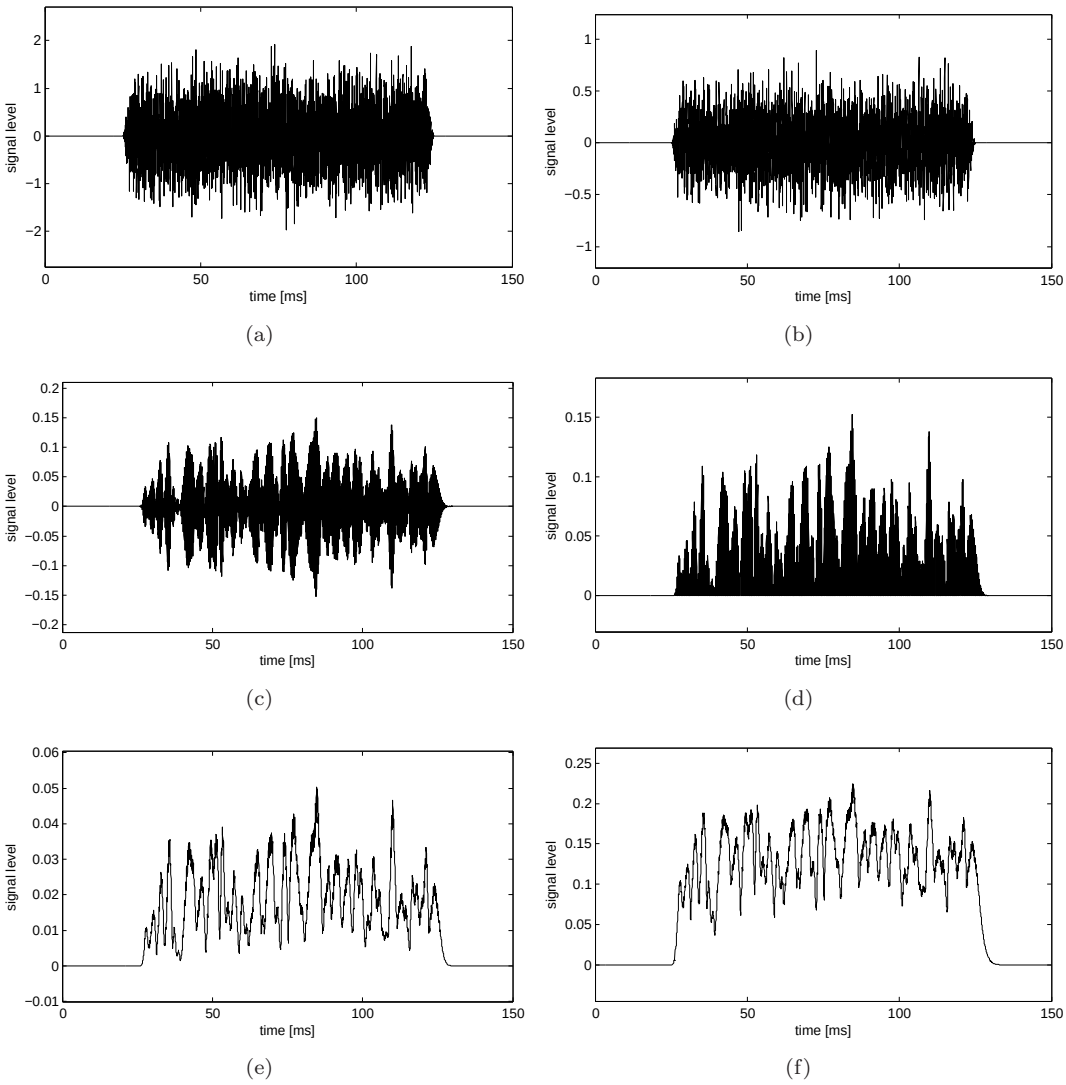


FIGURE 5.3: Signal transformation at each step of peripheral processes is shown. (a) Input signal (white Gaussian noise). (b) Bandpass filtered. (c) Filtered by gammatone filterbank (centre frequency of filter at 3075 Hz). (d) Half-wave rectification. (e) Lowpass filtered. (f) Square-root compression.

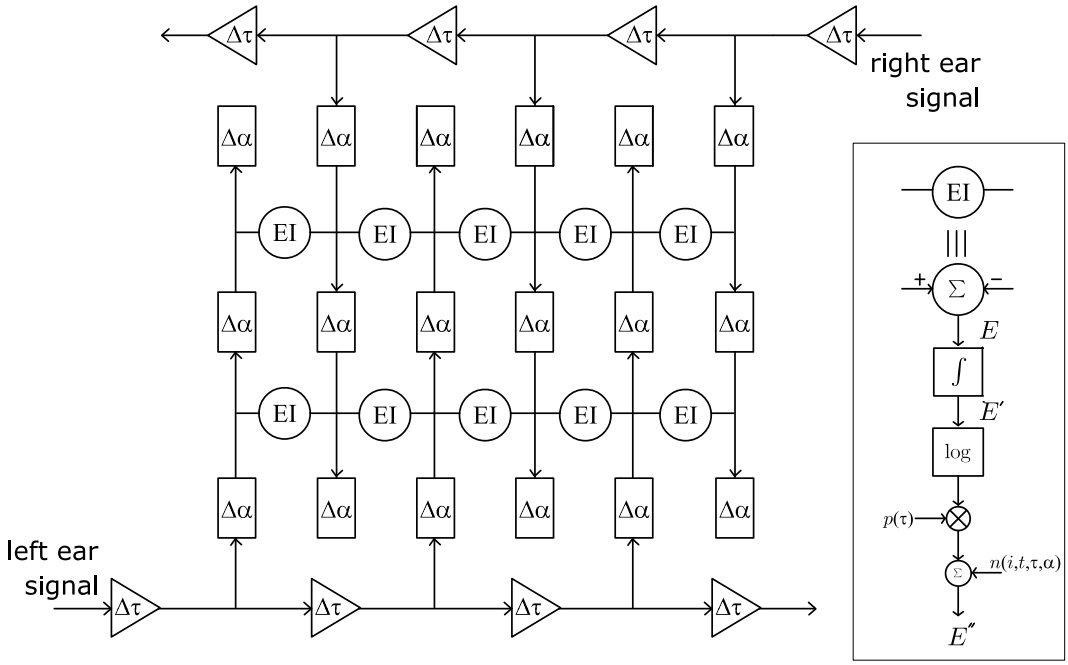
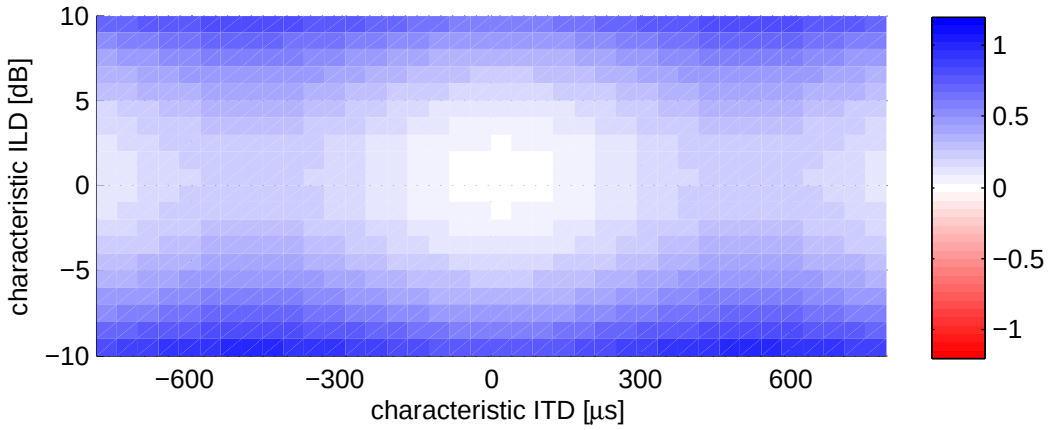
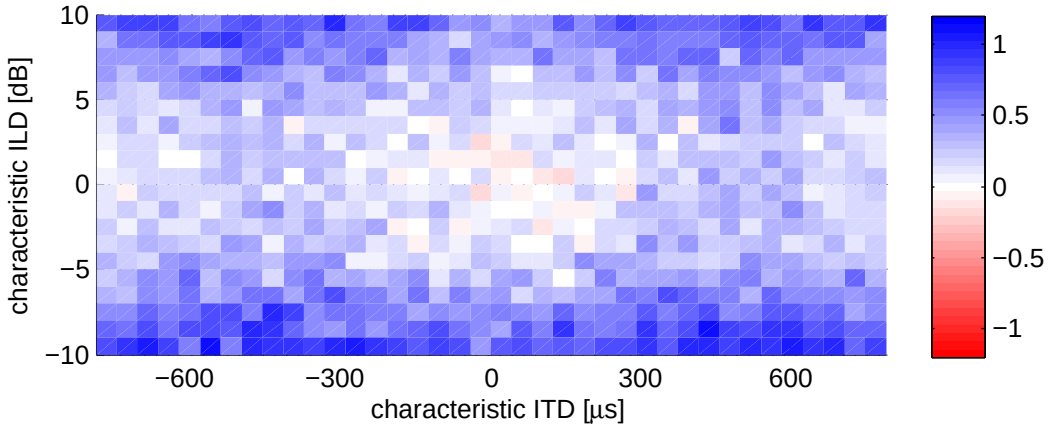


FIGURE 5.4: Two-dimensional network used by Breebaart et al. [1]. Operation in each EI cell is detailed in the box, where $p(\tau)$ represents the population of EI cells as a function of ITD, which, however, was not included in the current model. Also, the logarithmic compression was not considered. Instead, EI-patterns have been normalised by the energy of the input signals (see appendix B).

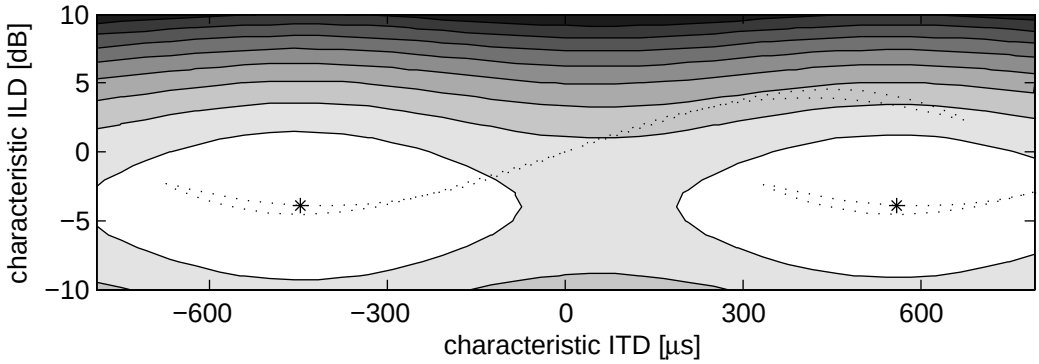


(a)

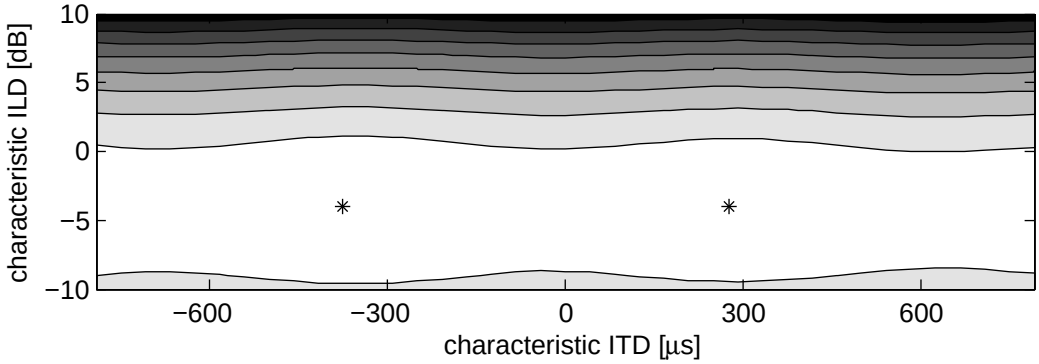


(b)

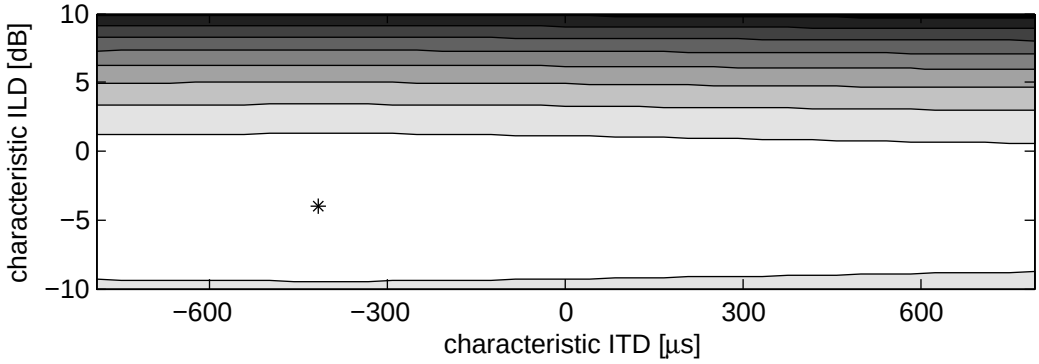
FIGURE 5.5: Influence of the noise mask $n(i, t, \tau, \alpha)$ for a EI-pattern computed at 993 Hz. (a) Before and (b) after the noise mask is added to the EI-pattern.



(a)



(b)



(c)

FIGURE 5.6: Examples of EI-patterns are shown for 45° at (a) 993 Hz, (b) 1534 Hz, and (c) 3075 Hz. Darker area indicates greater activity where points marked by asterisk indicate local minima of the pattern. The α axis (characteristic ILD) represents half the given external ILD due to the square-root compression in the peripheral processor (see Fig. 5.3). In panel (a), characteristic curves are superimposed to show that the minima of the patterns are actually on the curves.

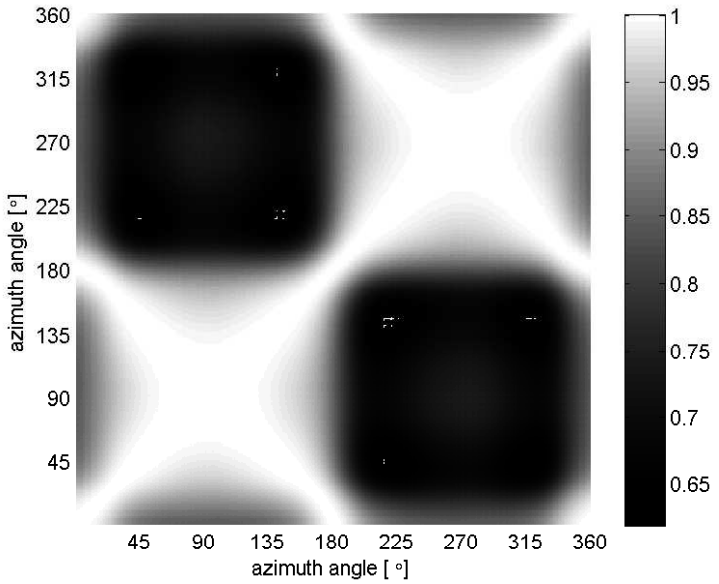


FIGURE 5.7: Normalised cross-correlation between pairs of EI-patterns across azimuth angle given at 458 Hz. (Scale of the colour contrast has been adjusted to clearly show the peaks of cross-correlation.)

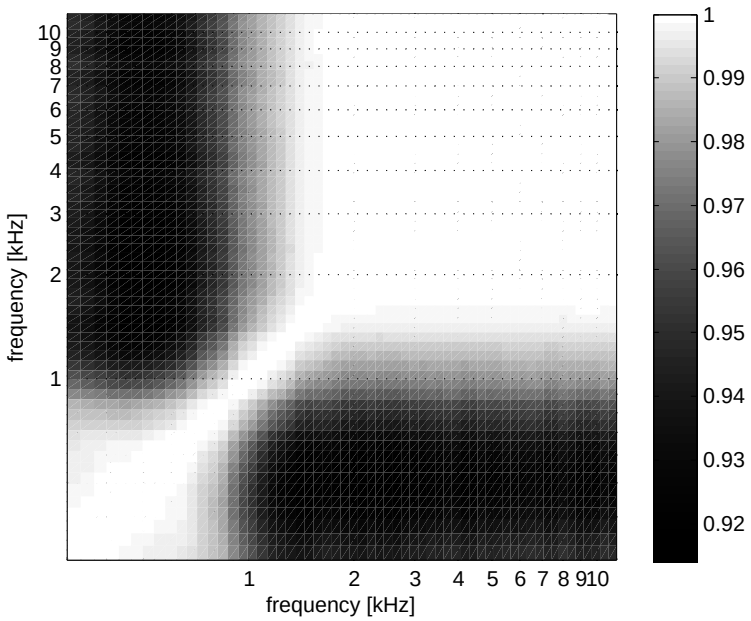
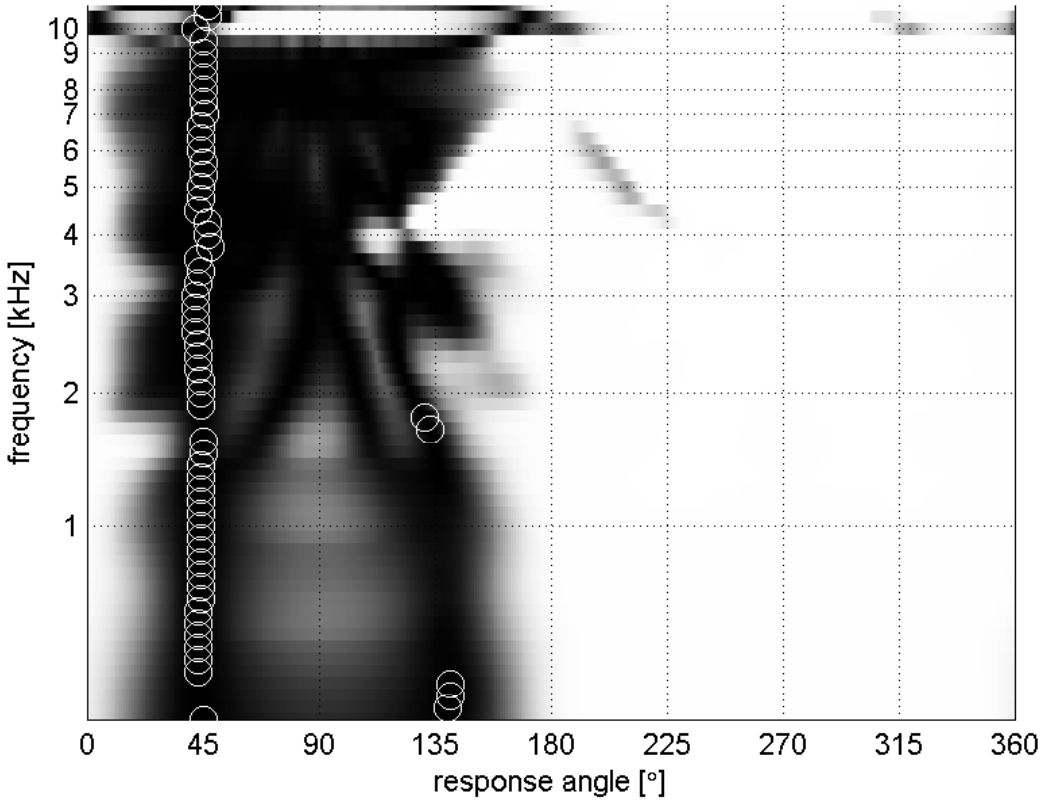
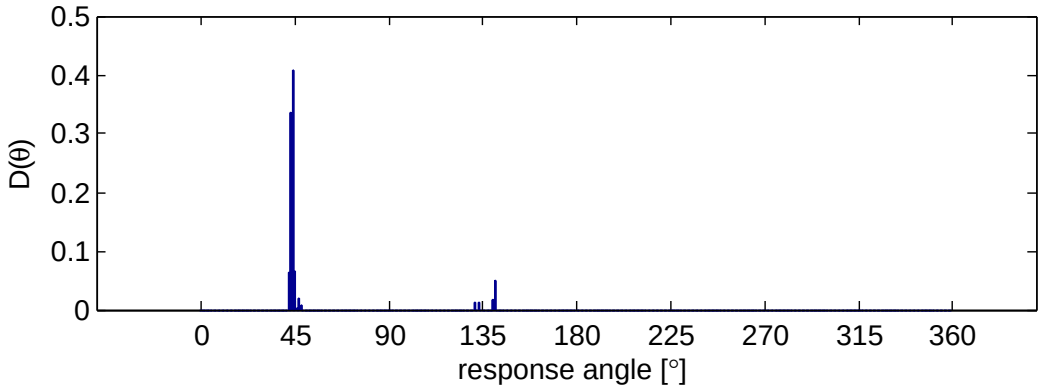


FIGURE 5.8: Normalised cross-correlation between pairs of EI-patterns for 0° as a function of frequency. (Scale of the colour contrast has been adjusted to clearly show the peaks of cross-correlation.)



(a)



(b)

FIGURE 5.9: (a) Cross-correlation between target EI-patterns and the predefined template is shown for each of 360 response angles and 60 frequency bands. (Scale of the colour contrast has been adjusted to clearly show the peaks of cross-correlation.) Estimates found in each frequency band are marked by circles, where a darker area indicates a higher correlation. Since a source is assumed at 45°, most responses are found near the target while the front-back confusion is also observed at 135°. (b) Probability function $D(\theta)$ has been plotted for target position at 45° . A secondary peak implying the front-back confusion is observed around 135°.

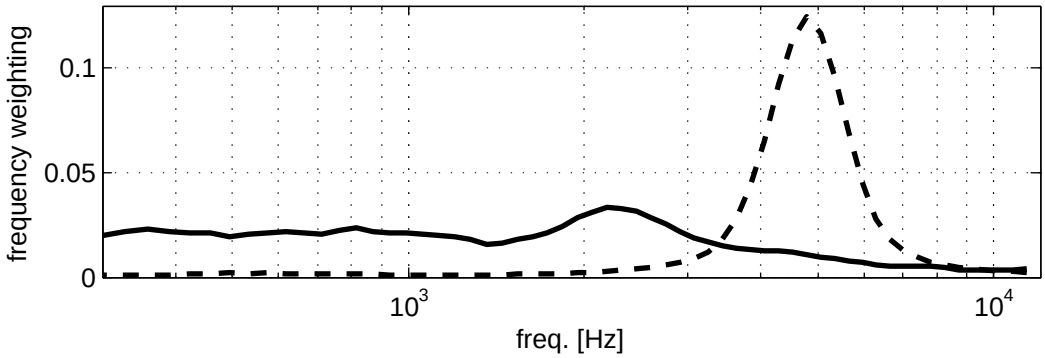
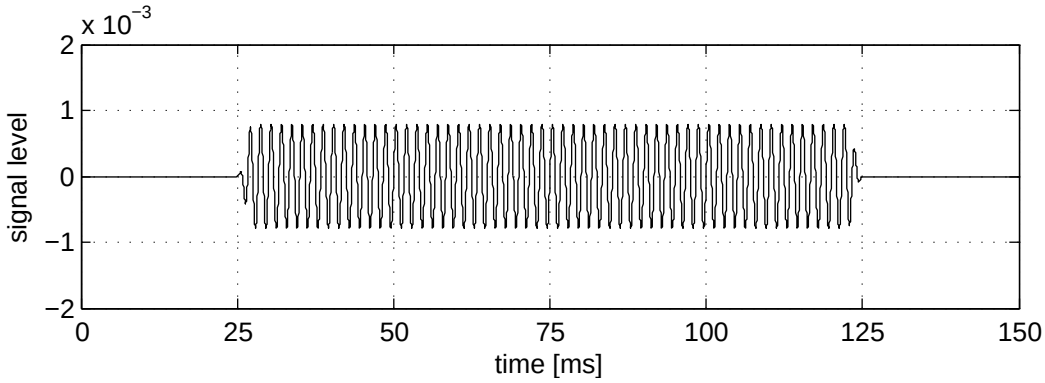
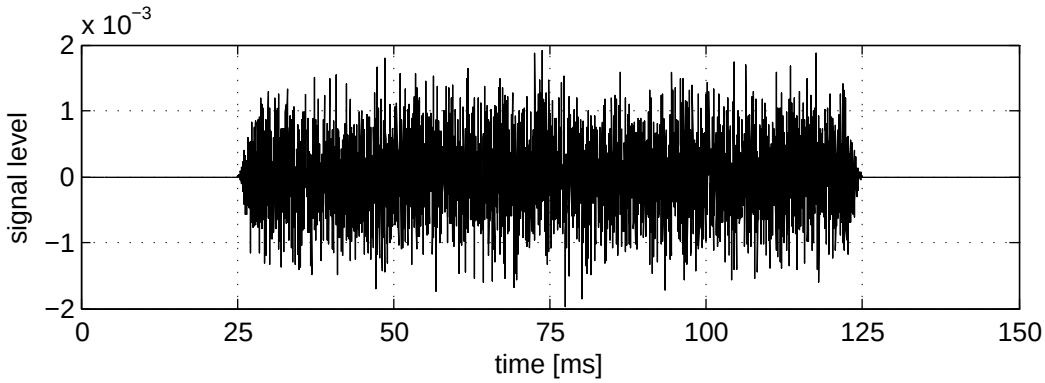


FIGURE 5.10: Examples of weighting functions $W(f)$ (normalised for comparison) for a white Gaussian noise (solid line) and a Gaussian tone burst at 5 kHz with 1-kHz bandwidth (dashed line).



(a)



(b)

FIGURE 5.11: Source signals used for the simulation which are 150 ms long with 100 ms of (a) a sinusoidal signal, for example, at 600 Hz (for the lateralisation simulation) or (b) a white Gaussian noise (for the localisation simulation).

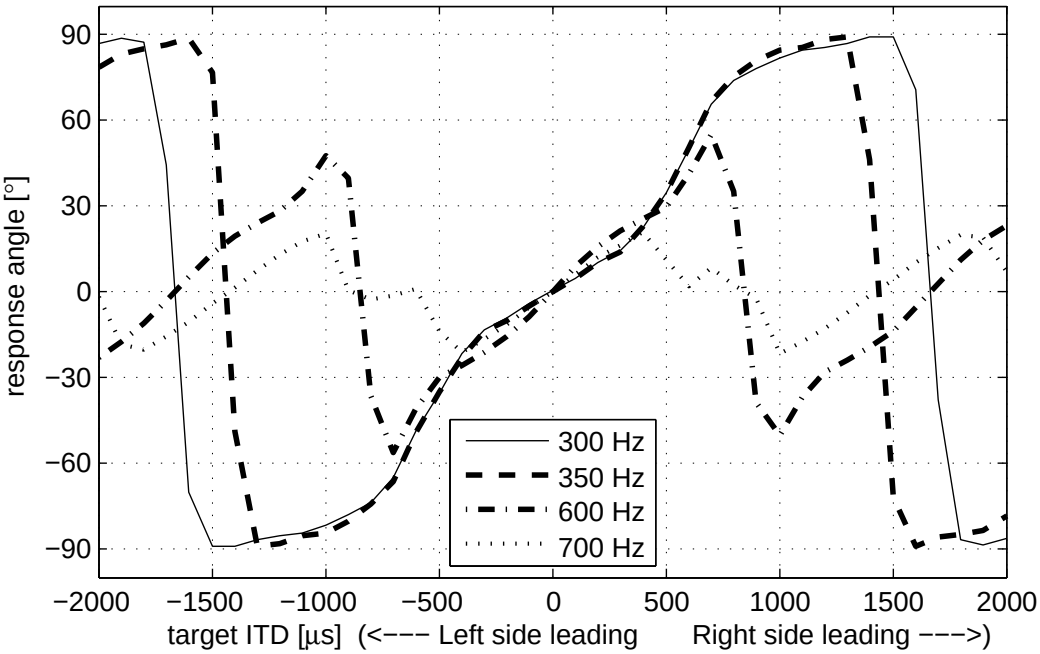


FIGURE 5.12: Model predictions for the lateralisation of pure tone signals at various low frequencies.

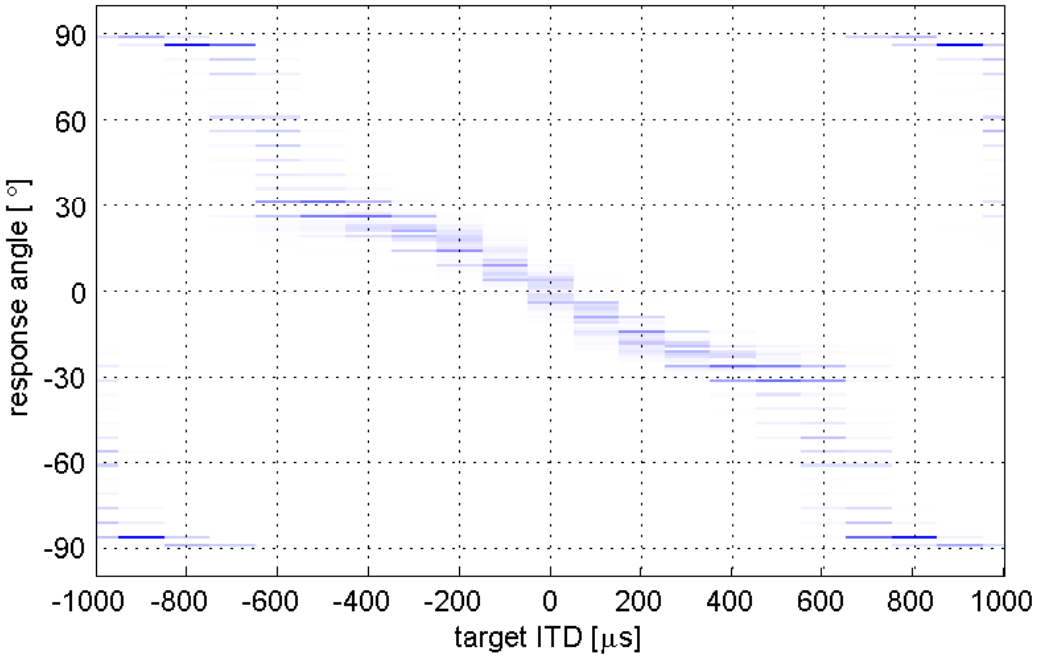
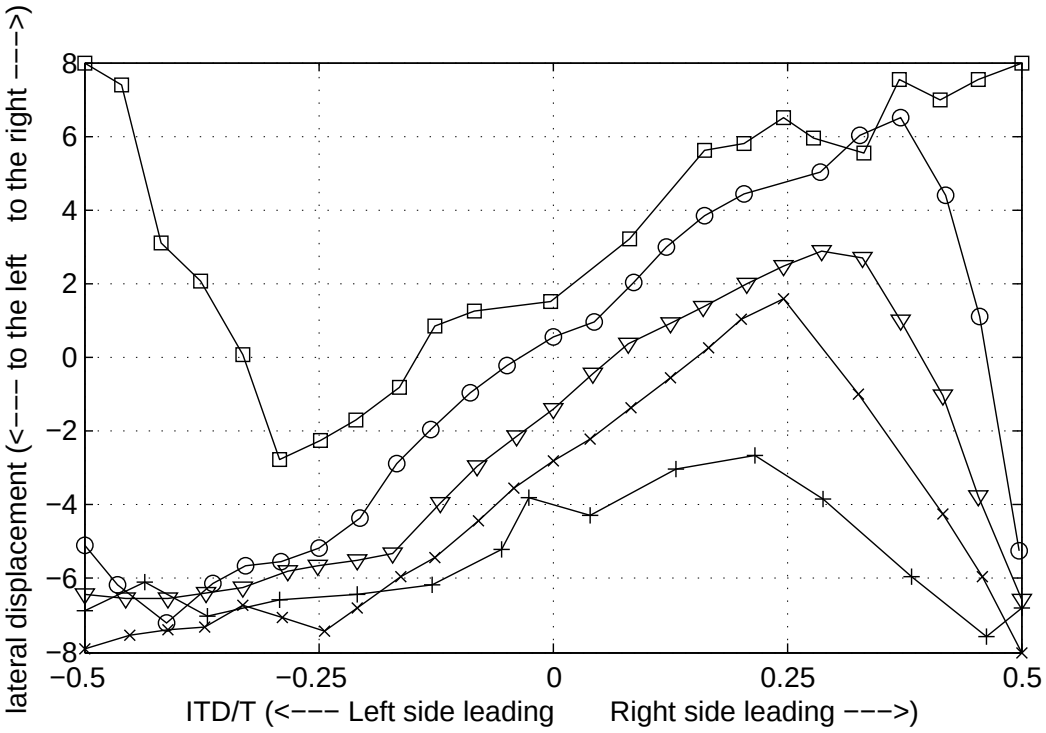
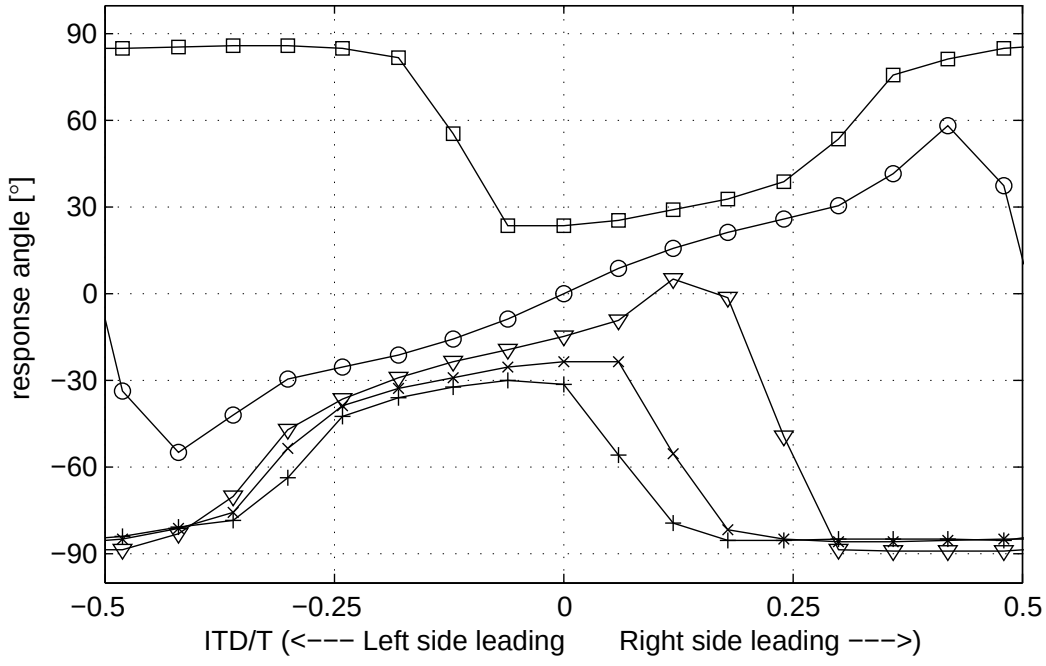


FIGURE 5.13: Model predictions for the lateralisation of pure tone signal at 600 Hz before averaging. Grey-scale indicates the relative frequency of the model responses along the vertical axis, which correspond to the target ITD every 100 μs shown on the horizontal axis.

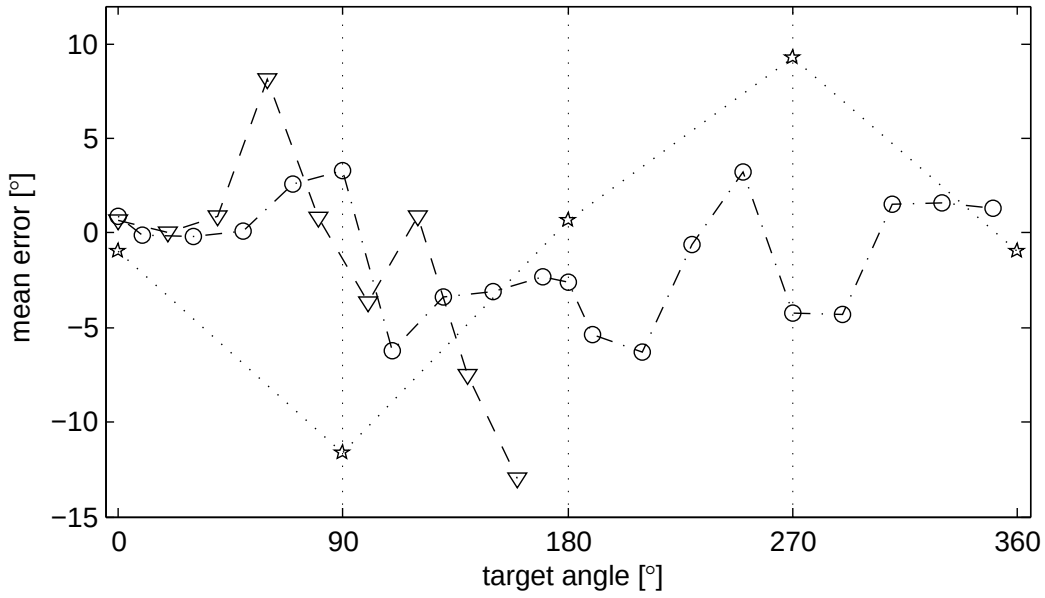


(a)

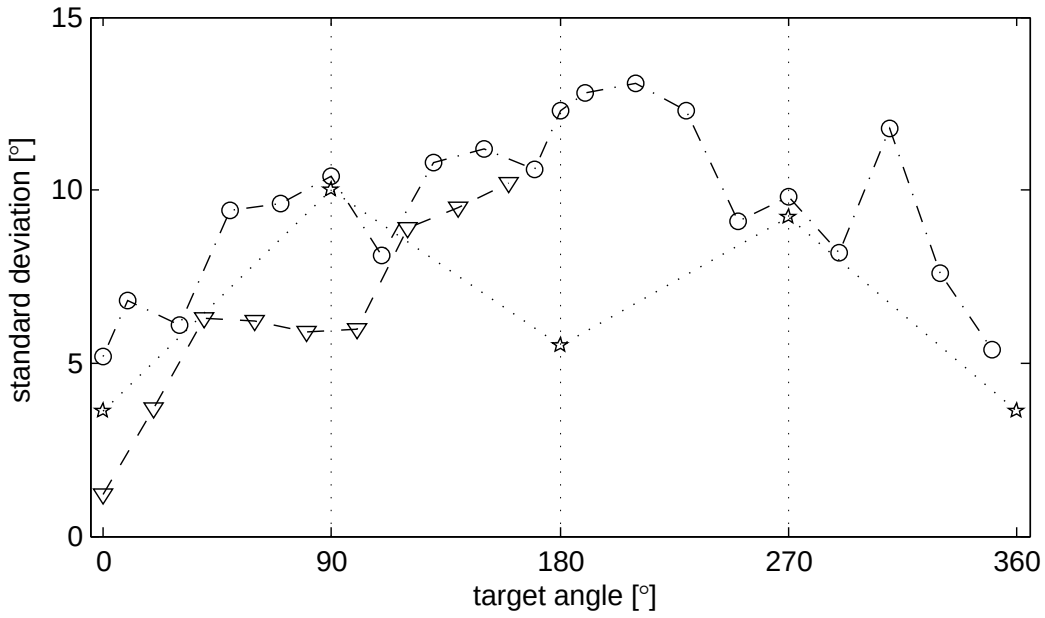


(b)

FIGURE 5.14: While the target ITDs are shown on the horizontal axis, judgements of image laterality for a 600-Hz pure tone are shown for various target ILDs: left channel louder by -9 dB (□), 0 dB (○), 6 dB (▽), 9 dB (×) and 12 dB (+). (a) Listening test results reproduced from Sayers [31] and (b) predictions of the current model.

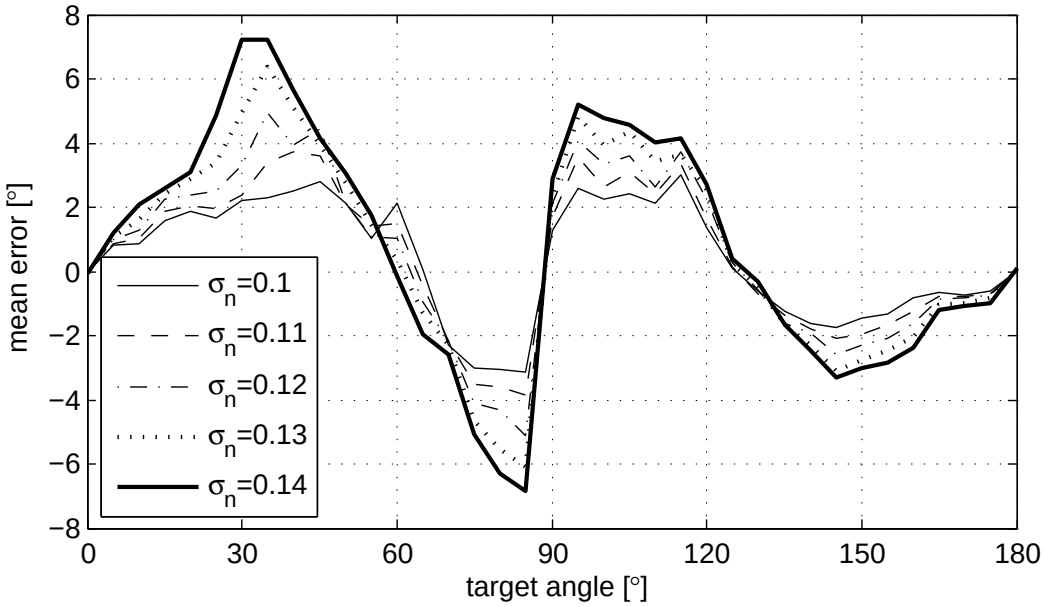


(a)

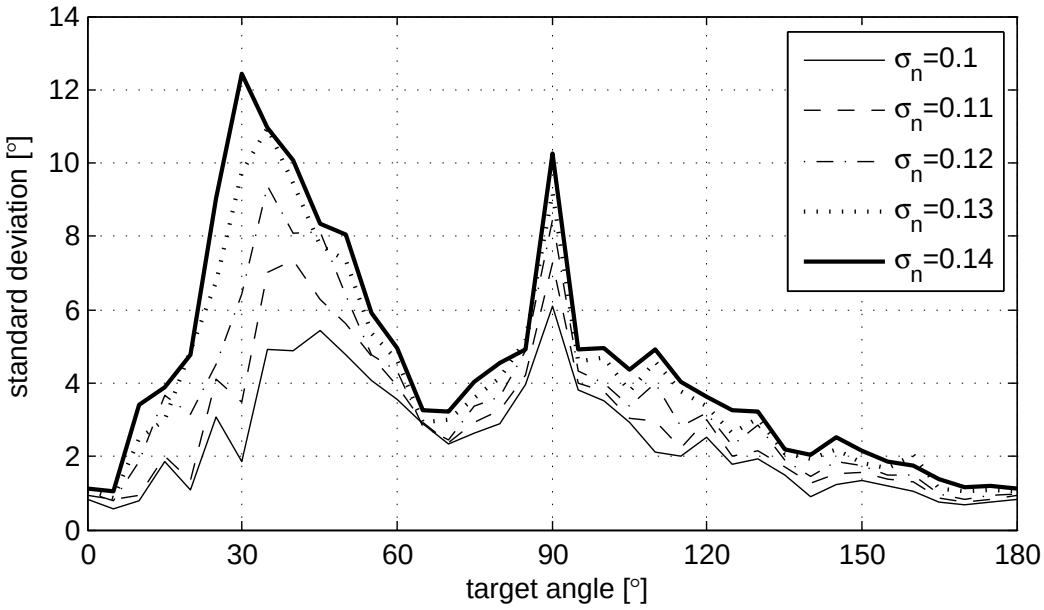


(b)

FIGURE 5.15: (a) Mean error and (b) standard deviation are shown from the result of the listening tests reported in Blauert [4] ($\star$), Carlile et al. [39] ($\circ$), Makous and Middlebrooks [38] ($\triangledown$). Positive error indicates that response angle is greater than target angle for $0° \sim 360°$. Data from Makous and Middlebrooks [38] correspond to 5°-elevation.

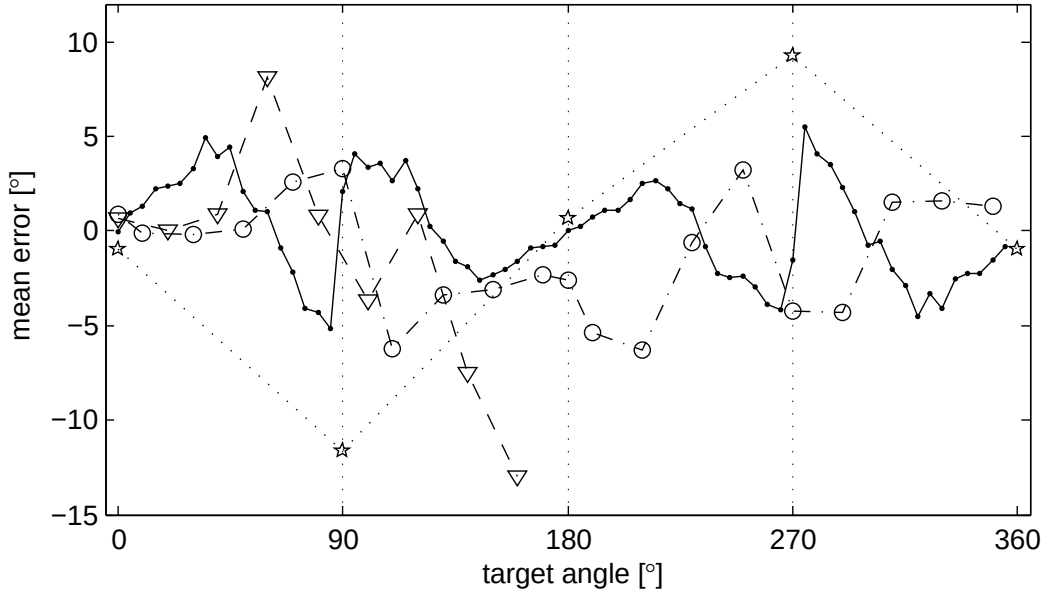


(a)

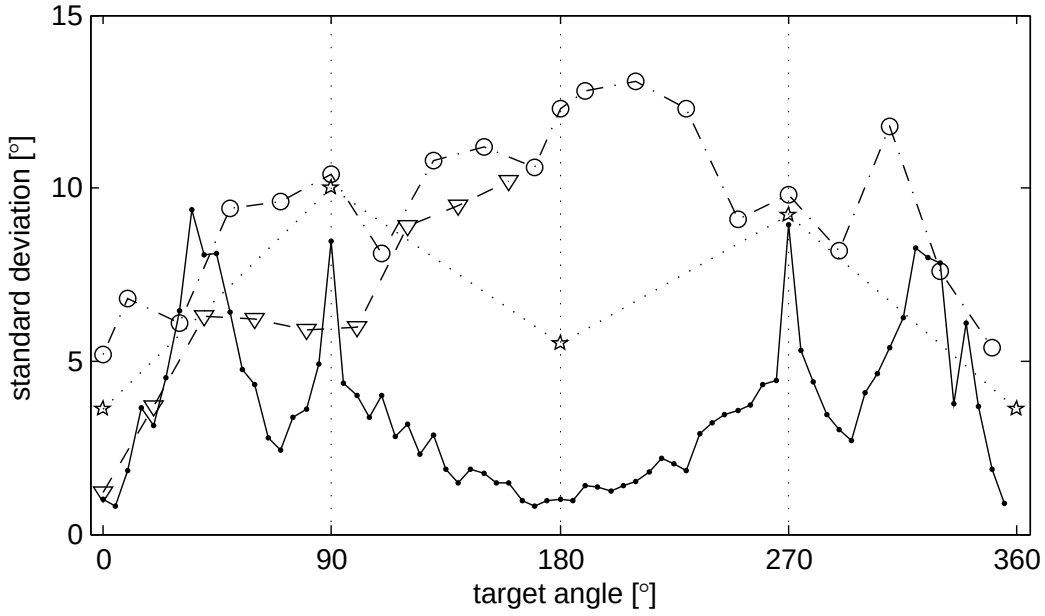


(b)

FIGURE 5.16: Influence of the internal noise on the localisation performance is shown for a white Gaussian noise as source signal. As σ_n increases, (a) mean error and (b) standard deviation increase.



(a)



(b)

FIGURE 5.17: Model performance in sound localisation is compared with some of the published listening test results. (a) Mean error and (b) standard deviation are shown for data points from Blauert [4] ($\star$), Carlile et al. [39] ($\circ$), Makous and Middlebrooks [38] ($\triangledown$), and the current model ($\bullet$) with $\sigma_n = 0.12$. Positive error indicates that response angle is greater than target angle for $0° \sim 360°$. Data from Makous and Middlebrooks [38] correspond to 5°-elevation.

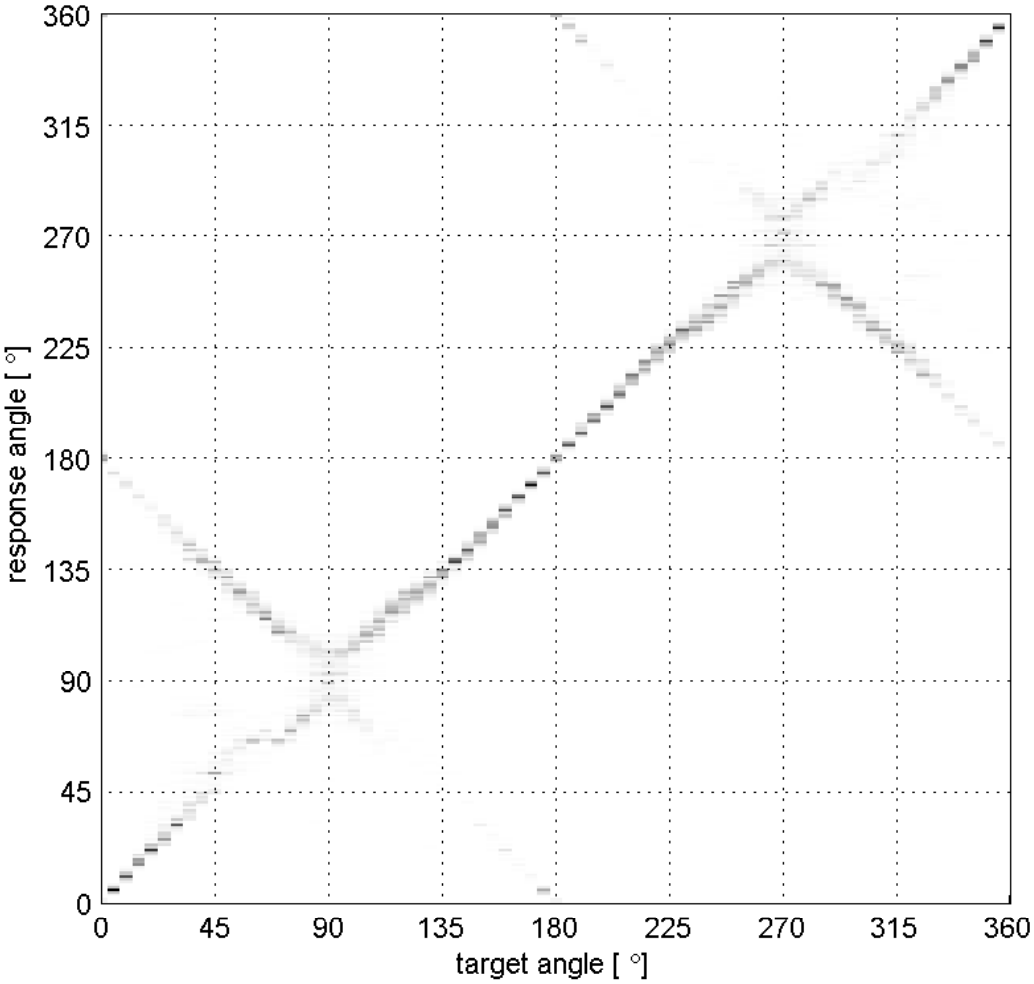


FIGURE 5.18: Model predictions ($\sigma_n = 0.12$) for a broadband sound source before the front-back confusion is resolved. Grey-scale indicates the relative frequency of the model responses along the vertical axis, which correspond to the target position at every 5° shown on the horizontal axis.

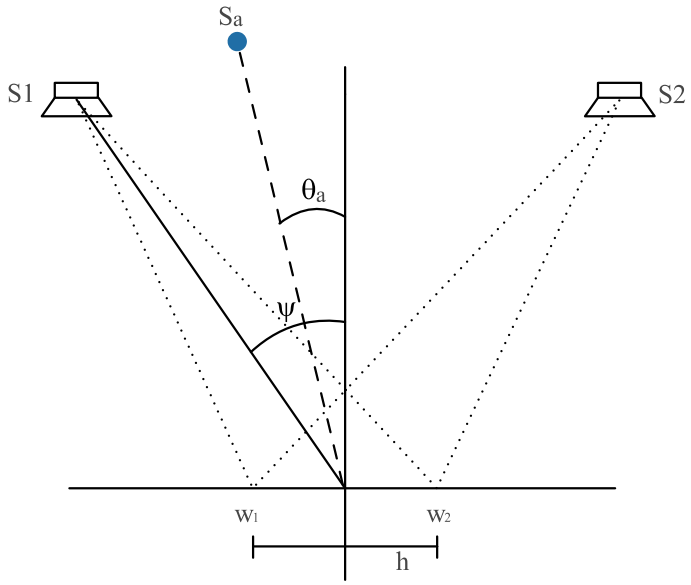


FIGURE 5.19: Configuration of stereophonic sound reproduction system. ψ, θ_a, S and w represent the half aperture angle between loudspeakers, the azimuthal location of the phantom image, and the locations of the loudspeakers and the receivers (ears), respectively. In conventional system, ψ is usually set to be 30°.

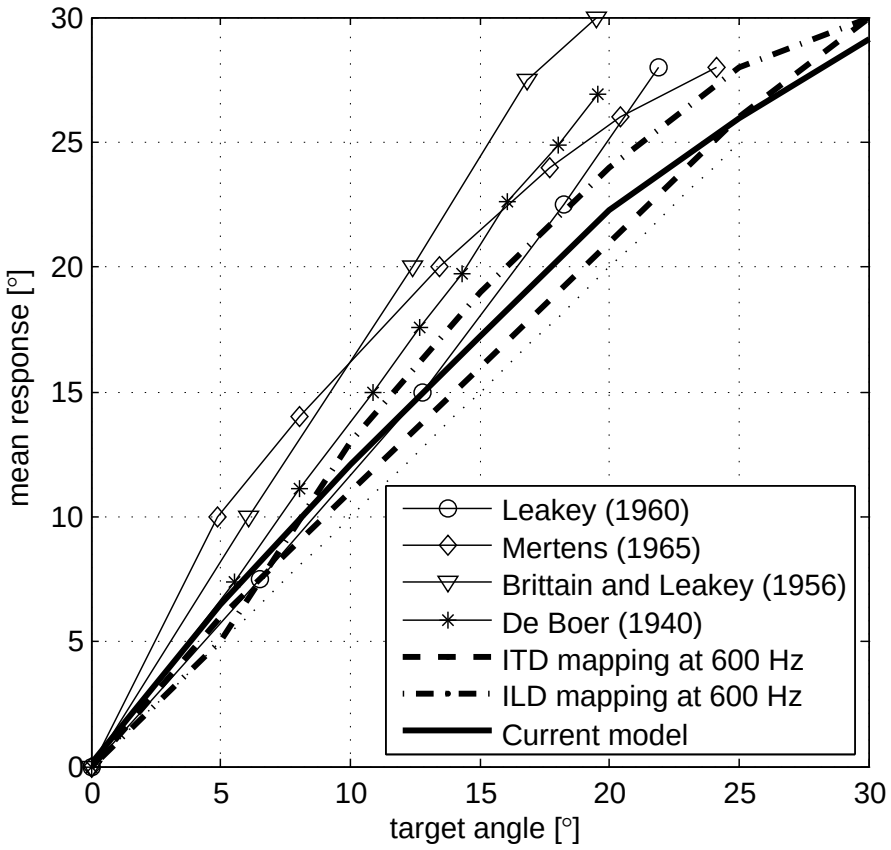
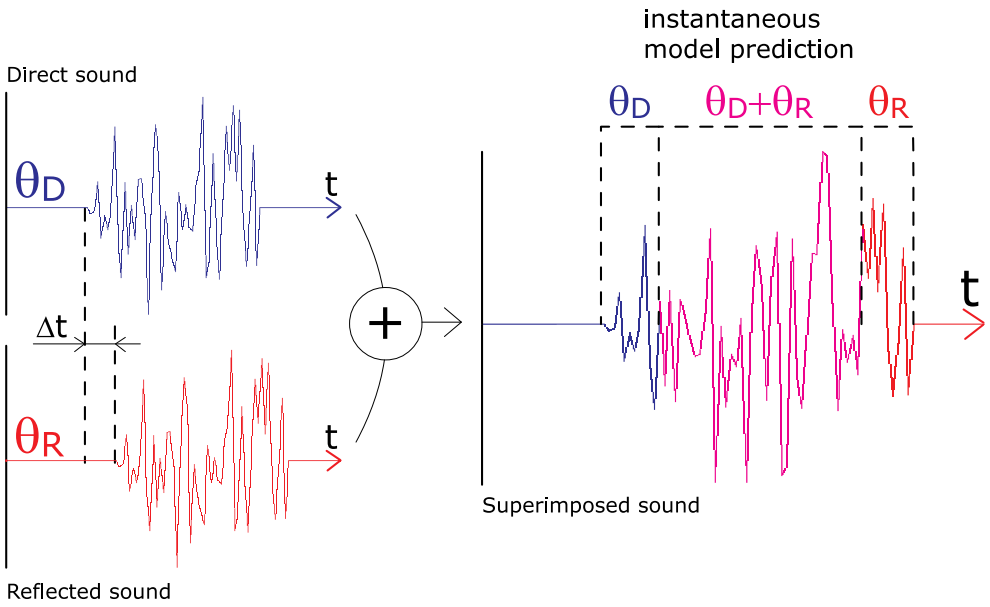
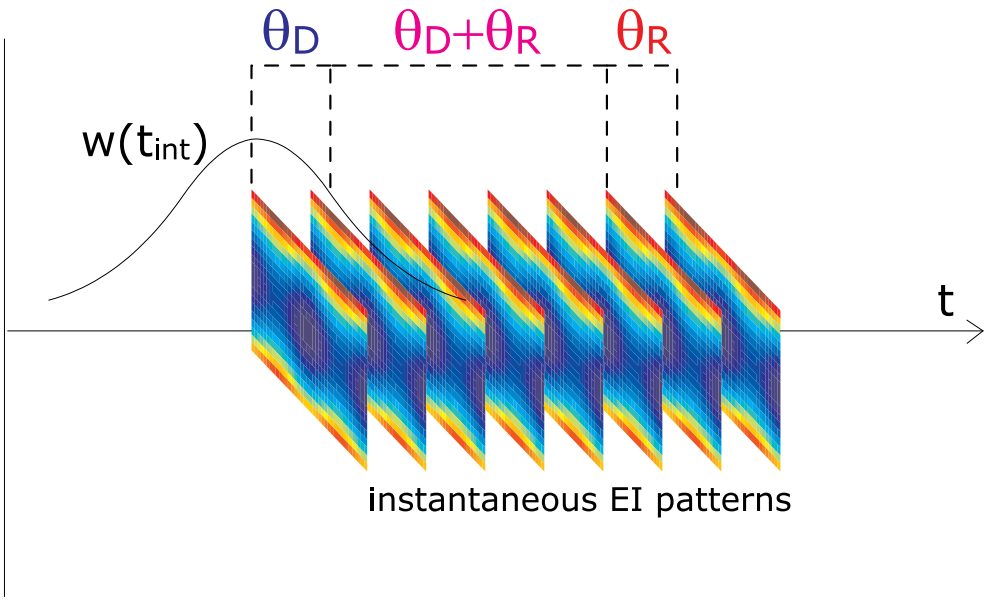


FIGURE 5.20: For the conventional stereophony based on the *sine* law, the averages of the model predictions are compared with the results of the listening tests cited in Rumsey [80] [from page 56; the horizontal axis has been modified from the intended ILD to the target azimuth angle using Eq. (5.8)]. In addition, two other estimates given by the ITD and the ILD mapping schemes are shown for a comparison.



(a)



(b)

FIGURE 5.21: (a) Model predictions for reverberant sound fields according to instantaneous EI patterns. Where the direct and the reflected sound waves are superimposed, the model prediction based on the instantaneous EI patterns might not reflect the precedence effect. (b) However, with the time integration according to Eq. (5.2), the influence of the EI patterns corresponding to the superimposed signals can be reduced.

Chapter 6

Listening test II - localisation of real and virtual acoustic images

6.1 Introduction

In chapter 5, a binaural hearing model has been suggested, which aims to predict the subjective judgements of acoustic image locations. Considering the EI-cell activity patterns [1] as the internal representation of the sound localisation cues, the model operates on the pattern-matching technique with a tentative frequency weighting scheme, and it has been shown to make reasonable predictions consistent with the test results in both real and virtual listening environments which have been reported in the literature.

The primary goal of the current listening test is to further investigate the reliability of the current model personalised for some of the subjects who have participated in the HRTF measurements described in chapter 3. In addition, it is intended that the spatial accuracy of virtual acoustic images can be evaluated in relation to the position and the angular aperture of the two-channel stereophonic systems, thus providing a practical insight to the optimal source positions for multichannel sound reproduction systems.

Human ability to localise sound sources has long been studied in many ways, and the results of some classical experiments are summarised in Blauert [4]. Dealing with mainly four directions on the horizontal plane, listeners in those days were instructed to move a loudspeaker to the position where they believed to be the target positions. In recent studies, the target area has been extended to two-dimensional spherical plane where localisation performance is investigated in both horizontal and vertical directions. Furthermore, there has been great improvement in the test equipment, and particularly,

the use of electromagnetic head-tracking device has enormously facilitated data acquisition at a greater level of accuracy (see, for example, Makous and Middlebrooks [38] and Carlile et al. [39]). Having achieved a reasonable account of the sound source localisation in normal listening situation, more recent studies examine different aspects of auditory spatial orientation, for example, the influence of the test sound level [81] and the presence of distractor [82].

As the relevant technology to provide virtual sound fields advances from the prototype stereophony system to recent multichannel systems, subjective evaluation of phantom images has been also of great research interest. As quoted in Rumsey [80], listening tests in early days have been mostly carried out to investigate the localisation of virtual images created by the conventional two-channel stereophony system. In more recent studies, the fidelity of virtual images has been tested for the lateral configuration of loudspeakers, where the evaluation and the optimisation of quadrophony and 5.1 channel sound reproduction systems were the primary objectives [83–85].

Essentially, the experimental study to be presented in this chapter is similar to the previous work briefly summarised above, where the subjective responses to acoustic images will be investigated in terms of spatial accuracy. Nevertheless, it is noticeable that, in the current listening tests, both real and virtual source localisation performance will be measured by the same participants in an identical test environment, so that the subjective judgement of virtual image positions may be investigated with reference to the baseline accuracy of real source localisation. In addition, a relatively wide range of stereophonic set-ups will be tested in the current experiment, including symmetric and asymmetric loudspeaker locations in listener's front, side and back, and accordingly, the accurate recordings of subjective responses obtained with the laser-beam assisted tracking device are expected to be very informative in the search for an optimal transducer configuration. Ultimately, it should be recalled that the primary objective of the current study is to compare the results of the subjective listening tests to the predictions of the established hearing model.

Following a brief introduction to the amplitude panning scheme employed in the test, a detailed description of the experimental arrangement will be made in section 6.2, where the test conditions will be categorised according to the proposed loudspeaker arrangements. In section 6.3, test results and the predictions of the pattern-matching model will be compared and discussed for each category of the conditions, while discussions will be followed in section 6.4 regarding the error analysis and the distinctive feature of the current model. Finally, some conclusions will be presented in section 6.5.

Contributors to this chapter: Kyeongok Kang (Electronics and Telecommunications Research Institute) and Filippo Fazi (ISVR) who participated in designing the

experiments, coded some part of the test interface in Matlab 7.0, and monitored actual listening tests.

6.2 Test method

6.2.1 Constant power panning

Using a pair of loudspeakers, virtual acoustic images presented in the current listening test were created based on the constant power panning (CPP) method [85]. Whereas the *sine* law [78, 79] (see appendix C) suggested in the original design of stereophony has been derived from the conversion of the inter-channel loudness ratio to the phase difference between two receiver positions in free space, the CPP method is simply established to guarantee that the total sound energy provided by the two transducers may remain constant as the position of a phantom image is controlled by the amplitude ratio to vary from one loudspeaker to the other.

The most elegant way of ensuring the constant power is the use of trigonometric identity [85, 86]:

$$\sin^2\theta_m + \cos^2\theta_m = 1 \tag{6.1}$$

where the amplitude gains for the 2-channel signals, g_1 and g_2 are given by

$$g_1 = \cos\theta_m \tag{6.2a}$$

$$g_2 = \sin\theta_m \tag{6.2b}$$

It is clear that, as θ_m varies from 0 to $\frac{\pi}{2}$, g_1 and g_2 vary from 1 to 0 and 0 to 1, respectively, while keeping the overall sound energy constant. The remaining task is to relate θ_m to the actual configuration of loudspeakers and the target location of a virtual image, which can be implemented by

$$\theta_m = \frac{\pi}{2} \times \frac{\theta_a - \theta_1}{\theta_2 - \theta_1} \quad , \qquad (\theta_2 \geq \theta_a \geq \theta_1) \tag{6.3}$$

where θ_1, θ_2 and θ_a represent the angular positions of the two loudspeakers and the phantom image, respectively, as illustrated in Fig. 6.1. In other words, θ_m is mapped between 0 and $\frac{\pi}{2}$ according to the ratio of the angular distance between the target position and one of the loudspeaker locations, to the angular aperture of the loudspeaker configuration.

6.2.2 Test arrangement

The current listening tests have been carried out in a small anechoic chamber located in Rayleigh building at the University of Southampton, which approximately measures $5\,m \times 5\,m \times 3\,m$ [see Fig. 6.2(b)]. Except for a part of the floor area where listener's

seat is positioned, all the surfaces of the room are treated with absorptive foam wedges to prevent sound reflections. Although the result of any detailed acoustic survey is not available, it is supposed that the test room can be regarded as being reasonably anechoic down to $200 \sim 300$ Hz. This chamber is annexed by a control room where most of the equipment was placed, and the experimenter had a CCTV facility to monitor the subject inside the room.

An array of 19 loudspeakers has been located in the test room at every 10° from 0° to 180° with respect to the room coordinate system shown in Fig. 6.2(a). (The manufacturer's data sheet claims that the frequency response of the transducer unit is reliable from 100 Hz to 20 kHz within ±2 dB.) The height of the array has been approximately adjusted to the average height of the subjects, where the radius from the loudspeaker to the array centre was measured to be 1.5 m. Since the visual cues given by the loudspeakers can bias the subjective judgments of acoustic image locations, the array has been covered by thin black curtains with rigid metal wires placed on top of each loudspeaker unit (see Fig. 6.3), which was extended beyond the loudspeakers at both ends approximately by $\sim 20\,cm$. In this way, empty space between loudspeakers can be fully disguised, and therefore, the locations of loudspeakers may not be recognised by the subjects.

A switch box has been custom-made for the current listening test, which is equipped with a micro-processor to separately route two-channel output from the PC soundcard to a selected pair of loudspeakers. While the micro-processor also receives a signal from a push-button that enables the subject to notify that he/she made a judgement, the switch box contains a built-in amplifier to provide sufficient power to the loudspeakers. A serial port has been used for the communication between the switch box and the PC with Matlab 7.0.

The source signal presented to the listeners is identical to what has been employed in sections 5.3.2 and 5.3.3 for the model simulation, where 6-ms rise-fall ramps have been applied to a 100-ms white Gaussian noise in a total of 150-ms sequence [see Fig. 5.11(b)]. The amplitude gains, g_1 and g_2, which were randomly given by the test design, have been then applied to this source signal, and the 2-channel output signals have been finally generated by a soundcard with D/A converter (RME ADI-2). When a single loudspeaker is used with a unit gain, the sound pressure level has been calibrated at the centre of the array to be 70 dBA.

For the subject to report the perceived image location, a head-tracking device, Polhemus

FASTRAK, has been employed in the current listening tests, which consists of a transmitter and a receiver, both connected to a control box. The transmitter can be considered as the origin of the coordinate system which, in this test, has been placed at the centre of the loudspeaker array below the subject's seat [see Fig. 6.3(a)], and the control box obtained translational and angular positions of the receiver, $(x, y, z, \text{azimuth}, \text{elevation}, \text{roll})$ relative to the transmitter. Since the current listening test is aimed to investigate the subjective perception of image locations on the horizontal plane, only x, y and azimuth information have been used.

Instead of wearing the receiver on the head, subjects held a wooden wand shown in Fig. 6.4(a) where the receiver and a laser pointer were attached to each end. In order to report the image location, the subject directed this device to where he/she perceived the acoustic image, and switched on the laser pointer to make a visible mark on the black curtain. Then, the subject pressed the push button [see Fig 6.4(b)], which triggered the control box of the head-tracker to send a single reading to the PC. Since the azimuth angle, θ_p' given by the head-tracker is not identical to the perceived image position, θ_p as illustrated in Fig. 6.4(c), a vector sum has been computed from the position vector (x, y) and the radius of the loudspeaker array.

A total of 10 university personnel (7 male and 3 female) have been paid for their participation in the current listening test, who are identified in the following sections as SA, SC, SD, SE, SF, SG, SH, SI, SJ and SK. The first five subjects also participated in the HRTF measurements (chapter 3) and the listening test for the lateralisation of dichotic pure tones (chapter 4), while the distal-region HRTF has been additionally measured for the subject SG (not presented in chapter 3). In pure tone audiometry which has been carried out recently or during the course of the current test, all participants have shown acceptable hearing ability across the audible frequency (less than 20dB hearing level).

This experimental study has been approved by the Safety and Ethics Committee of the Institute of Sound and Vibration Research (ISVR), University of Southampton (Approval number: 774).

6.2.3 Test conditions

In the current listening tests, there were four categories of test conditions. For the following specifications for each category, all angular locations have been represented in the subjective coordinate system where the 0° indicates the listener's front with the positive increment given to the right-hand side. It is also recommended to refer to Fig. 6.1 for the convention of symbols.

- **Category 1** - *Localisation of a single sound source*

  While only a single loudspeaker was being used throughout the two repeated sessions, a total of 10 subjective judgements (5 in each session) have been obtained for target locations from 0° to 180° at every 10°. Test results in this category can be regarded as the individual baseline performance of the sound localisation task.

- **Category 2** - *Five representative centre angles, θ_c with varying angular aperture, ψ*

  The centre of the two loudspeakers, θ_c was designed to be 0°, 50°, 90°, 130° and 180°, the interval between which is either 40° or 50°. The angular distance between the two transducers, 2ψ was 60° for θ_c =0°, 90° and 180°, while it was 40° and 60° for θ_c =50° and 130°. Having configured an active pair of loudspeakers, the target image location θ_a has been now controlled to vary from θ_1 to θ_2 every 10° except for the case of θ_c =0° and 180° where left-right symmetry has been assumed. As the total number of conditions is 39, test results in this category are expected to reveal the efficiency of conventional stereophony ($\theta_c = 0°$ with $\psi = 30°$) and similar arrangements in various positions. In addition, the influence of angular aperture can be investigated, which is expected to depend on the centre angle, θ_c.

- **Category 3** - *Detailed investigation for the lateral positions*

  In a preliminary study, it has been shown that it is hard to create convincing virtual acoustic images for lateral target positions. For this reason, the third category of the listening tests has been designed to investigate the influence of the angular aperture, ψ where θ_2 is always fixed at 90°. Test results in this category for 25 test conditions can be analysed in an attempt to give an optimal loudspeaker configuration for phantom images at lateral positions.

- **Category 4** - *Two adjacent loudspeakers in the 5.1 channel configuration*

  In the conventional 5.1 channel configuration (ITU-R BS.775), loudspeakers are located at 0° (C), 30° (R), 110° (RS), 250° (LS) and 330° (L). In the last category of the current test, two adjacent pairs of the loudspeakers at above five locations have been tested such that (θ_1, θ_2) was $(0°, 30°)$, $(30°, 110°)$, $(110°, 250°)$, where other combinations were discarded due to left-right symmetry. For each selection of loudspeakers, θ_a varied from θ_1 to θ_2 at every 10°. The listening tests in this category with 21 test conditions will investigate the efficiency of the conventional 5.1 channel system based on the constant power panning method.

While the loudspeaker array designed for the current listening test covers only a half circle, the above test conditions require loudspeaker configurations in subject's front,

side and back. Accordingly, the orientation of the listener's seat has been adjusted for certain test conditions, and, to minimise the number of seat relocations, the 4 categories of the test conditions have been rearranged to be in the 6 sessions listed in table 6.1, where the number of test conditions has been also balanced to be between 19 and 24.

In the beginning of each session, the subject's seat has been oriented as shown in Fig. 6.5, and then a voice message signalled the start of the session. A single trial started by another voice message instructing the listener to align his/her head to the 'X' mark on the 0° position (shown in Figs. 6.2(b) and 6.4 as a red cross on a small white piece of paper) with respect to the subjective coordinate system, after which a test signal randomly chosen from the test conditions listed in table 6.1 was played over selected loudspeakers. On hearing the test signal, the listener was instructed to point to the perceived image position with the wooden wand, confirm the location by the red laser beam, and press the push-button to confirm the decision and to move on to the next trial. All the procedure could be monitored by the CCTV facility and the Matlab interface shown in Fig. 6.6.

Category*	**No.**	(θ_c, ψ)	(θ_1, θ_2)	θ_a, (min, max)	**Session***	**Remarks**
1 (19)	1	n/a	n/a	$(0°, 180°)$	1 (19)	single source
2 (39)	2	$(0°, 30°)$	$(-30°, 30°)$	$(0°, 30°)$	2 (20)	5 centre angles
	3	$(180°, 30°)$	$(150°, 210°)$	$(150°, 180°)$	3 (24)	
	4	$(90°, 30°)$	$(60°, 120°)$	$(60°, 120°)$	4 (20)	
	5	$(50°, 20°)$	$(30°, 70°)$	$(30°, 70°)$	2 (20)	
	6	$(50°, 30°)$	$(20°, 80°)$	$(20°, 80°)$		
	7	$(130°, 20°)$	$(110°, 150°)$	$(110°, 150°)$	3 (24)	
	8	$(130°, 30°)$	$(100°, 160°)$	$(100°, 160°)$		
3 (25)	9	$(80°, 10°)$	$(70°, 90°)$	$(70°, 90°)$	5 (21)	θ_2 fixed at 90°
	10	$(75°, 15°)$	$(60°, 90°)$	$(60°, 90°)$	4 (20)	
	11	$(70°, 20°)$	$(50°, 90°)$	$(50°, 90°)$	5 (21)	
	12	$(65°, 25°)$	$(40°, 90°)$	$(40°, 90°)$		
	13	$(60°, 30°)$	$(30°, 90°)$	$(30°, 90°)$		
4 (21)	14	$(15°, 15°)$	$(0°, 30°)$	$(0°, 30°)$	2 (20)	5.1 ch. configuration
	15	$(70°, 40°)$	$(30°, 110°)$	$(30°, 110°)$	4 (20)	
	16	$(180°, 70°)$	$(110°, 250°)$	$(110°, 180°)$	3 (24)	

TABLE 6.1: Test conditions listed according to the categories described in section 6.2.3, where the second column is the **test condition number** specific to the particular loudspeaker configuration. *Numbers in the parentheses indicate the total number of test conditions and that of trials in each category and session, respectively. See Fig. 6.1 for the convention of symbols.

In each session, all test conditions have been repeated 5 times while being randomly presented to the subject, where session 1 for the localisation of a single real sound source has been repeated twice, giving a total of 10 subjective responses for each target location. 20 trials have been regarded as a block of tests, which was normally completed within 5 minutes, and a 5-minute break has been given to the subject every 1 or 2 blocks of trials. All subjects spent less than 1 hour on different days to complete a single type of session which is composed of up to 120 trials (190 for session 1).

6.3 Test results compared with model predictions

In this section, the results of the localisation listening tests will be presented along with the predictions of the pattern-matching model suggested in chapter 5. For the model simulations, EI-pattern templates have been first established with the individual HRTFs of the 6 subjects, SA, SC, SD, SE, SF and SG (see chapter 3), and the pattern-matching model has been run on Matlab 7.0 to give predictions for all categories of the test conditions listed in table 6.1. For each target acoustic image created with specific transducer configurations, the input gains, g_1 and g_2 were obtained according to Eq. (6.2) and applied to a white Gaussian noise shown in Fig. 5.11(b). Then, the sound propagation from the transducers to the listener's ears has been accounted for by the individual HRTF database to finally produce the synthesised binaural signals which are the input to the model. 500 model runs have been made with the internal noise, $n(i, t, \tau, \alpha)$ being a Gaussian random process ($\sigma_n = 0.12$). For the details of the pattern-matching model and relevant model parameters, readers are referred to section 5.2.

In the following two subsections, comparative analyses will be made between the listening test results and the model predictions separately for the localisation of real sound sources (category 1) and the localisation of virtual images (categories 2 to 4).

6.3.1 Localisation of real sound source: category 1

In order to present the acquired data effectively, regardless of their specific distributions, box-plots have been employed in Fig. 6.7 where the results of the single loudspeaker localisation are shown for the 6 subjects and corresponding model predictions. In each unit of a box-plot, the vertical edges of the box represent the first and the third quartiles, while the median and the highest/lowest values are indicated by the line within the box and the upper/lower whiskers, respectively. In addition, outliers beyond the whiskers are denoted by separate markers. In Fig. 6.7, the blue and the red plots drawn for the listening test results and the model predictions, respectively, have been paired for the convenience of comparison at each target position from 0° to 180° at every 10°.

It is of primary interest to investigate the listening test results (blue), and, firstly, it is observed in Fig. 6.7 that, despite the inter-subject variability, the reported image positions are mostly below the actual target locations. As such an underestimation of the source location can be found across all subjects including the other 4 participants whose data are not shown in Fig. 6.7, it is suggested that there could have been systematic biases involved either in the test procedures including the way the subject was positioned

or performed the required tasks, or in the geometrical accuracy of the measurement set-up, the latter of which is, however, unlikely. A detailed discussion regarding this issue will be made later in section 6.4.1.

In addition to the accuracy of the localisation, the variance of the reported image locations varies with subjects as shown in Fig. 6.7, but there appears to be a common trend that the variance increases as the target position approaches to 90°, from which it starts to generally decrease to 180°. In terms of both median and variance of the responses, most subjects are observed to have poorer localisation performance for the target locations in the rear hemisphere compared to their mirror-imaged positions in the front, and, even in the frontal hemisphere, the spatial resolution in the range from 0° to approximately 40° is exceptionally good, which might be regarded as the auditory counterpart of the retinal fovea responsible for sharp central vision. Nevertheless, there are a few remarkable cases, particularly with the subject SA where sound sources in the back were localised as correctly as those in the front.

The predictions made by individual models anticipate, as shown in red in Fig. 6.7, that sound sources can be localised equally accurately in the frontal and the rear hemispheres, and the target positions around 90° will be mostly overestimated, which contrasts to the subjective test results. For these reasons, the individual agreement between the two results is not outstanding, although there are many target conditions where the paired median values are found NOT to be statistically different as the associated box-plots are vertically overlapped.

It is also interesting that the frequency of front-back confusion is very much higher in the model simulation than in the listening test results as shown in Fig. 6.7. Indeed, there are just a few occasions in the test where listeners reported mirror-imaged responses, for example, at 0° and 10° for SC and at 130° and 170° for SF. This might possibly be linked to the absence of physical means to control the head position, thus head movement helping listeners to resolve the confusion, although listeners were instructed to direct to a reference point during the stimulus presentation.

On the other hand, from Figs. 6.8(a) and (c) in which the responses of all the 10 subjects are presented, it seems that the other 4 participants committed front-back confusions relatively more frequently than the 6 individuals shown in Fig. 6.7. Furthermore, as the dashed boxes in Figs. 6.8(a) and (b) highlight the responses corresponding to the front-back confusion, it is remarkable that mirror-imaged responses are mostly found for the target positions in the frontal hemisphere both in the listening test results and in the model predictions. Since the responses for the front-back confusion have been mainly made by those subjects whose pattern-matching models are not available to contribute to the histogram shown in Fig. 6.8(b), it is unlikely to conclusively confirm the consistency

between the two results regarding the specific direction of more frequent front-back confusions. Nevertheless, it is equally likely that the model simulations possibly reflect the actual auditory process where subtle internal errors result in the front-back confusion, perhaps more often in the frontal direction (see section 5.3.2).

Fig. 6.8(c) summarises the above observations as all the subjective responses and the model predictions are represented by box-plots in blue and red, respectively. The underestimation of the subjective responses and the overestimation of the model responses are clearly shown around 90° target position, while the population of the mirror-imaged responses can be also compared between the two results. It is noteworthy that there are no outliers in the model predictions for the target positions between 50° and 90° where the front-back confused responses greatly contribute to the significant variance, implying almost equally bimodal distributions.

Due to this bimodality in both subjective test data and model predictions, the analysis of mean responses can be made only after the front-back confusion is resolved. Being consistent with the way the mirror-imaged responses have been corrected so far in chapters 4 and 5, 90° (and 270°) has been considered to be the critical target location, at which corresponding responses are NOT resolved, whilst those for all other target positions are corrected to be in the same hemisphere as the target.

As the error bars indicate the 95% confidence intervals, Fig. 6.9 shows the averages of the subjective responses (blue) and the individual model predictions (red) after the front-back confusion is resolved. Similar to the box-plot analysis made in Fig. 6.7, it is apparent that the subjective judgements mostly underestimated the actual target positions, which is, however, not prominent in the model predictions. Furthermore, it is observed that the front-back correction resulted in an undesirable side-effect, especially for the target positions around 90°, perhaps between 70° and 120°, where even those responses normally distributed around the true target position were mirror-imaged due to the large variance. Given the large variance of the responses around the lateral target positions, it is uncertain whether the unconditional front-back correction with respect to 90° is inevitable, since it is true that such a correction can severely affect the final statistics of the data. In particular, the averages of the subjective responses shown in Fig. 6.9 make significant jumps between 90° and 100° target positions, which are apparently attributed to the side-effect of the front-back correction. Such a jump is also observed in the model predictions, most prominently with the models for SD and SE, but on a much smaller scale, and the comparison between the two results implies that the underestimated responses in the subjective test results are mainly responsible for the much steeper increases of the averages around 90°.

In each panel shown in Fig. 6.9, the averages of the model predictions, presumably the best estimates of the population means, are hardly found within the confidence intervals given by the subjective test data, which is also observed in the overall statistics shown in Fig. 6.10. The unsatisfactory agreement between the model simulation and the listening test results can be attributed either to the limited predictive scope of the current model or to the unknown systematic bias in the subjective tests. Comparison with the similar subjective test results reported in the literature suggests that the latter might be the actual case, as depicted in Fig. 6.11. In panel (b) of the figure, the standard deviations of the subjective responses from all the 10 participants are in a good agreement with those reported in the studies by Blauert [4], Makous and Middlebrooks [38] and Carlile et al. [39], while the model predictions show relatively lower variances as already pointed out in section 5.3.2. However, the mean error plot shown in Fig. 6.11(a) illustrates that the subjective responses in the current listening test do not agree with the recent data reported by Makous and Middlebrooks [38] and Carlile et al. [39] where Blauert's data perhaps require more target positions for a proper comparison. Being able to find similar up-and-downs in the mean response plots for both current and previous listening test results, it is suggested that the current test data are generally lower than the published data by about 5° to 10°, which implies that there could have been some systematic biases in the test results. As mentioned before, this issue of the underestimated target positions will be further dealt with in section 6.4.1.

The analysis of variance (ANOVA) has shown that both individual subjective responses and model predictions can NOT be regarded as being sampled from common normal distributions except for the target position at 0°, which suggests that the sound location judgements both in the listening tests and the model simulations were unique to each participant and his/her own pattern-matching model.

6.3.2 Localisation of virtual acoustic images: categories 2-4

Firstly, in Fig. 6.12, the test results and the model predictions are presented for the conventional stereophony system with a pair of loudspeakers at -30° and 30°, where the responses have been corrected for front-back confusion. As the averages and the 95% confidence intervals are represented by errorbars in Fig. 6.12(a) for each individual subject, it is clear that the subjective responses are mostly below the target positions indicated by the black dashed line, similar to the case of the localisation of a single loudspeaker presented in the previous section. Except for the subject SD, the mean responses for the 0° target position are found approximately where designed, but as the target moves to the 30° loudspeaker location, the underestimation becomes significant in general where the maximum mean error appears to be up to -10°. The statistics of the

responses vary from subject to subject, and in particular, the test result for the subject SF fluctuates most severely, which is particularly reflected by the widest confidence interval for the 10° target [see Figs. 6.7(e) and 6.9(e) for the baseline performance of this subject].

The pattern-matching models for individual subjects shown in Fig. 6.12(b) predict that localisation responses will be relatively close to the target positions, although there can be inter-subject variance. For example, mean errors are expected to be only up to ±5°, which tapers off as the target moves to the right-hand side loudspeaker. Apparently, the agreement between the subjective responses and the predictions of the corresponding model has been found unsatisfactory, which was the common case for all other loudspeaker configurations listed in table 6.1 (see section 6.2.3). Therefore, in the following paragraphs, only the global statistics will be analysed and compared between the model simulations and the subjective tests, where the prediction of a new analytical model will be additionally presented for a comparison. Based on assumption of the free-field sound propagation at a single frequency, this model essentially computes the interaural phase difference (IPD) given by the sound signals presented by a pair of loudspeakers, and equate it to the IPD function established for a single sound source, returning the estimate of the corresponding source azimuth angle. Readers are referred to appendix C for the details of this **IPD model**. In the current study, predictions of the IPD model have been obtained at 600 Hz, while the reliability of this analytical model will be discussed in section 6.4.3 in comparison with the current pattern-matching model.

- **Category 2 - Five representative centre angles**

  The results shown in Fig. 6.12 have been collected across subjects, and rearranged in Fig. 6.13. In panel (a), all subjective judgements have been presented as a 2D histogram where the scale of the colour contrast represents the relative frequency in each bin at every 1° along the vertical direction. Superimposed on this histogram, means and 95% confidence intervals of the subjective responses have been denoted as thick errorbars. Similarly, simulation results from all individual models have been illustrated in panel (b) of Fig. 6.13, where a thick dashed line has been distinctively used. Finally, the two results along with the predictions by the IPD model (thin dashed line) have been put together in panel (c) for a comparison, in which the results of the ANOVA for the pattern-matching models and the subjective responses can be additionally found as circled points. For example, the circled model mean at 30° target position indicates that it is statistically NOT rejected that predictions from the 6 individual models may originate from a common population, strongly implying the similarity between the simulation

results. From Fig. 6.13 to Fig. 6.19, the same plotting scheme has been employed to present the results of the virtual source localisation.

Centre angle, $\theta_c = 0°$ (Fig. 6.13) - It is clearly shown that the subjects underestimated the designated target positions, where the global mean errors are up to about -6°. The pattern-matching model as well as the IPD model predict that the stereophonic images will be perceived slightly closer to the median plane, but not as significantly as the subjective judgements. It is interesting to compare this result to the listening test data reported in the literature where the *sine* law has been employed to create the phantom images. As illustrated in Fig. 5.20, the stereophonic images given according to Blumlein's original idea [78] have been found to be perceived at greater azimuth angles than the intended location, both in the published subjective tests and the predictions of the current model. On the contrary, those images created by the constant-power panning method are perceived and predicted to be below the target positions, although the extent of the underestimation has to be considered in conjunction with the possible response bias in the test. If it is true that in the current listening test there was such a systematic bias so as to underestimate the target position, and if, therefore, the results of the pattern-matching model can be considered to be relatively reliable, the comparison of the model predictions for the two systems suggests that the constant-power panning method can present more reliable images with slightly higher spatial accuracy than the *sine* law.

Centre angle, $\theta_c = 180°$ (Fig. 6.14) - Compared to its counterpart in the frontal hemisphere, this rear set-up of the conventional stereophony system gives rise to more variance in the subjective perception of the image positions as the responses are spread widely across the target locations. The results of the ANOVA also reflect the greater variance of the responses where the statistical test has found that the judgements can NOT be considered to be different between subjects. In addition, it is observed that the mean of the subjective responses changes from under- to overestimation across about 170°, which could not be predicted by either model, where the predictions of the current model are very compact and mostly equivalent to or slightly greater than the actual target positions.

Centre angle, $\theta_c = 90°$ (Fig. 6.15) - When loudspeakers are located symmetrically with respect to the frontal plane, the amplitude panning method seems to be incapable of presenting convincing virtual images. Despite the varying amplitude ratio, subjective responses shown in Fig. 6.15(a) are mainly found around the louder transducer, slightly underestimated, except for the target position at 90° where the amplitude gains are equal. Since the original idea of the Blumlein's stereophony with the conventional frontal set-up was to convert the inter-channel

amplitude ratio to the interaural phase difference, it is apparent that the position of virtual image can not be controlled with the modified lateral set-up where the two loudspeakers signals reaching each ear are always in-phase due to the identical path lengths, thus the IPD being invariant, regardless of the amplitude ratio (see appendix C). Therefore, the predictions from the IPD model were constantly at 60°, which were unintentionally made consistent with the actual subjective responses only after the front-back correction. On the contrary, it is remarkable that the pattern-matching model has made very good predictions for the phantom image positions with the lateral loudspeaker configuration. As Fig. 6.25 shows the raw results of the model simulation before the front-back correction, it is suggested that the current model successfully incorporated the associated interaural level difference (ILD) to resolve the ambiguity of the IPD, where the ILD is often regarded as the by-product of the amplitude panning scheme only in the high frequency region resulted from the head-shadowing (see section 6.4.3 for further discussion). In addition, the comparison between the simulation and the listening test results shown in Fig. 6.15(c) gives a strong indication that the subjective responses could have been biased to underestimate the target locations, as the difference between the average responses is relatively constant throughout the target positions.

Centre angle, $\theta_c = 50°$ (Fig. 6.16) - Mixed in some bins according to the relative frequencies, the two colours in Figs 6.16(a) and (b) represent the results given for a fixed centre position at 50° but with different angular apertures, 40° (blue) and 60° (red). To examine the subjective responses first, it is clear that the underestimation of the target positions are more significant with the wider-aperture loudspeaker configuration. Furthermore, it is suggested that, as the target moves towards the loudspeaker on the far side, the perceived position quickly migrates to the side of the louder transducer, the 'speed' of which is closely related to the spacing between the loudspeakers. In other words, the identical change in the target position will give rise to a greater increase in the perceived location for the target positions near the far-side loudspeaker, and the rate will be higher when the loudspeaker span is wider. Both observations can be also made for the simulation results shown in Fig. 6.16(b), where the distinction between the two loudspeaker configurations is more prominent, although the extent of underestimation is, similar to the previous test conditions, less than that in the actual listening test results.

Centre angle, $\theta_c = 130°$ (Fig. 6.17) - Being the counterpart to the previous test configuration with 50° centre angle, the results shown in Fig. 6.17(a) appear to be similar to those presented in Fig. 6.16(a), except for the relatively greater variance observed at the intermediate target positions. On the other hand, the model predictions shown in Fig. 6.17(b) are in clear contrast to those given for the

frontal hemisphere in Fig. 6.16(b), where the target positions are overestimated, and more overestimated with a wider loudspeaker span. The IPD model gives the same indication of overestimation, and in fact, there are some factors that could challenge the subjective test results. For example, the vertical distance between the two mean plots in Fig. 6.17(a) is very small compared to that in Fig. 6.16(a), and at 150°, the order of the plots is even swapped, perhaps suggesting that the mean responses for the wider angular aperture might be actually greater. Also considering the greater variance and the more significant underestimation for the rear target positions shown in the baseline localisation performance, a further experimental study might be able to reassess the reliability of the model predictions for these particular loudspeaker configurations.

- **Category 3 - θ_2 fixed at 90°**

  Preliminary simulation studies revealed that the lateral images provided by the two front-back symmetric sound sources may not be reliable, which have been confirmed to be true in the subjective listening tests in test condition no. 4 (see table 6.1). Accordingly, the current loudspeaker configurations have been designed to investigate the possibility of delivering convincing lateral sound images at the cost of an additional loudspeaker at 90°.

  Fig. 6.18(a) shows the results of the listening test where a total of five test conditions (for five loudspeaker spans) are separately colour-coded. Due to the heavy mixture of the different colours, it is difficult to obtain any usable information from the 2D histogram. However, the errorbars representing the mean responses and the 95% confidence intervals show a clear relationship between the subjective responses and the loudspeaker spans. While all the mean plots are below the dotted reference line, the extent of underestimation becomes more significant for wider angular apertures, where the rate of image shift to the lateral side is also higher. It is not surprising to see that this result is consistent with one of the category-2 test conditions shown in Fig. 6.16, since both cases involve a loudspeaker positioned at the relatively far side.

  The model simulations shown in Fig. 6.18(b) are very impressive in predicting the features in the subjective test results except for the overall downside shift. The vertical order of the mean plots is consistent in both results, and the spacings between nearby lines are also comparable. In addition, the predicted slopes of the mean responses near the far side are very similar to the empirical values, although 'the region of convergence' seems to have shifted to be between 80° and 90°, contrasting to the range from 70° to 80° in the listening test results.

From both model simulations and subjective listening tests, it is clear that an additional transducer at 90° will increase the reliability of the virtual images in the lateral region, which, undoubtedly, further improves as the angular distance between the transducers becomes narrower.

- **Category 4 - Conventional surround system with 5.1 channels**

  It is first recalled that, similar to the way commercially available 5.1 channel systems operate, only the nearby pairs of loudspeakers have been employed in the current listening tests to create virtual images in the associated angular range, and the localisation judgements have been obtained only in the right hemisphere, assuming left-right symmetry.

  Fig. 6.19(a) shows the results of the subjective tests, which have been combined from the three different test configurations, colour-coded in blue, red and green for C-R ('centre' at 0° - 'right' at 30°), R-RS ('right' at 30° - 'right surround' at 110°) and RS-LS ('right surround' at 110° - 'left surround' at 250°), respectively. Firstly, the performances of the C-R configuration is considered to be equivalent to the similar set-up for conventional stereophony shown in Fig. 6.13, where the narrower loudspeaker span does not seem to have improved the image fidelity. In a similar comparison between the R-RS configuration and the set-up of loudspeakers at 30° and 90° presented in Fig. 6.18, it has been commonly found that the perceived image location slowly moves to the side up to 70° target position, and then it suffers from a rapid jump between 80° and 90°.

  For the target range beyond 110° for the RS-LS configuration, subjective responses quickly move towards the median plane, and it seems that subjects had difficulties in locating the target sound images in the rear, resulting in greater variance compared to the test conditions in the frontal hemisphere. Also in terms of the mean responses, it is noticeable that subjects tended to overestimate the target position in the rear hemisphere, even when judging the position of the source at 180°, which has been similarly observed in the results of the 60°-span rear set-up shown in Fig. 6.14 (test condition no. 3). Since the IPD cues provided by the same left-right symmetric loudspeaker set-up will be almost identical regardless of whether it is positioned in the front or in the rear, the inaccuracy and the instability of the subjective responses for the rear sound images can be solely attributed to the hearing process, the motor-sensory process or the physical limitation of body movement, although it is beyond the scope of the current study to define the relative dominance of these factors.

It is remarkable to see that the pattern-matching model describes the subjective responses very well in Fig. 6.19(c) except for the 'problematic' rear target positions. The underestimation followed by the rapid migration to the side has been reasonably predicted for the frontal targets, and the slight underestimation also observed in the beginning of the third mean plot (green) is also consistent between the two results, at least qualitatively. In particular, the model predicts that the perceived image locations can be ambiguous for the target images around 90°, where the predicted mean values show a prominent peak. For example, different inter-channel amplitude gains intended to present images at approximately 90° and 110° can be perceived as an identical virtual image located around 105°. Such degeneracy of the image positions is also identified in the subjective test results in the similar range of target positions, where the relatively wide confidence interval also reflects the ambiguity of the responses.

It is of further interest to compare the current results for the 5.1 channel surround system to similar recent data reported in the literature. In Fig. 6.20, subjective test results obtained by Martin [84] have been reproduced in panel (a), where the horizontal axis has been converted from inter-channel amplitude ratio (IAD) to the target angle in accordance with Eq. (6.2). As the current listening test results and the model predictions are also presented as box-plots in panels (b) and (c), respectively, it is first observed that the subjective responses obtained in this study are significantly downshifted relative to Martin's data [84], and the extent of underestimation appears to be approximately up to 10°, which is consistent with the estimation suggested in the analysis of the single source localisation results in section 6.3.1.

In addition to the overall downshift, the current test results are distinguished from those published in a few more respects. For example, in Martin's data, the rapid jump of the perceived image location is identified roughly in the middle of the loudspeaker positions, where a large variance can be found at 70° target position with very considerable distance between whiskers. However, the current subjective listening tests imply that the region of indeterminate image locations is significantly inclined to the side, so that the greatest uncertainty may be found around the 90° target position. In other words, the rapid jump of the virtual image takes place when the IAD is nearly 0 dB in Martin [84], but it is identified in the current results when the IAD slightly favours the far side loudspeaker, which is also consistent with the model predictions shown in Fig. 6.20(c). Additionally, the current test results more clearly shows that the mean responses for the far-side target positions will have a prominent peak around 90° [see Fig. 6.20(b)], which has been related to the image degeneracy, and, as discussed above, the model

predictions in panel (c) also support the particular shape of the mean plot for the lateral target angles.

6.4 General discussion

6.4.1 Underestimated target position

In section 6.3, it has been occasionally suggested that, compared to similar listening tests reported in the literature, subjective judgements in the current listening tests could have been biased to underestimate the target image locations. For example, as shown in Fig. 6.11, the mean responses obtained in the single source localisation tests were significantly less than the values suggested in the studies by Makous and Middlebrooks [38] and Carlile et al. [39], while the responses for 5.1 channel configuration were also found to be consistently below those reported in Martin [84] (see Fig. 6.20). In addition, the pattern-matching model as well as the IPD model gives a similar indication, particularly by predictions made for the lateral set-up of 60°-span loudspeakers, as presented in Fig. 6.15.

Arguably, auditory spatial perception is not self-representative, but usually requires other sensory process to manifest its processing results. For instance, eye movement can be triggered to 'see' the point where the hearing system located a sound source, and in many cases, head or whole-body movement is also involved in such an operation. Therefore, the results of the sound localisation listening tests reflect not only the accuracy of the auditory spatial perception but also that of motor-sensory process and its neural liaison with hearing process. As the physical limitation in reporting the sound sources in the back can be an additional factor to influence the results of the localisation tests, it is arguably very difficult to investigate the performance of the hearing process exclusively.

However, it is still meaningful to compare the results of similar listening tests, since the required tasks are mostly identical to each other, where the characteristics of all the relevant brain processes and the physical operations are assumed to be equally incorporated. Having suggested so, the most likely factor that could have resulted in the relative underestimation in the current test data is considered to be the absence of training schedule. Whereas there was no training session carried out in the current test, experimental studies reported by Carlile et al. [39] and Makous and Middlebrooks [38] included considerable time of pre-test training where subjects received visual feedback for their judgements of sound source locations. In those tests, training sessions were designed to minimise subject's possible use of eye movement which was suggested to disrupt the measurement results recorded with the head-tracking device. While such training programmes must have effectively reduced the bias effect of concern, it is also suggested that the subjects' sound localisation accuracy could have been improved, perhaps their auditory/motor-sensory spatial maps being restructured. Nevertheless, considering the difficulty in isolating the individual cognitive factors in the subjective

tests, it is arguable whether the human performance of auditory sound localisation can be better investigated with trained or naïve subjects.

It is also possible to relate the subjective bias in the current listening tests to the absence of head-restraint. Although subjects were instructed to keep their head position at the array centre during the stimulus presentation followed by the reporting procedure, it is unlikely that the perfectly correct position is achieved without any monitoring facility. Particularly, it is likely that, in the beginning of each trial, subjects maintain upright posture paying attention to the upcoming stimulus (who could be already off centre though), but after the signal being presented, they might slightly ease off in using the pointing device, which often involves a whole-body translational or rotational movement. Assuming that subjects remember and report the direction of perceived sound image with respect to their head at the time they were exposed to the sound, post-stimulus movement could give rise to a systematic bias effect to the test results. For example, Fig. 6.21 shows the case where listener makes forward-backward movement, Δy from the array centre after the stimulus. Although the perceived source position is θ_p [panel (a)], the reported position after the displacement will be $\widetilde{\theta_p}$ [panel (b)] as recorded by the head-tracking device with reference to the origin. Where Fig. 6.21(c) shows the relationship between the true subjective judgement θ_p and the biased recording $\widetilde{\theta_p}$, it is noticeable that the target position can be underestimated by the listener's post-stimulus movement in the forward direction, which is the usual direction of listener displacement for a relaxed posture. Assuming $\Delta y = 20\,cm$, the subjective judgements obtained for the single source localisation tests have been tentatively compensated and redrawn in Fig. 6.22 where the agreement to the published data has been considerably improved, especially in terms of the mean errors. It is unlikely that such a significant positioning error, $\Delta y = 20\,cm$ could be unnoticed by the experimenter, but the above error analysis appears to be still useful, considering the case where possible pre-stimulus positioning error could be combined with the post-stimulus displacement.

6.4.2 Influence of visual cues on the subjective response

As presented in section 6.2.2, the loudspeaker array has been covered to prevent any visual cue to affect the test results. Black curtains treated with thin pieces of metal wire (see Fig. 6.3) successfully disguised each loudspeaker so that participants' attention may not be visually attracted to result in discrete responses. Nevertheless, since the loudspeaker array was only a half circle, it has been of concern, despite the considerable end-margin of the metal-wire treatment, whether there would be subjective bias in responding to real or virtual sound images which were positioned near the first or the last transducers. Especially, the single source localisation tests could be more vulnerable

to this possible 'edge-effect' where even the loudspeakers at 0° and 180° had to be used for stimulus presentation.

In order to investigate the presence of the edge-effect, the test results for the condition no. 1 have been compared to those for no. 14 (see table 6.1), where the subjective responses to a real sound source at 0° have been obtained in both cases, but by using different loudspeakers at far-side (no. 1) and in the middle (no. 14). Fig. 6.23 shows the statistics of the subjective judgements for 0° source position where box-plots in blue and red indicate the results using the far-side and the middle loudspeakers, respectively. Although the distance between the first and the third quartiles appears to be greater for some participants when the source in the middle of the array was in use, it is unlikely to be concluded that there has been an edge-effect associated with the judgement of the far-side source position, since, for 6 participants (SA, SD, SE, SF, SI and SJ), it is NOT rejected with 5% significance level that the median values given in two cases reflect an identical population, as the blue and the red boxes plotted for those participants are vertically overlapped. A similar analysis for 180° source location could have been very helpful, which is, unfortunately, not available due to the lack of the test data where the mid-position loudspeaker was employed.

6.4.3 Pattern-matching model vs. IPD model

The comparison between the subjective responses and the simulation results depicted in panel (c)'s in Figs. 6.13 through 6.19 seems to indicate that the IPD model might be able to give as successful predictions as the pattern-matching model, and this could imply that the IPD model based on a simple analytic equation (see appendix C) is far more efficient than the pattern-matching model that requires very heavy computation. However, there are a few important distinctions between the two models, which make the pattern-matching model outstanding. Firstly, the IPD model is based on the free-field assumption to consider the transfer characteristics from the transducers to the listeners' ears, which has been shown in many studies to be insufficient to take into account the complex interaction between the sound field and the subject's head and torso. In addition, the IPD model is valid only at low frequencies where the interaural phase difference is not ambiguous, and even in the low frequency region, its predictions heavily depend on the frequency as shown in Fig. 6.24. Therefore, the IPD model requires an additional companion process to handle the localisation information present at high frequencies, especially the ILD cues given by the head-shadowing effect which is, in the amplitude panning scheme, considered to be a by-product.

For the reasons suggested above, there can be test conditions where the IPD model is incapable to make reliable predictions, and in fact, the test result with the front-back symmetric loudspeaker positions is the extreme example where the free-field assumption has failed to operate. Since the IPD given by Eq. (C.11) for the loudspeaker set-up at 60° and 120° (test condition no. 4) is constant regardless of the target position while the high-frequency ILD has not been taken into account, the prediction by the IPD model at 600 Hz was constant at 60° [the red circles in Fig. 6.25(b)], which has been unintentionally made indicative of two positions at 60° and 120° only after the front-back correction [see Fig. 6.15(c)]. On the contrary, it has been shown in Figs. 6.15(c) and 6.25(b) that the pattern-matching model successfully copes with the extreme condition, making predictions that reflect the actual subjective responses even before the front-back correction. Such an inherent capability to resolve the front-back confusion can be attributed to the fact that the pattern-matching model takes into account both ITD and ILD information from low to high frequencies where the two important localisation cues are considered to be closely related to each other with reference to their natural combinations that can be slightly different for the sound sources in the front and in the rear.

Fig. 6.26 presents a detailed account of how the pattern-matching model incorporates the ILD information at high frequencies. The grey-scale of the 2D histogram in panel (a) shows the cross-correlation between the EI-template and the target EI-patterns that have been obtained for a virtual sound image positioned at 20° with a conventional stereophony system (test condition no. 2). As the white circles indicate the estimate of the target position made in each auditory frequency band, the model predictions below approximately 1500 Hz are similar to those given by the IPD model that are shown in Fig. 6.24, where, in both cases, the estimated target position slightly increases with frequency (if front-back confusion is assumed to have been corrected). On the other hand, the pattern-matching model gives a somewhat confusing prediction between 2 kHz and 3 kHz that the perceived image positions will be scattered around 60°, which might reflect the transition from the ITD-dominated to the ILD-dominated region for auditory spatial processing, often associated in the literature with the gradual loss of phase information. Above 3 kHz, however, the local estimates indicate again that subjects will perceive a phantom image approximately at its target position, where the ILD arbitrarily given by the head-shadowing effect is considered to be the main localisation cue.

Despite the puzzling estimates in the mid-range frequency, the weighting scheme employed in the current model successfully operated to obtain the probability function of target location shown in Fig. 6.26(b), since there are more auditory frequency bands populated at low frequencies, while the frequency weighting function associated with a white Gaussian noise (see Fig. 5.10) is relatively uniform up to 3 kHz. It is suggested

that subjective listening tests investigating the influence of various frequency contents of the sound source can be designed as future work, which are expected to confirm the above discussed reliability of the current model, particularly whether the predictions given between 2 kHz and 3 kHz are trustworthy.

6.5 Conclusion

In this experimental study, subjective judgements of acoustic image locations have been investigated for real and virtual sound sources. A white Gaussian noise stimulus has been presented over one or a pair of transducers installed on a half-circle array, where the perceived image locations have been reported by using a pointing device equipped with a head-tracker. For virtual sound sources, stereophonic images have been created based on the constant power panning (CPP) method.

The target location in the single source localisation test has been mainly underestimated by the subjects up to 14° on average, which was consistently observed in the test of virtual source localisation. From the comparison to the similar tests reported in the literature, it has been suggested that the absence of training schedule and the lack of position-monitoring system could be responsible for the prominent underestimation of the acoustic target position.

For the conventional stereophonic system, the CPP method has been found to create a better (although slightly underestimated) virtual image than the *sine* law in terms of the spatial accuracy, while the latter method has been reported in the literature [80] to give phantom images often overestimated by listeners. Furthermore, the target image position has been found to be more underestimated as the loudspeaker span becomes wider, particularly when the CPP method is implemented by the stereo system located to the side of the subject. In extreme case where loudspeakers are positioned front-back symmetric at 60° and 120°, it has been found that the amplitude panning method fails to control the location of a virtual image. Finally, the test results of the conventional 5.1 channel set-up showed that the virtual image makes a rapid jump on the relatively far-side in a R-RS configuration, where the indeterminacy of the image position also increases. Compared to the similar study by Martin [84], the current test results have been found to be equivalent, or perhaps better, in describing the actual subjective judgements.

Although the individual link could not be established between subjective judgements and his or her own model predictions, the global statistics of the simulation results showed a reasonable agreement with the listening test results for most of the test conditions investigated in the experiments. The reliability of the model has been also demonstrated in many cases, and in particular, the comparison to the IPD model showed that the current pattern-matching model successfully incorporated the ILD information across frequency, resolving the ambiguity in the IPD, and thus front-back confusion. A further experimental study using sound sources with various frequency contents is expected to

validate the reliability and the extended predictive scope of the pattern-matching model, which have been partly confirmed in the current study.

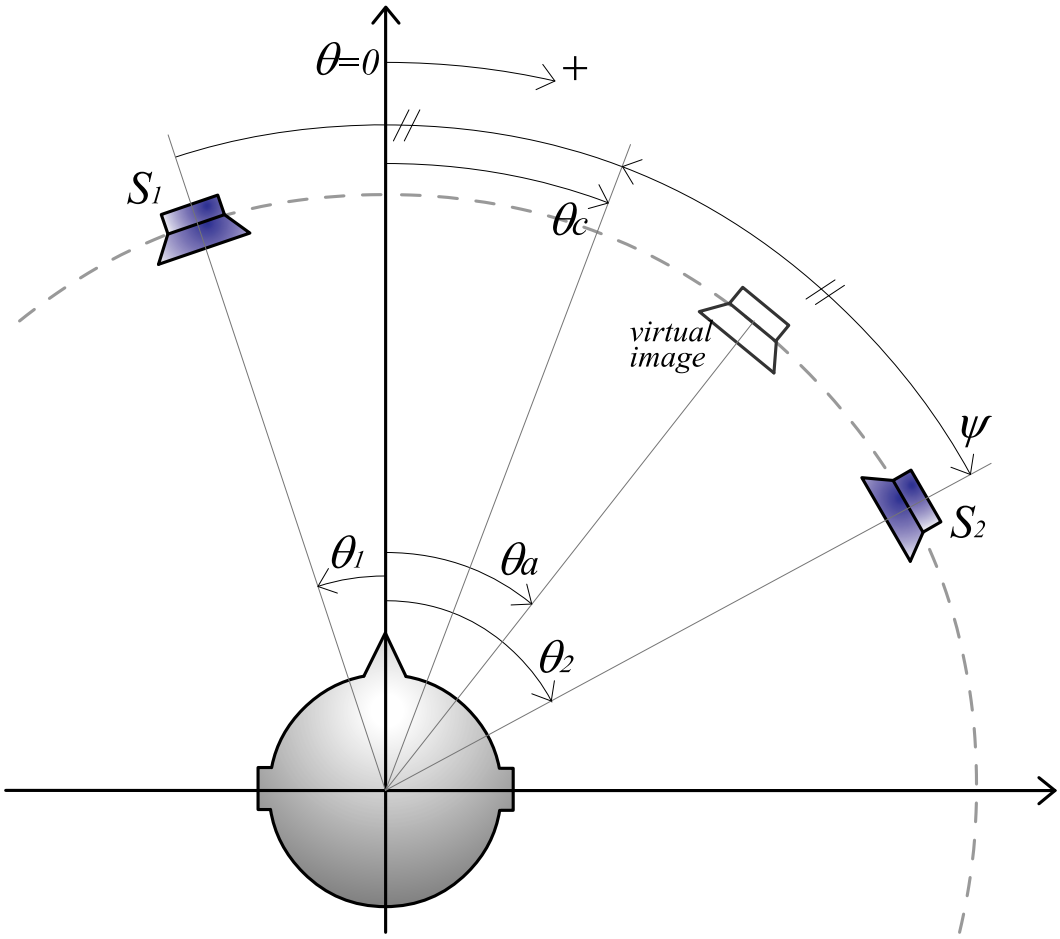
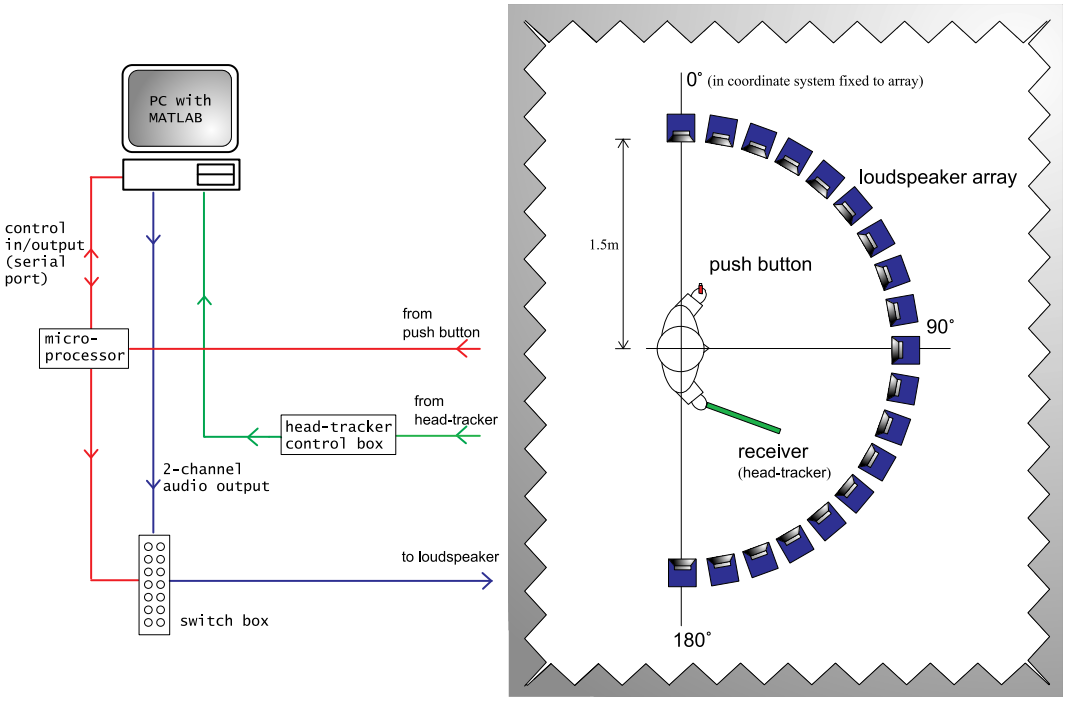


FIGURE 6.1: Stereophonic configuration to create phantom images. S represents each loudspeaker while a virtual image is positioned between the two transducers. θ_c represents the centre position of the two loudspeakers. All angle notations are made with respect to the coordinate system where the subject's front is 0°. Symbols will be used consistently throughout this chapter.

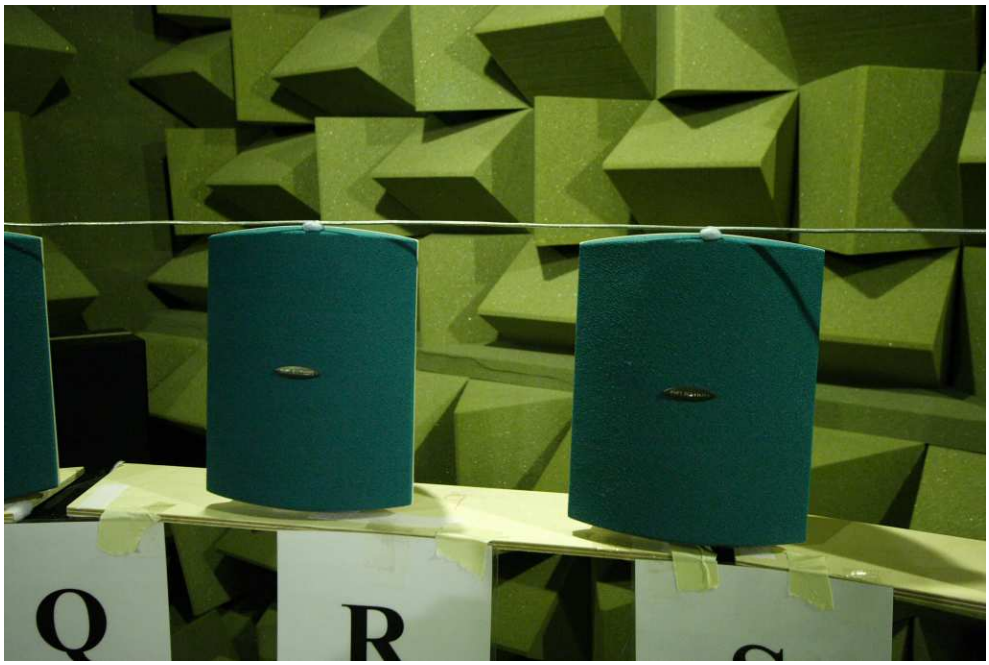


(a)

(b)

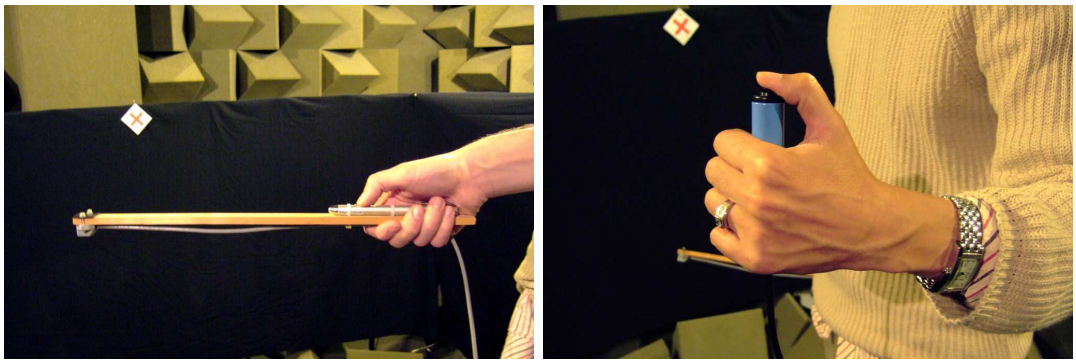
FIGURE 6.2: (a) Diagram illustrating the test and the control rooms. In this diagram, 0° in subjective coordinate system is in the same direction as 90° in room coordinate system. (b) Photograph taken on site.

(a)

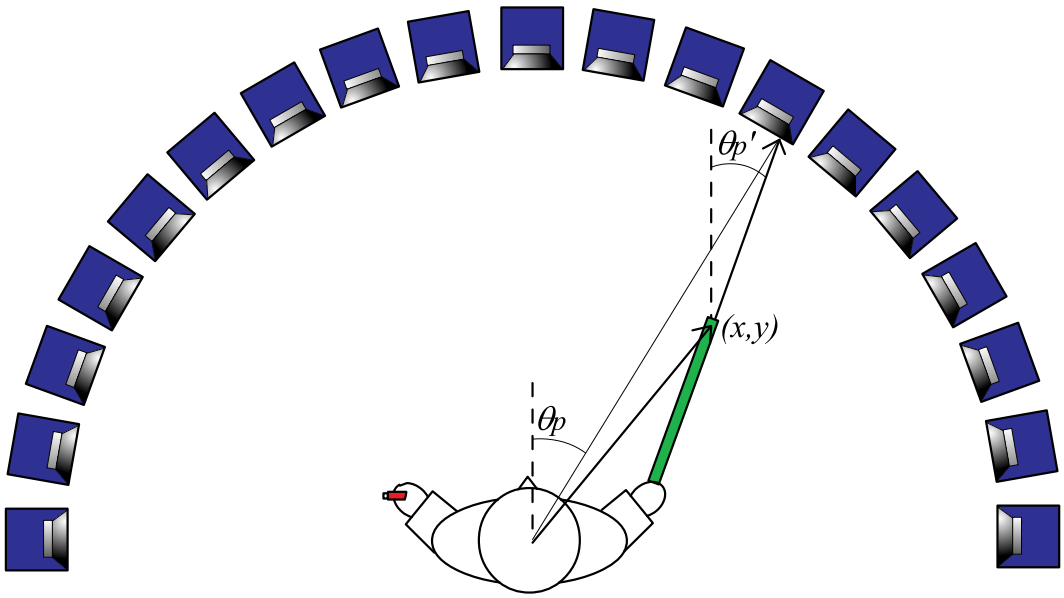


(b)

FIGURE 6.3: Treatment to remove the visual cues of the loudspeaker locations. (a) The loudspeaker array has been covered by black curtains (b) where pieces of metal wires were attached on top of each unit to disguise the space in-between. In panel (a), the transmitter of the head-tracking device is shown as it is positioned at the centre of the half-circled array below the subject's seat.



(a) (b)



(c)

FIGURE 6.4: (a) Pointing device equipped with the head-tracking receiver and a pen-shaped laser-beam. (b) Push button to confirm the judgement. (c) The position and the angle recordings by the head-tracking device have to be converted to the response angle, θ_p with respect to the subjective coordinate system.

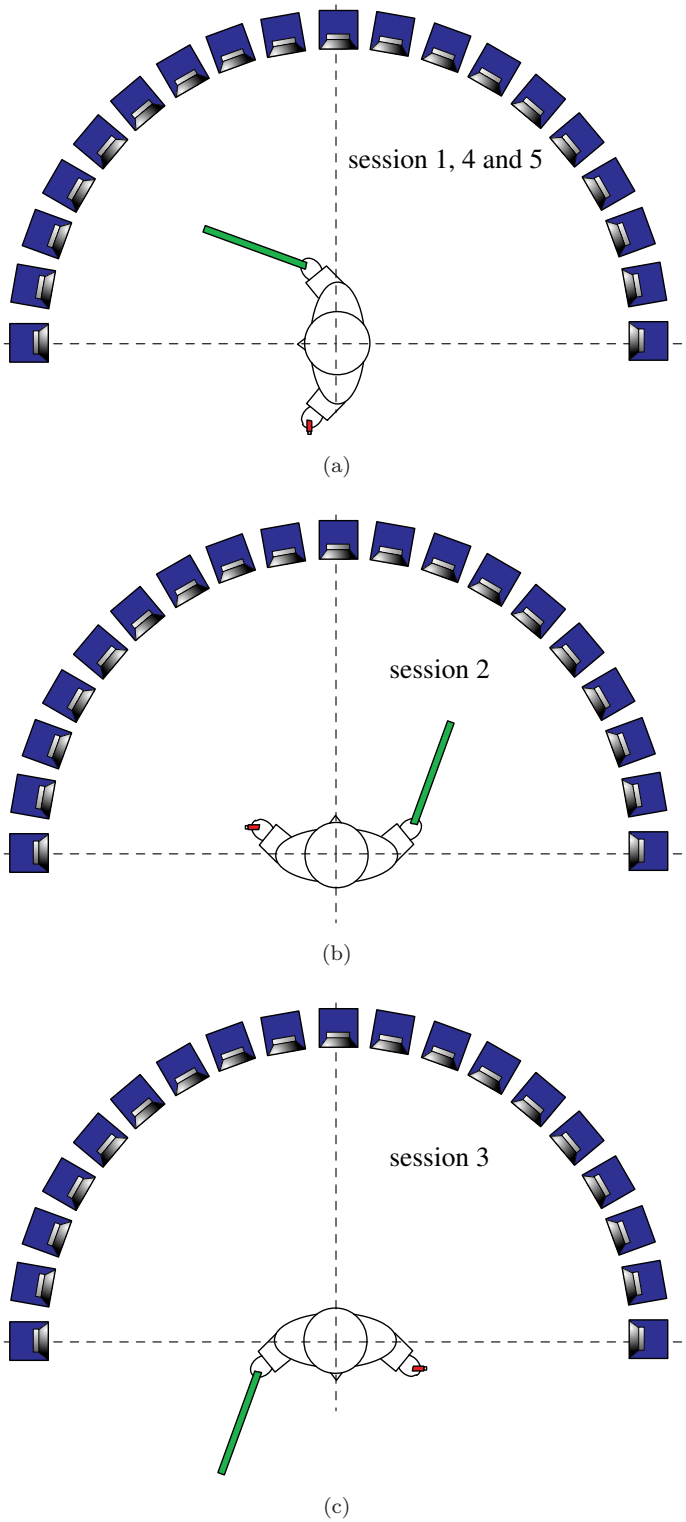


FIGURE 6.5: Orientations of the subject in accordance with the test conditions for each session (see table 6.1).

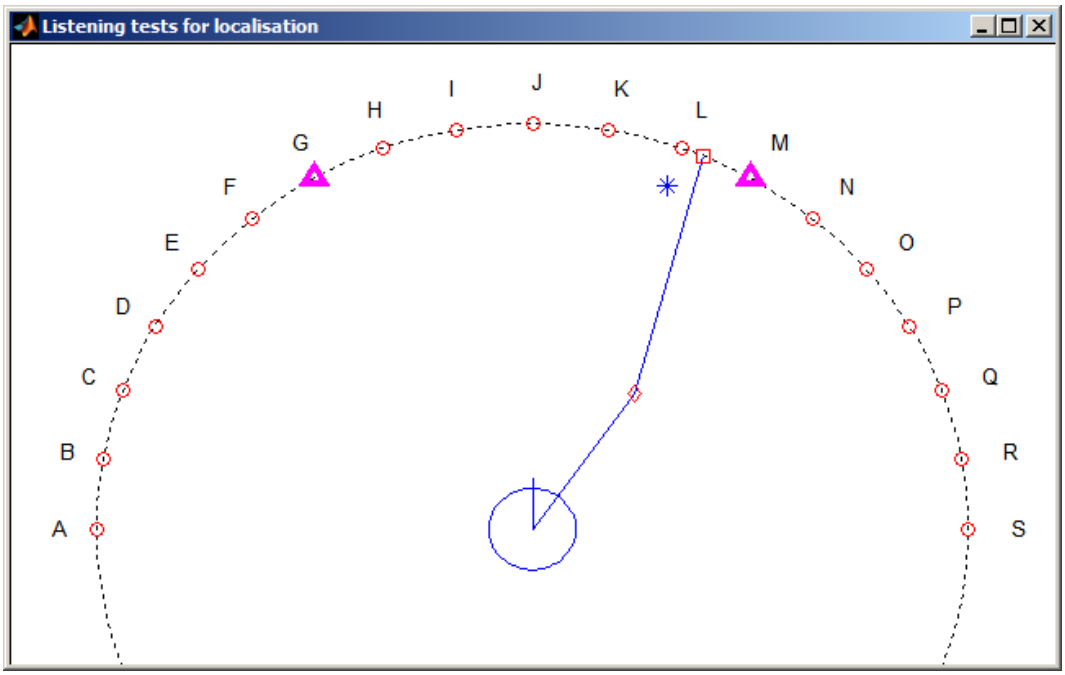


FIGURE 6.6: Matlab interface to show the test procedures. While the direction of the head is indicated by the circle at the centre of the array, and the positions of the loudspeakers are marked by small circles, the thick triangles, asterisk, diamond and the rectangle indicate the active loudspeakers, target image location, position of receiver unit (tip of the wand) and the perceived image location, respectively.

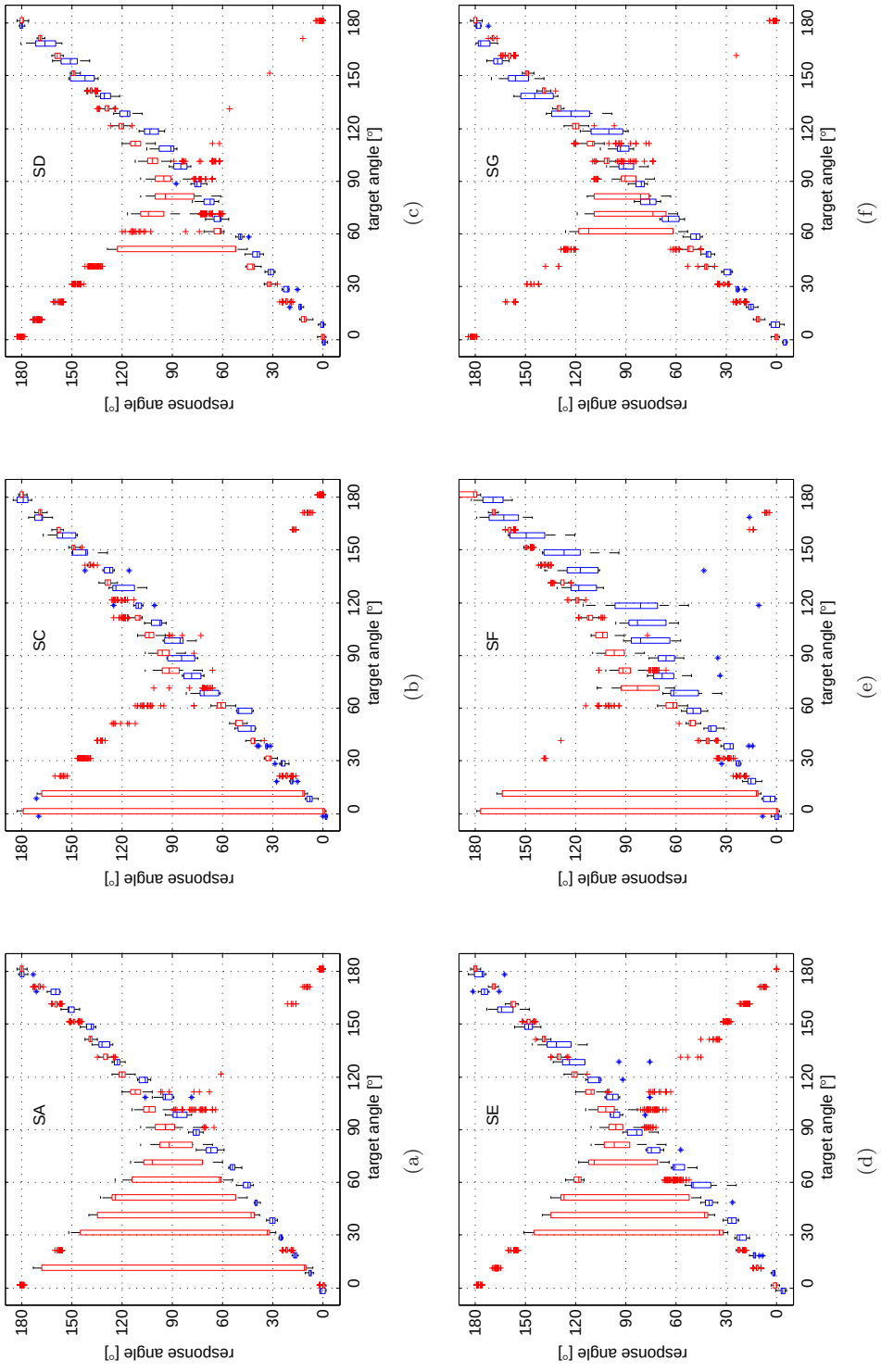
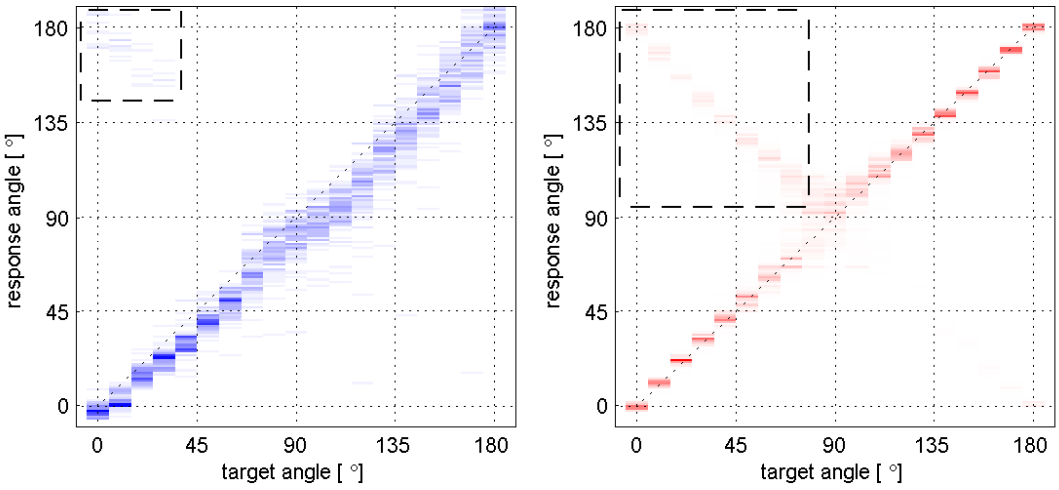
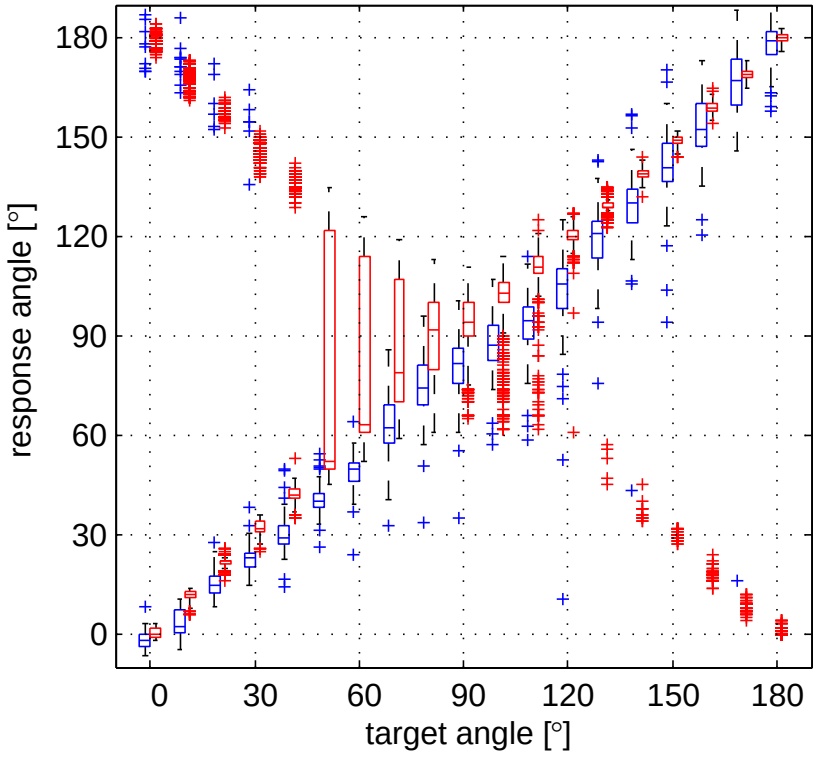


FIGURE 6.7: Box-plots comparing the subjective responses (blue) and the model predictions (red) of the real source localisation test. Only the results for the 6 subjects are presented, who participated in the HRTF measurements. The first/third quartiles are represented by the edges of the boxes while the median and the highest/lowest values are indicated by the line in the box and the two whiskers, respectively. Outliers beyond the highest/lowest values are marked by $*$ (subject) and $+$ (model). Subjects' initials are denoted in the upper mid of each plot.



(a) (b)



(c)

FIGURE 6.8: Results of the real source localisation test. (a) Listening test data from all 10 subjects. (b) Predictions of the pattern-matching models established for the 6 participants. The scale of the colour-contrast indicates the relative frequency in each bin along the vertical direction. Dashed boxes represent the mirror-imaged responses corresponding to front-back confusion. (c) Box-plots plotted for all subjective judgements (blue) and the predictions of models for the six subjects (red).

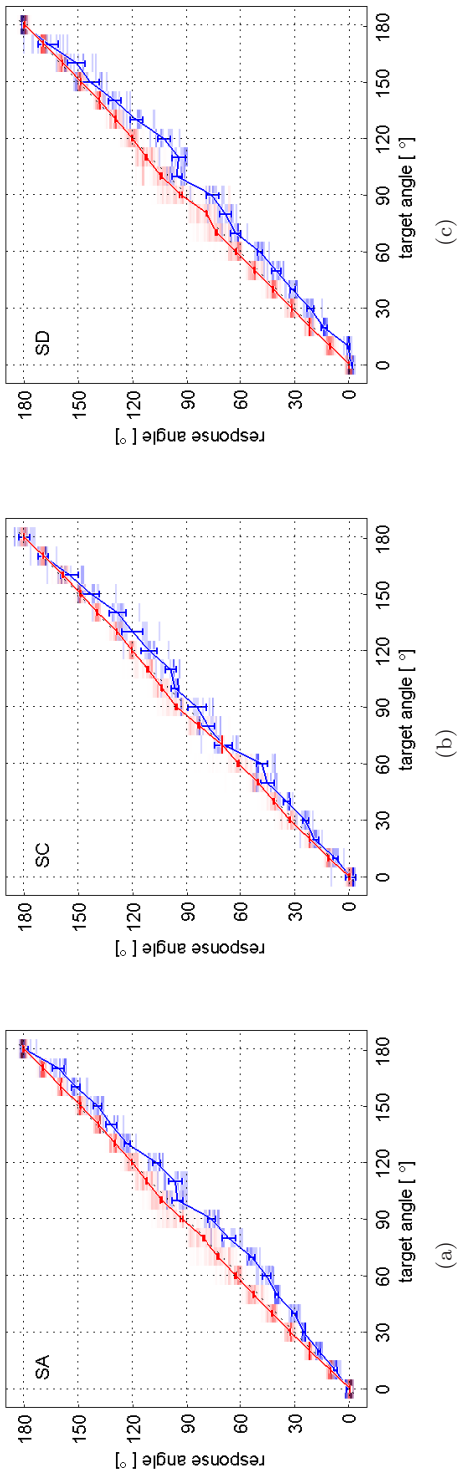


FIGURE 6.9: Errorbars indicating the mean responses and the 95% confidence intervals of the subjective judgements (blue) and the model predictions (red) for the real source localisation after front-back confusion is corrected. Colour-coded 2D histograms under the mean response plots represent relative frequency of the responses. Subjects' initials are denoted on the top left corner of each plot.

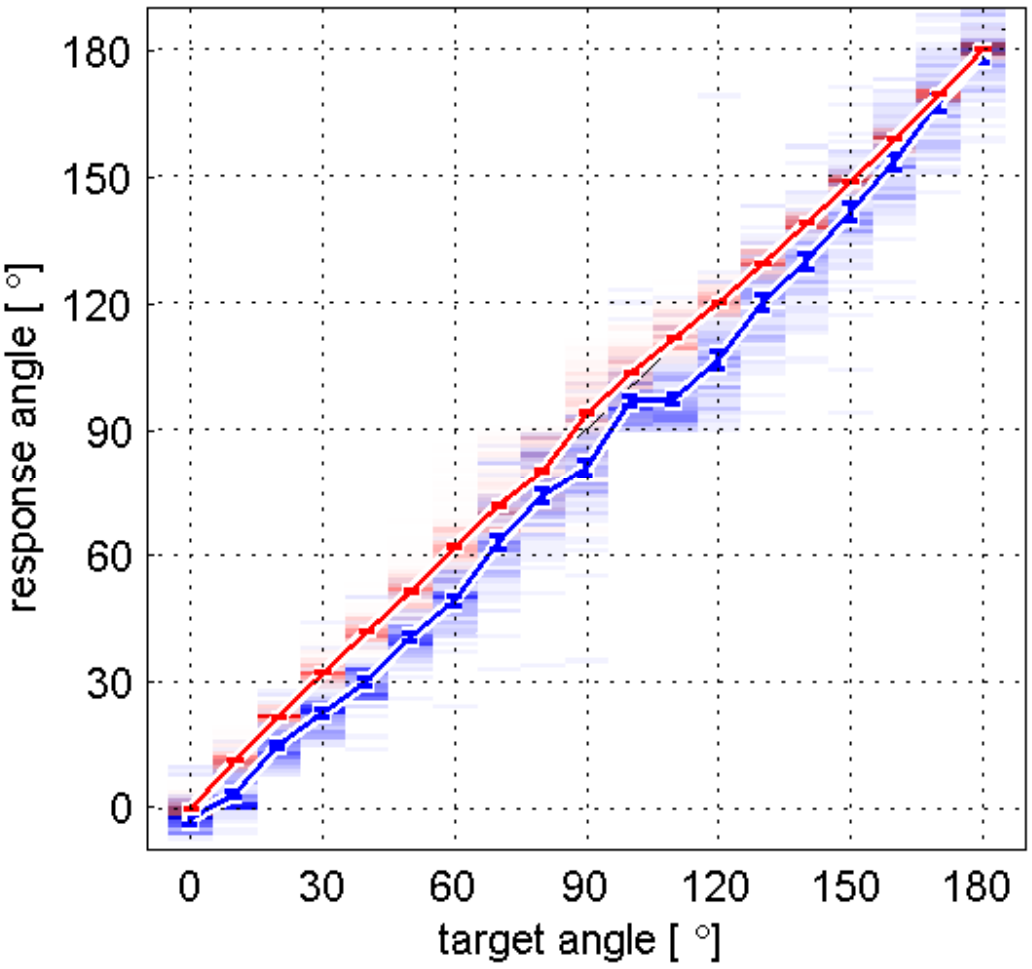
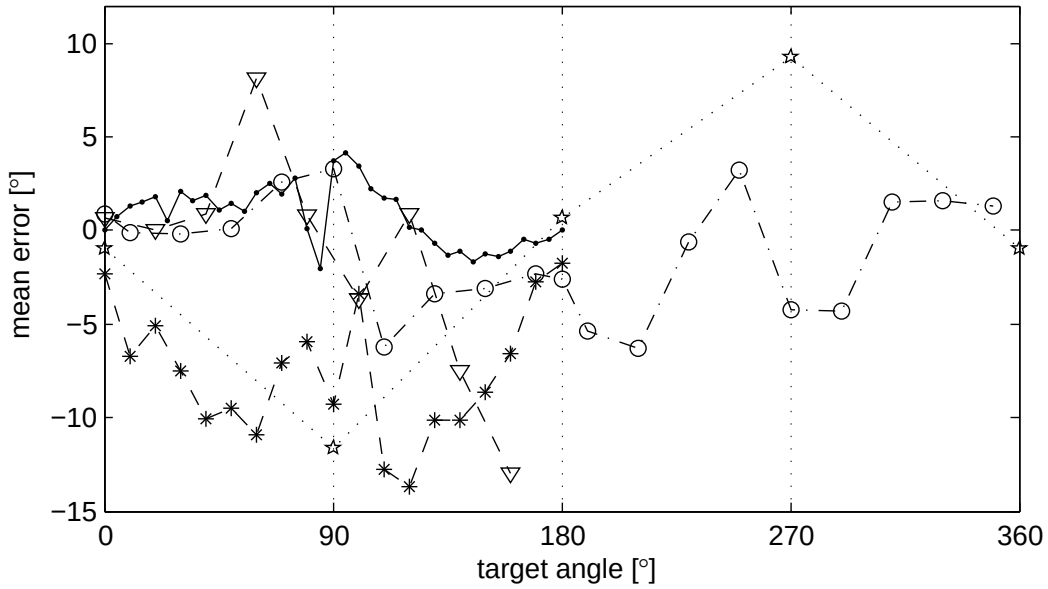
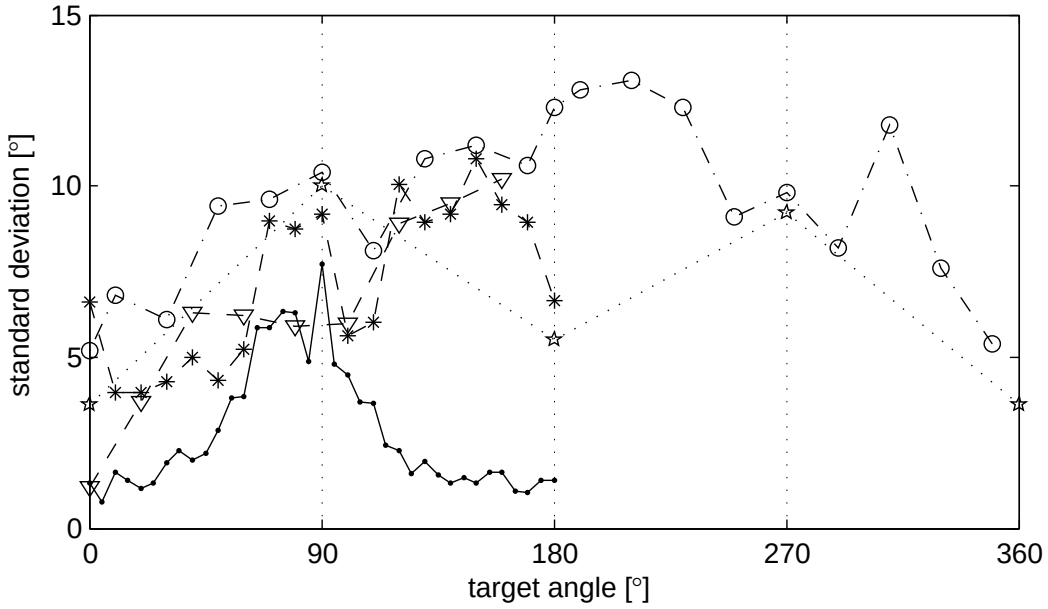


FIGURE 6.10: Errorbars indicating the mean responses and the 95% confidence intervals of the subjective judgements (blue) and the model predictions (red) for the real source localisation after the front-back confusion is corrected. Results for all subjects are presented.

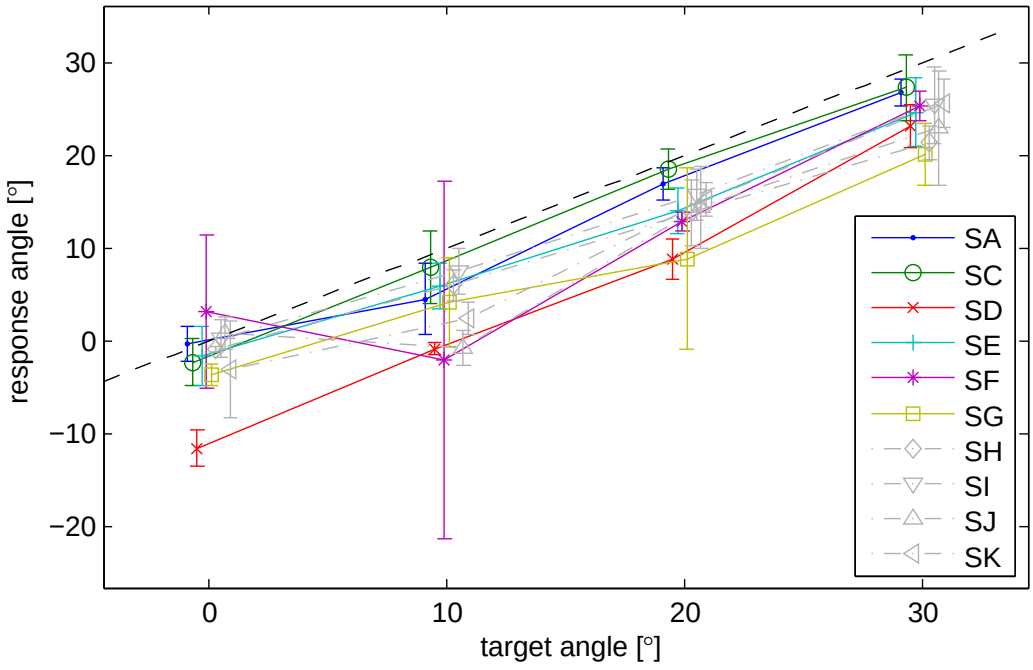


(a)

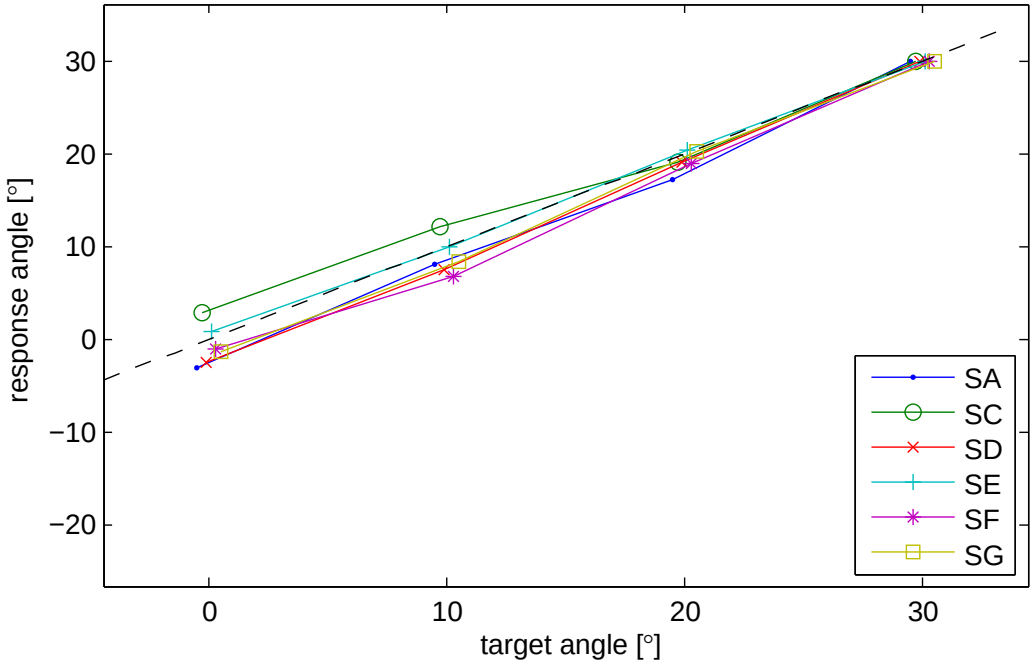


(b)

FIGURE 6.11: (a) Mean error and (b) standard deviation of the single source localisation test results are shown from Blauert [4] ($\star$), Carlile et al. [39] ($\circ$) and Makous and Middlebrooks [38] ($\triangledown$), while the results of the current tests ($\cdot$) and the model predictions ($*$) are also presented. Positive error indicates that response angle is greater than target angle for $0° \sim 360°$. Data from Makous and Middlebrooks [38] correspond to 5°-elevation.



(a)



(b)

FIGURE 6.12: Virtual source localisation ($\theta_c = 0°$, $\psi = 30°$): Errorbars indicate the mean responses and the 95% confidence intervals of (a) the listening test results and (b) the model predictions for individual subjects.

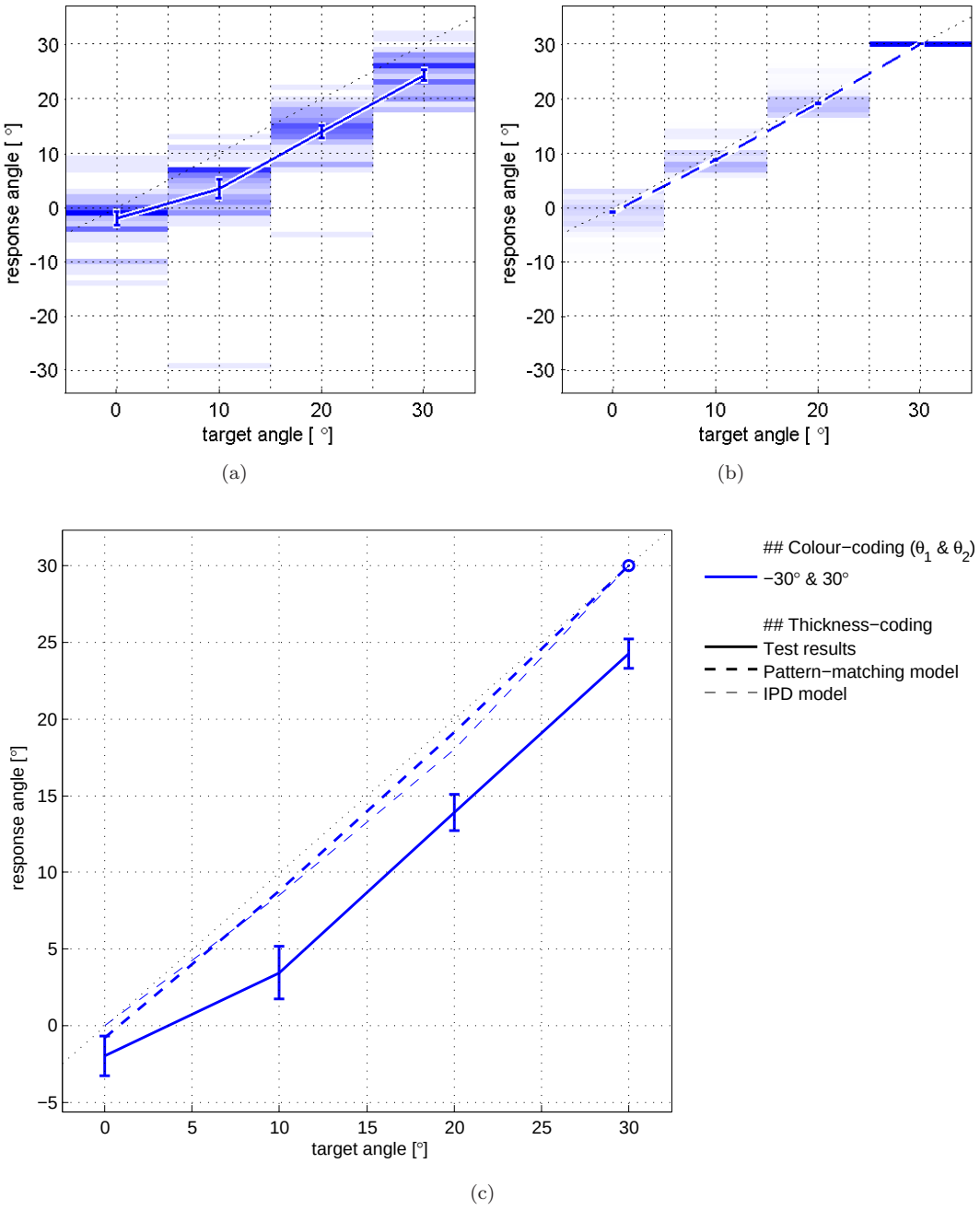


FIGURE 6.13: Virtual source localisation ($\theta_c = 0°$, $\psi = 30°$): (a) Subjective responses (10 subjects) and (b) model predictions (6 models) represented by 2D histograms with superimposed errorbars. (c) Comparison between the subjective responses (thick solid line), the predictions of the current model (thick dashed line) and the IPD model (thin dashed line).

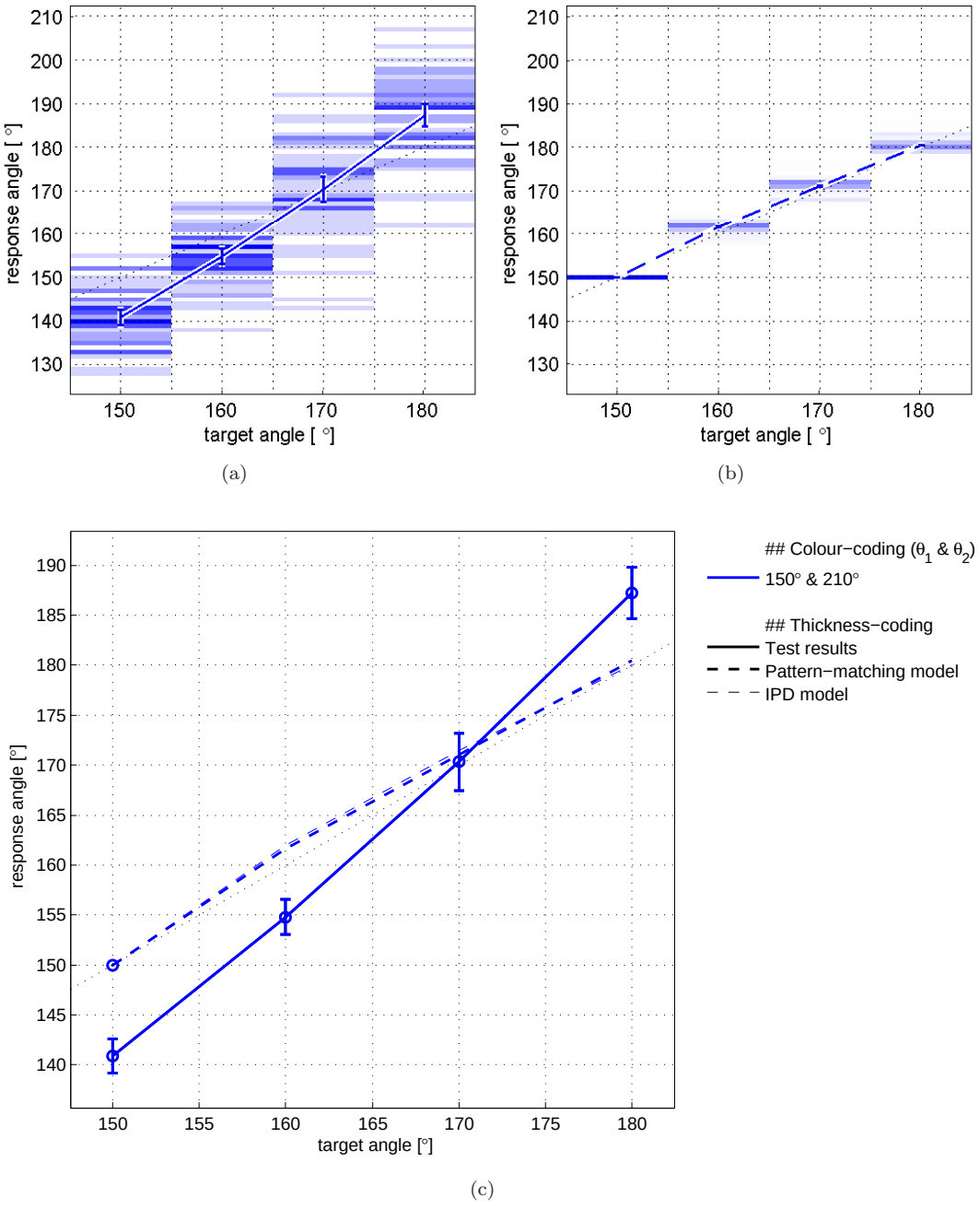


FIGURE 6.14: Virtual source localisation ($\theta_c = 180°$, $\psi = 30°$): (a) Subjective responses (10 subjects) and (b) model predictions (6 models) represented by 2D histograms with superimposed errorbars. (c) Comparison between the subjective responses (thick solid line), the predictions of the current model (thick dashed line) and the IPD model (thin dashed line).

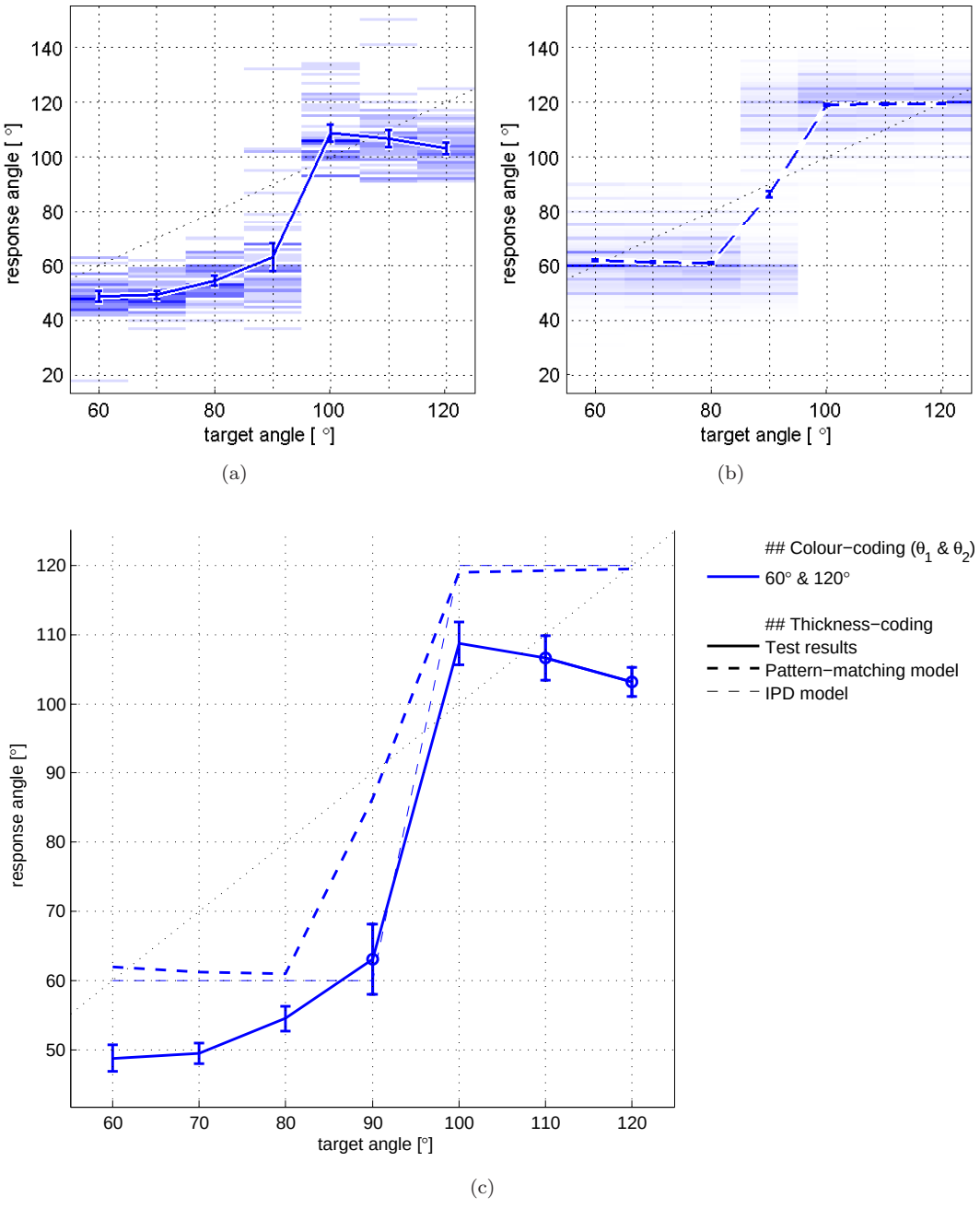


FIGURE 6.15: Virtual source localisation ($\theta_c = 90°$, $\psi = 30°$): (a) Subjective responses (10 subjects) and (b) model predictions (6 models) represented by 2D histograms with superimposed errorbars. (c) Comparison between the subjective responses (thick solid line), the predictions of the current model (thick dashed line) and the IPD model (thin dashed line).

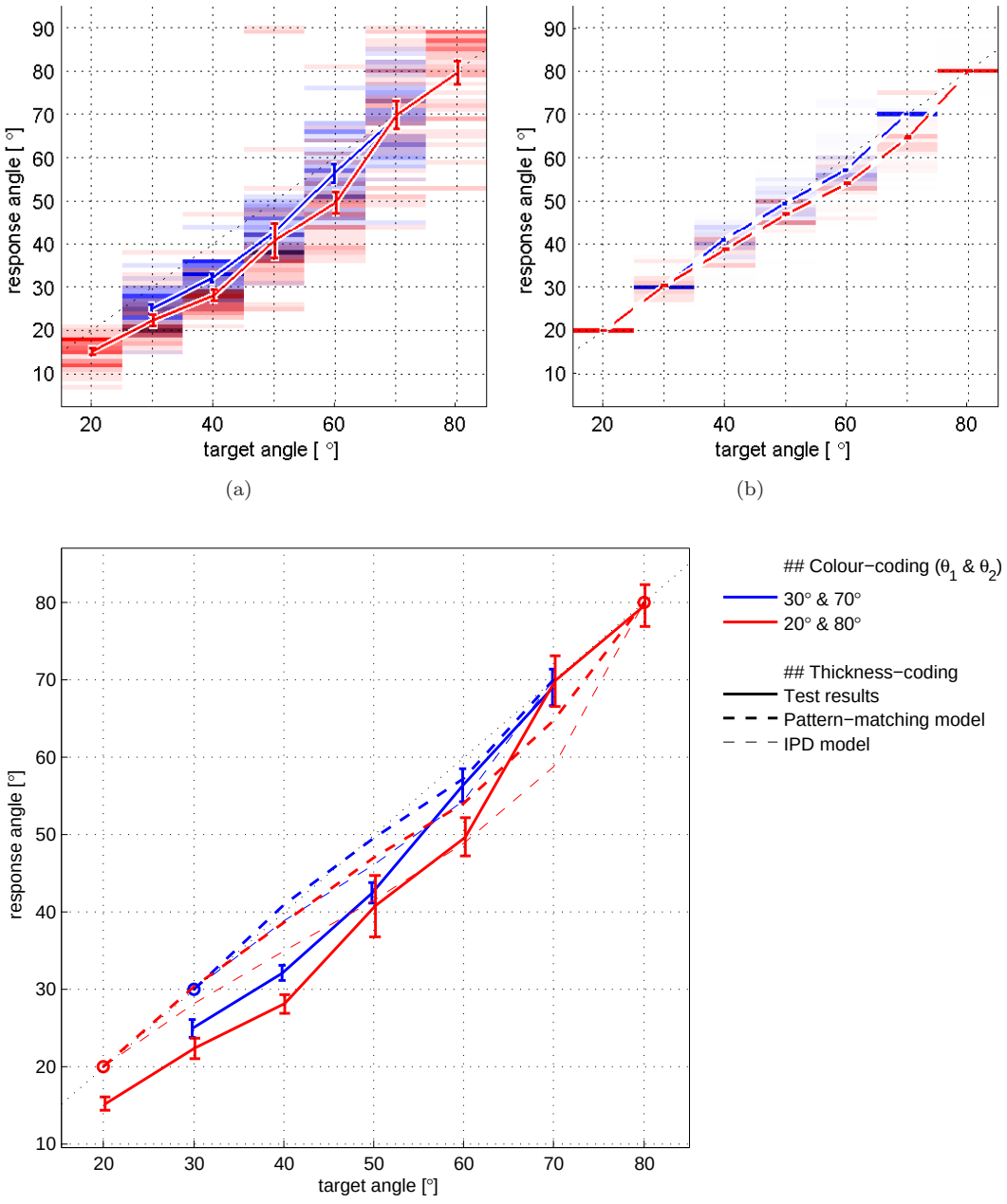


FIGURE 6.16: .
]Virtual source localisation [$\theta_c = 50°$, $\psi = 20°$ (blue) & $30°$ (red)]: (a) Subjective responses (10 subjects) and (b) model predictions (6 models) represented by 2D histograms with superimposed errorbars. (c) Comparison between the subjective responses (thick solid line), the predictions of the current model (thick dashed line) and the IPD model (thin dashed line).

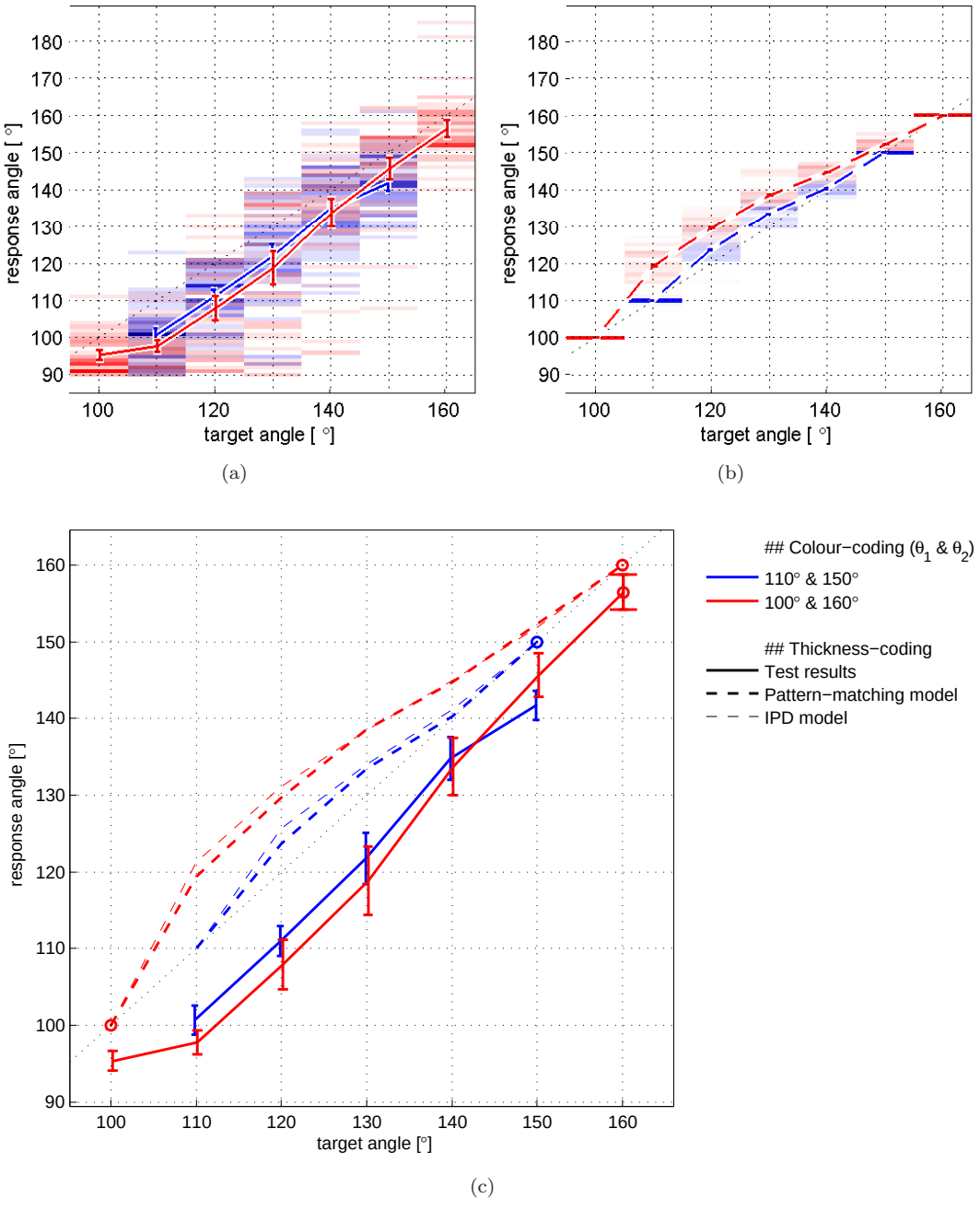


FIGURE 6.17: .
]Virtual source localisation [$\theta_c = 130°$, $\psi = 20°$ (blue) & $30°$ (red)]: (a) Subjective responses (10 subjects) and (b) model predictions (6 models) represented by 2D histograms with superimposed errorbars. (c) Comparison between the subjective responses (thick solid line), the predictions of the current model (thick dashed line) and the IPD model (thin dashed line).

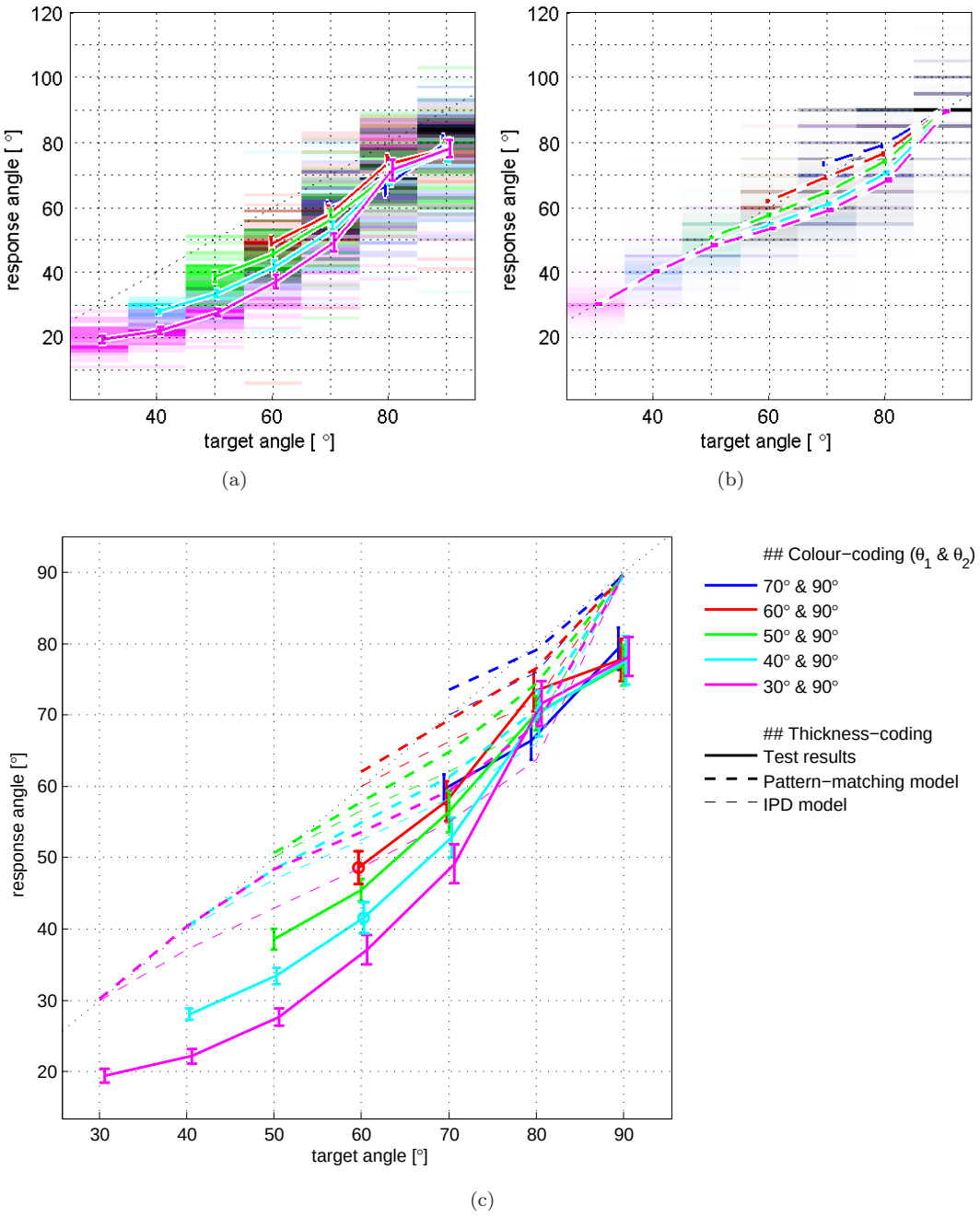


FIGURE 6.18: Virtual source localisation ($\theta_2 = 90°$, $\psi = 10°$ to $30°$ as denoted in legend): (a) Subjective responses (10 subjects) and (b) model predictions (6 models) represented by 2D histograms with superimposed errorbars. (c) Comparison between the subjective responses (thick solid line), the predictions of the current model (thick dashed line) and the IPD model (thin dashed line).

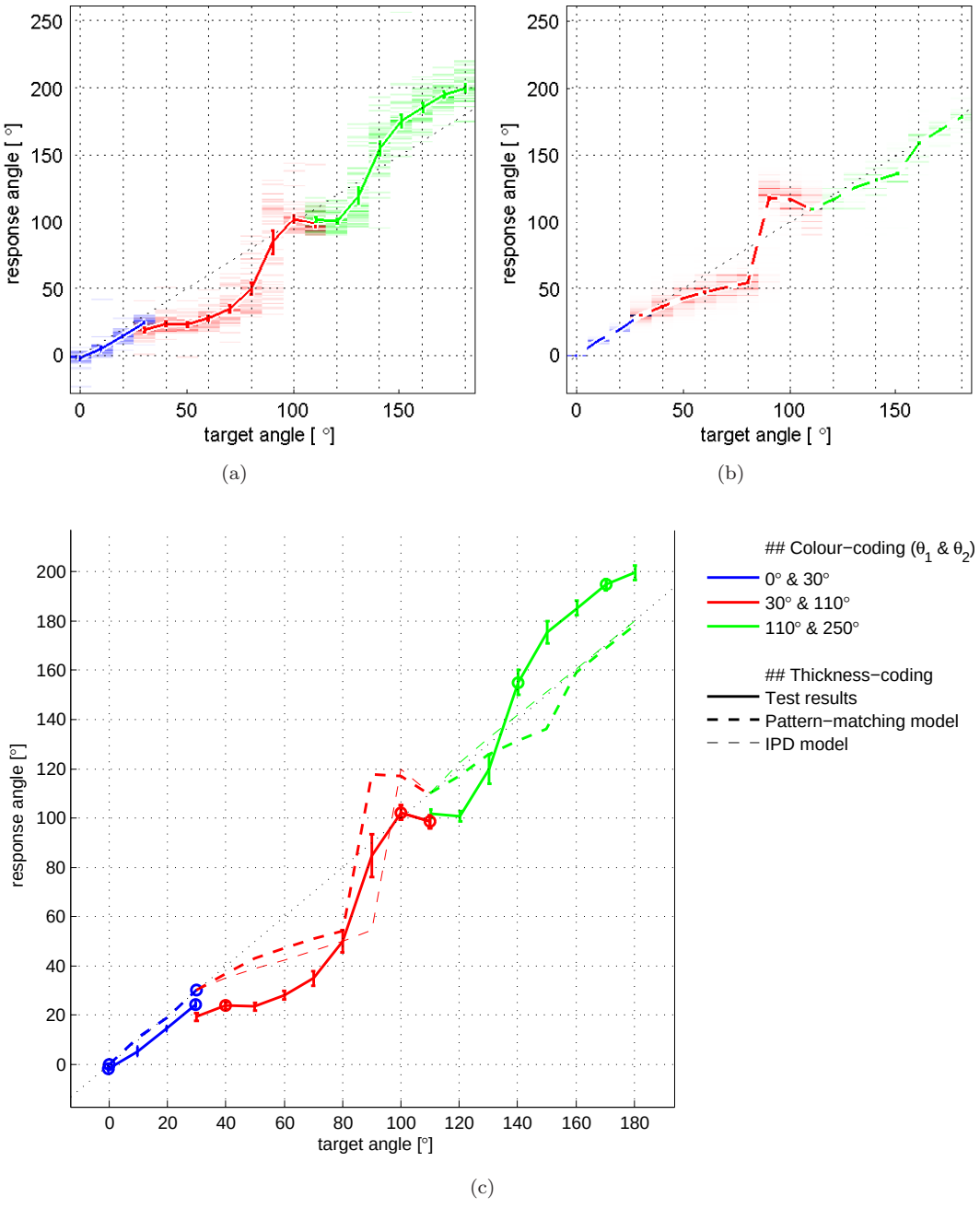


FIGURE 6.19: Virtual source localisation (5.1 channel configuration at 0° (C), 30° (R), 110° (RS) and 250° (LS); blue for C-R, red for R-RS and green for RS-LS): (a) Subjective responses (10 subjects) and (b) model predictions (6 models) represented by 2D histograms with superimposed errorbars. (c) Comparison between the subjective responses (thick solid line), the predictions of the current model (thick dashed line) and the IPD model (thin dashed line).

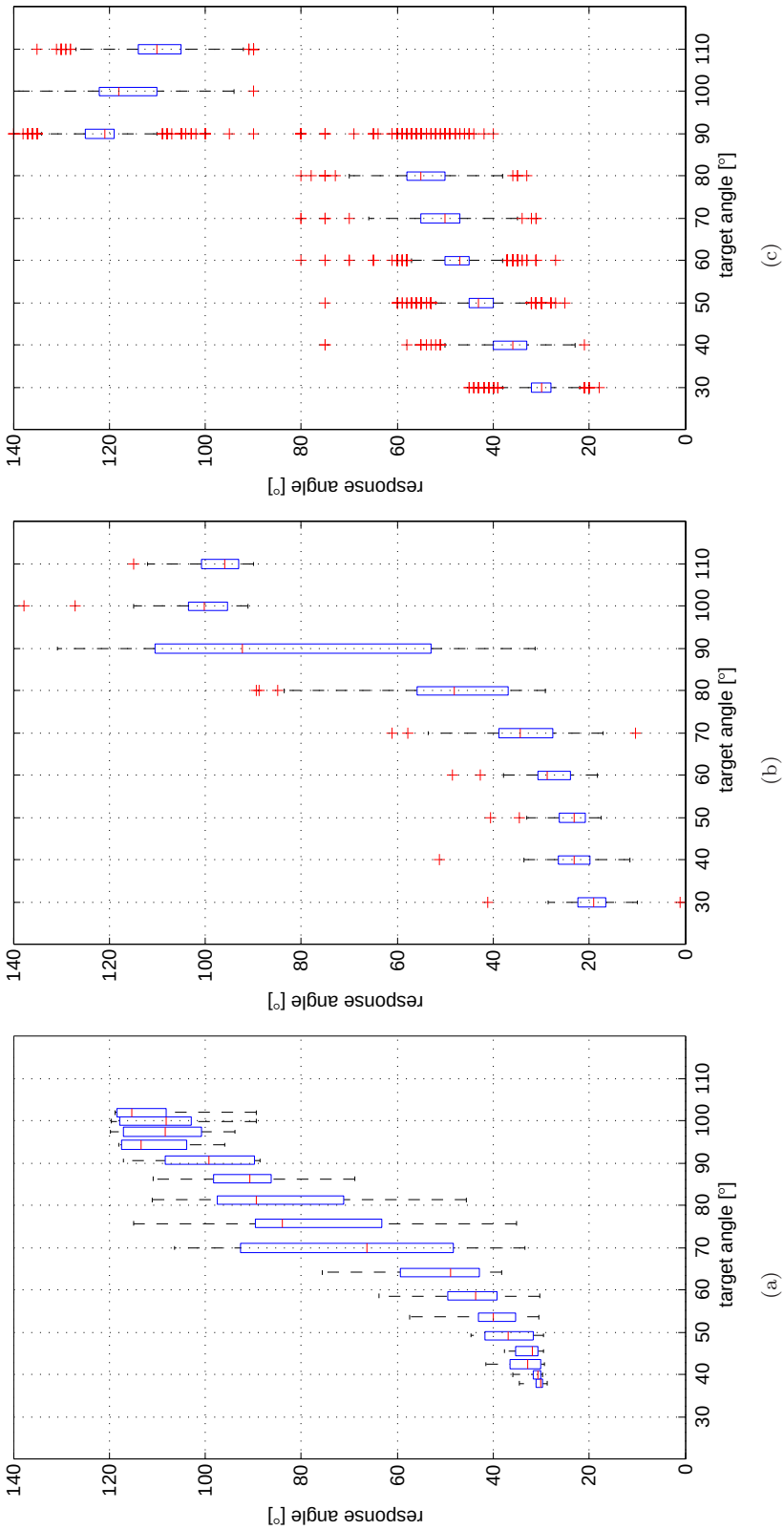
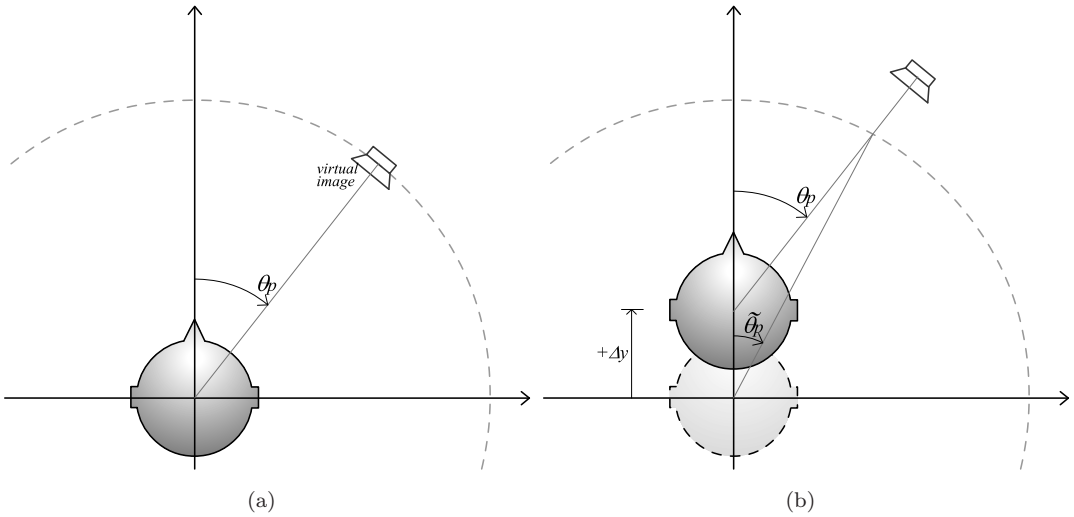
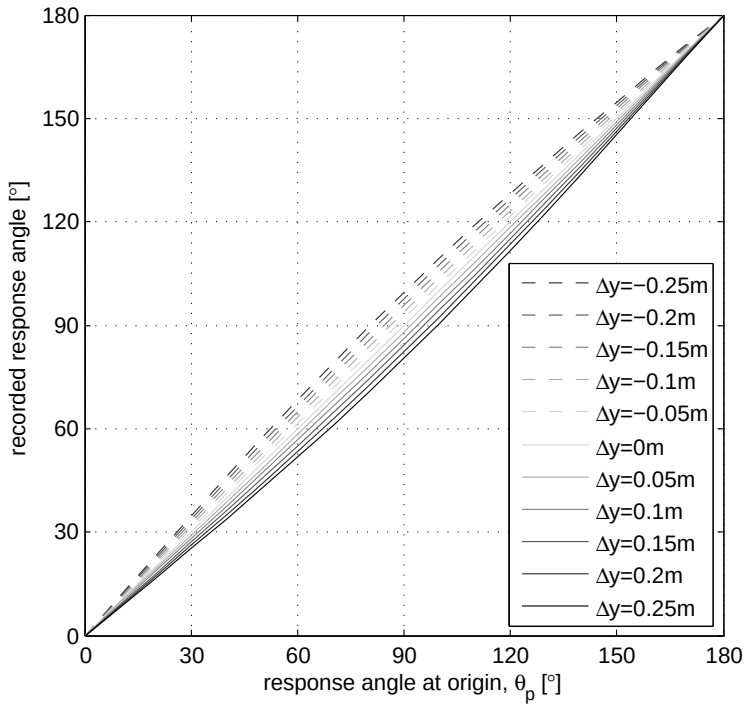


FIGURE 6.20: Box-plots indicating the subjective judgements of image positions for R-RS loudspeakers of 5.1 channel configuration. (a) Data by Martin [84] with the horizontal axis converted from inter-channel amplitude difference to the target location obtained in accordance with Eq. (6.2). (b) Results of the current listening tests. (c) Predictions of the pattern-matching model.

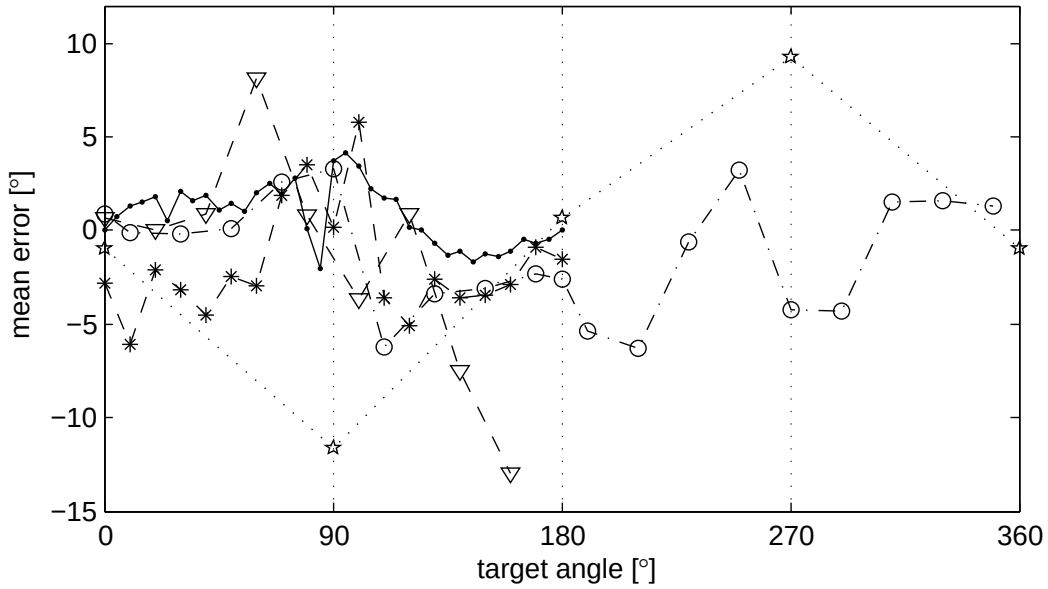


(a) (b)

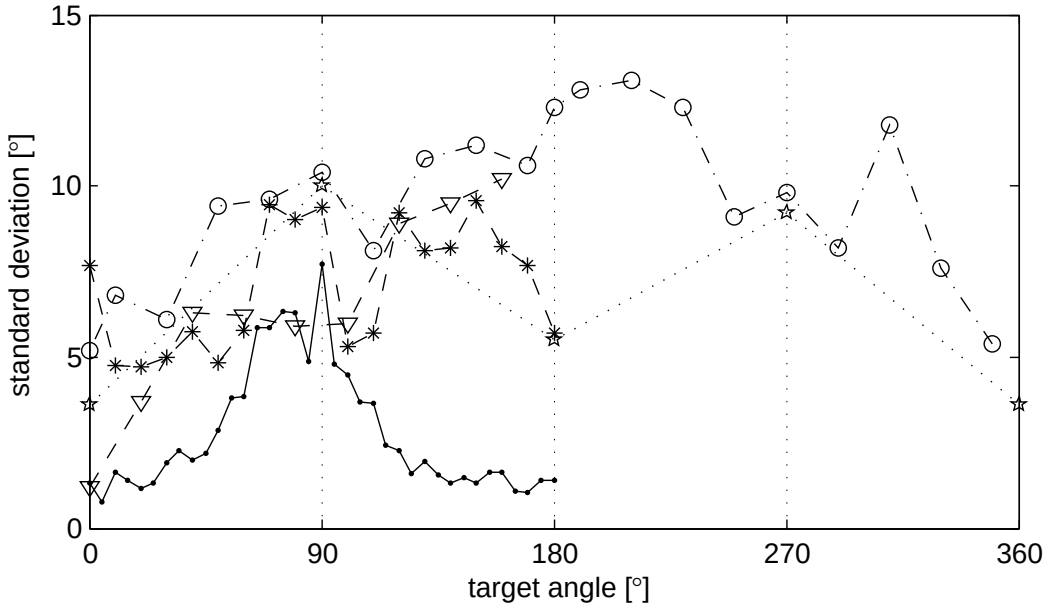


(c)

FIGURE 6.21: (a) Perceived image position, θ_p at the centre position. (b) Recorded image position, $\widetilde{\theta_p}$ inconsistent with the perceive position, θ_p due to the forward/backward displacement of the subject. (c) Mapping function relating the recorded location to the perceived image position for various displacements, Δy.



(a)



(b)

FIGURE 6.22: Assuming $\Delta y = 0.2\,m$ in Fig. 6.21, subjective judgements have been compensated to redraw Fig. 6.11 for (a) the mean responses and (b) the standard deviations in the real source localisation tests. [Blauert [4] ($\star$), Carlile et al. [39] ($\circ$), Makous and Middlebrooks [38] ($\triangledown$), current tests ($\cdot$) and the model predictions ($*$)]

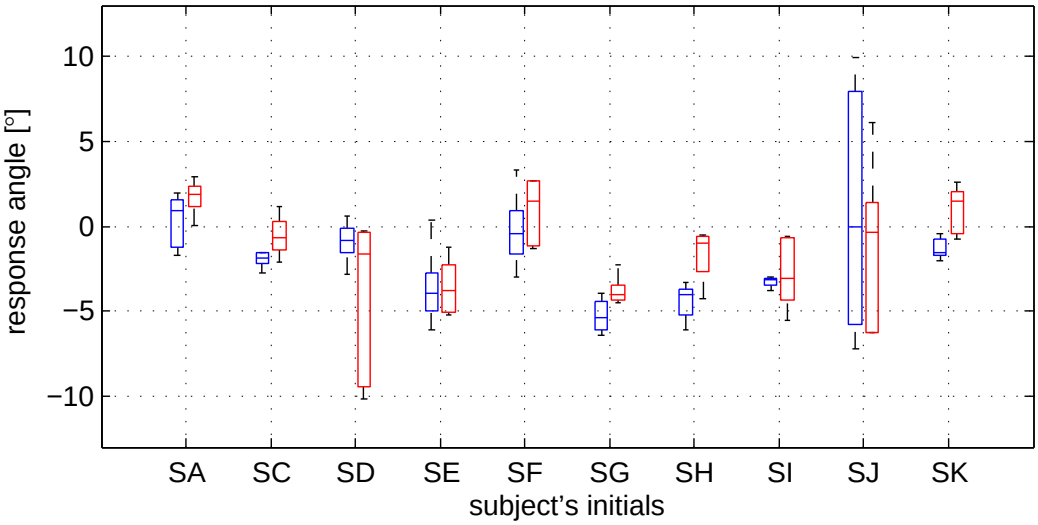


FIGURE 6.23: Box-plots for the comparison between the results of the single loudspeaker localisation tests using loudspeakers at the far-side (blue) and in the middle (red). Sound source is at 0°

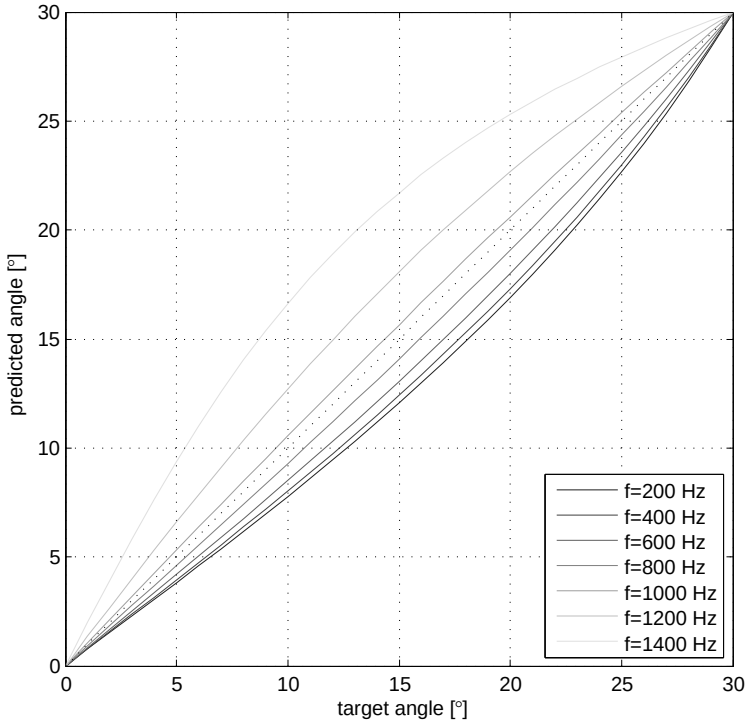
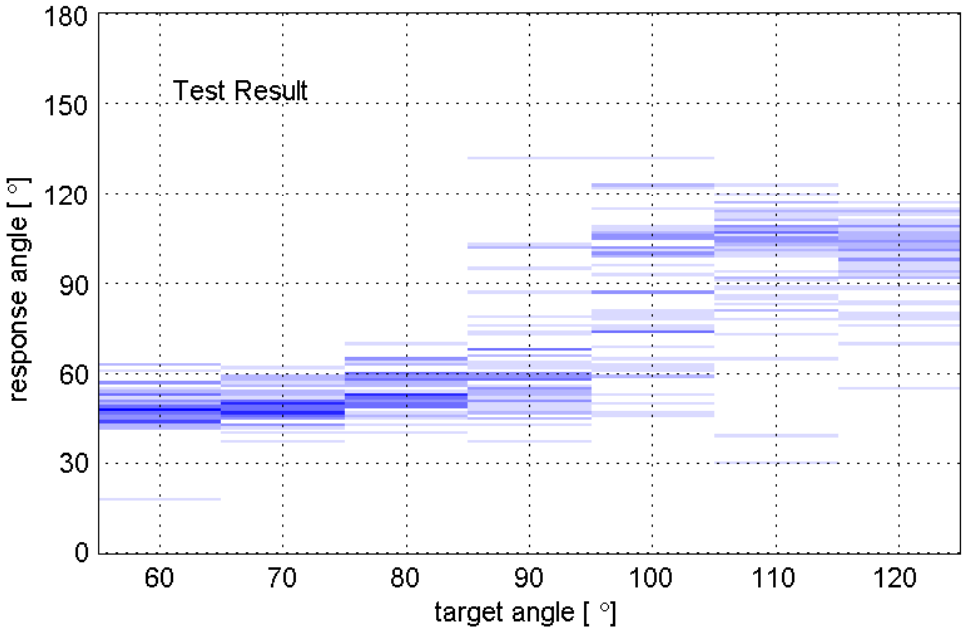
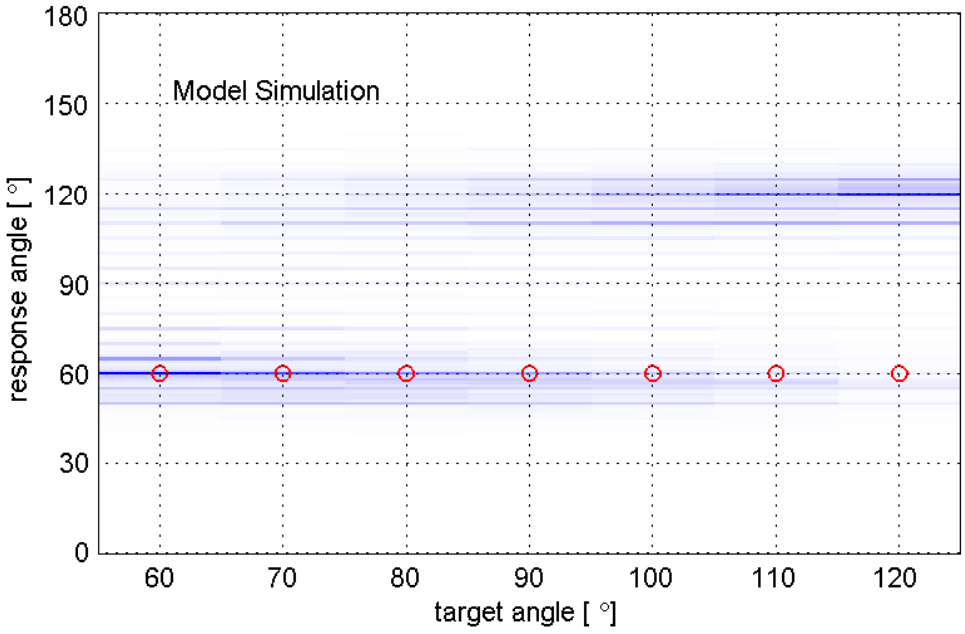


FIGURE 6.24: Predictions of the IPD model at various frequencies for the conventional stereophonic system based on the constant-power panning method.

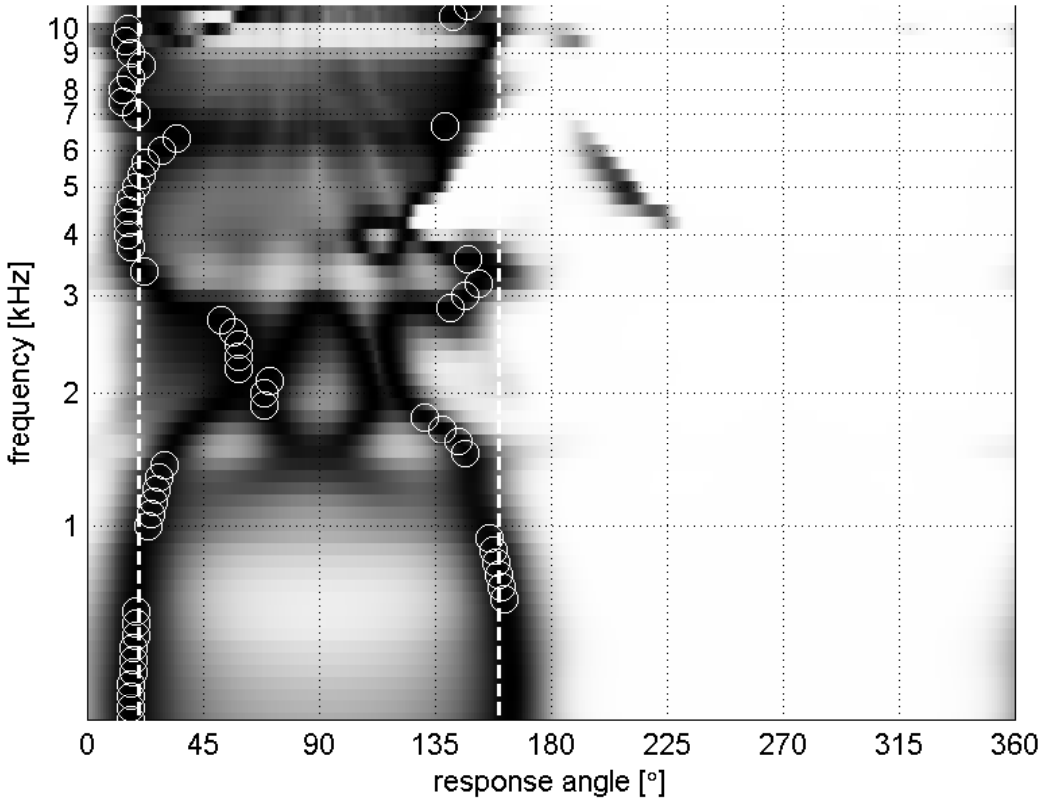


(a)

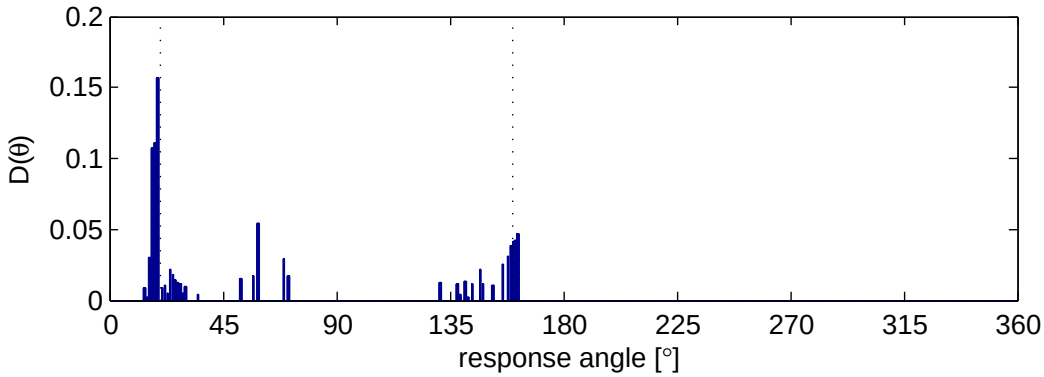


(b)

FIGURE 6.25: Raw results before the front-back correction for the front-back symmetric loudspeaker configuration presented in Fig. 6.15. (a) Subjective responses and (b) predictions of the pattern-matching model with red circles indicating the predictions made by the IPD model.



(a)



(b)

FIGURE 6.26: Central processes of the pattern-matching model presented for the virtual image at 20° created by the conventional stereophony system. (a) Cross-correlation between the EI-template and the target EI-patterns where white circles indicate the decisions made in each auditory frequency band. Dashed white lines represent the target and its mirror-imaged positions. (b) Probability function of source location obtained by weighting the local decisions shown in panel (a).

Chapter 7

Conclusion

In the current study, two binaural hearing models, the characteristic-curve (CC) model and the pattern-matching (PM) model have been suggested for the prediction of the subjective perception of inside- and outside-head acoustic image locations in the horizontal plane. The two models are similar in that, for the central processes, the free-field cues are considered to be the memory of human spatial hearing to which the internal representations of the target stimulus are compared. Such a comparison produces the estimate of the acoustic image position in azimuth angle at a single frequency or in a frequency band. In particular, the characteristic curve model takes the interaural time difference and the interaural level difference as the intermediate input to the central processor. The nearest-neighbour to those localisation cues has been found on the characteristic curve that is the collection of all possible combinations of ITD and ILD arising in the free-field listening environment. In addition, the pattern-matching model considered a whole curved surface in (τ, α, EI'') coordinate space as an internal representation of the auditory scene in a single auditory frequency band, and compared it to the EI-template in terms of cross-correlation. The EI-template is, again, the collection of all possible EI-cell activity patterns for sound sources in the horizontal plane, the resolution of which is limited by the resolution of the HRTF.

On the other hand, there are many distinctions between the two models. Firstly, the pattern-matching model includes all three processes of spatial hearing, peripheral, binaural and central processes, whereas the characteristic curve model is only focused on the central processes, assuming that ITD and ILD are already given by the lower level processes. In terms of the predictive scope, the CC model operates only at a single frequency due to the restriction in the establishment of the characteristic curves across frequency. However, the EI-patterns contain both waveform and envelope ITD information, and, with an experimental frequency weighting scheme, decisions made in each

auditory frequency band can be combined to produce a probability function of target location, from which a single estimate can be obtained even for a broadband sound source. Finally, it is noteworthy that internal processing errors have been taken into account in the CC model by two independent random noise component that are added to the target ITD and ILD, while a noise mask has been applied to the EI-pattern in the PM model.

The predictive scope of the two models have been further investigated in two separate listening tests, where, in order to emphasise the strong point of each model, the subjective judgements of lateralities of low-frequency pure tone signals have been obtained and compared to the predictions of the CC model. In addition, the PM model has been applied to the analysis of the localisation listening tests where broadband real and virtual acoustic images have been presented to the subjects. In order to establish the computational models for some of the participants, individual HRTFs have been measured in advance to provide the memory of the past localisation operation, that is the characteristic curve and the EI-template for the CC model and the PM model, respectively.

Unfortunately, it was difficult to establish strictly quantitative and subject-specific links between the listening test results and the predictions of the associated hearing models mainly due to 1) the large variance near the critical ITDs in case of the lateralisation tests and 2) the significant underestimation of the target positions in the localisation tests, the second of which was especially prominent compared to similar test results reported in the literature. Nevertheless, qualitative agreement between the subjective responses and the model predictions was quite remarkable, where features in the lateralisation of pure tones and the localisation of broadband real and virtual sound sources were described well by the CC and the PM models, respectively. For example, in the lateralisation study, the critical ITD values given in the listening tests have been found to be consistent with those predicted by the CC model. Also, the empirical laterality curves were very similar to those simulated in terms of the vertical distance between nearby curves representing different target ILDs, particularly when the ITD is small in the absolute sense compared to the critical ITDs. In the localisation of virtual images created by various stereophonic arrangements, the relationship between the loudspeaker angular aperture and the extent of position underestimation has been described well by the PM model. In addition, the subjective evaluation of the conventional 5.1 channel surround system has been successfully predicted by the PM model where it has been suggested that the results in the current study are much more extensive than some previous studies in the literature.

Statistical analysis has shown that the agreement between subjective responses in the laterality tests and predictions of the subject's own CC model is reasonable, where the

success rate has been approximately between 30% and 60%, and could be up to 70%, depending on the subject and the target ILDs. In the localisation tests, however, the discrepancy between the subjective responses and the individual model predictions was quite significant. It is suggested that the prominent contrast between the two cases can be probably attributed to the nature of each listening test. As a matter of fact, the lateralities of the dichotic pure tones have been investigated by a matching task where subjects found an 'acoustic' pointer that was most consistent with the target image location. However, the outside-head location of the acoustic image in the localisation tests has been examined by subjects who used a 'visible' pointer to report the perceived image position. In other words, in the former test, auditory perception could be represented by a reference auditory stimulus, where the matching process involves, arguably, only the hearing process. However, in the latter case, auditory perception had to be represented visually, requiring other sensory processes and physical operation, thus inevitably adding to the uncertainties from each process, although it is beyond the scope of this study to discuss the relative influence of each factor.

It is not intended in the current study to suggest the presence of specific neural structures or processing mechanisms, for example, the nearest-neighbour finding or the pattern-matching procedures, which are still active topics of research in auditory neurosciences and physiology. Indeed, the models are purely based on assumptions, mainly the essential role of the past localisation memory established by the free-field ITD and the ILD information, which are nevertheless very common ideas readily accepted by the hearing research community. As a matter of fact, despite the importance of 'naturally combined' binaural cues, there has been little effort to obtain the estimate of auditory image location simultaneously from both interaural disparities, and the central processes of the CC and the PM models described in the current study are considered to be unique. Therefore, considering the relatively successful predictions of the important features of human spatial hearing, and the simplicity and the flexibility of the two models in handling both inside- and outside-head localisation problems in a single framework, it is regarded as worth investigating the predictive scope of the current models in an extended range of listening environments. In addition, it is expected that the current models may be applied for the design and evaluation of spatial audio systems, possibly predicting how individual listeners would appreciate the reproduced sound field.

Appendix A

Spatial interpolation of the HRTF

Recent studies of the spatial interpolation of head-related transfer functions (HRTF) databases include those by Evans et al. [50], Langendijk and Bronkhorst [87] and Takeuchi [88]. Whilst the latter two studies refined the spatial resolution of HRTFs in the frequency domain by linearly interpolating the magnitude and the phase responses separately, Evans et al. [50] performed a spherical harmonic transformation of the measured HRTFs carefully sampled over an entire spherical surface, and then, recreated HRTFs at any angular location by an inverse transform. They applied this scheme both in the time and frequency domains and achieved reasonable recreation and interpolation performances in terms of the resulting mean square errors.

From the above listed studies, it is reasonable to consider three independent parameters that establish a certain HRTF interpolation scheme: (1) Whether it is computed in the time- or frequency-domain (**domain**); (2) Whether the actual interpolation is linear or uses other schemes such as the spherical harmonic transformation (**algorithm**); (3) Whether the onset times of the raw head-related impulse responses (HRIR; the time domain representation of HRTF) are first equalised in the time-domain (**onset-equalisation**). Normally, the application of onset-equalisation has to be determined in the first place, followed by the decisions regarding the domain and the algorithm.

Abbreviation	Algorithm	Domain	Onset-equalisation
LT	Linear	Time	No
LF	Linear	Freq.	No
FT	FFT (trigonometric)	Time	No
FF	FFT (trigonometric)	Freq.	No
LTeq	Linear	Time	Yes
LFeq	Linear	Freq.	Yes
FTeq	FFT (trigonometric)	Time	Yes
FFeq	FFT (trigonometric)	Freq.	Yes

TABLE A.1: Table of abbreviation for the interpolation schemes characterised by the algorithm, domain of computation and the equalisation of the onset times.

Combining the three parameters can result in various interpolation schemes, particularly with a variety of algorithms. However, in this section, only two algorithms, linear and trigonometric interpolations will be investigated, where the trigonometric interpolation (or the FFT interpolation) can be regarded as the 1-dimensional simplification of the spherical harmonic transformation used in Evans et al. [50]. Therefore, a total of 8 schemes will be compared for HRTF interpolation in the horizontal plane, which will be abbreviated as shown in table A.1.

Each parameter is implemented in the following manner.

- **Onset-equalisation**: Onset time can be obtained from the peak of the cross-correlation function between a pair of nearby HRIRs. For example, for the HRIRs shown in Fig. A.1(a), panel (b) in the same figure shows the relative onset-time acquired, which, however, contains a step-wise increase/decrease for some azimuth angles, limited by the sampling frequency. Therefore, it is necessary to oversample the HRIRs in advance, 10 times the original sampling frequency in the current study, so that the onset time calculation may give a smoother curve as shown in Fig. A.1(c) (see the increase in the sample numbers due to the oversampling).

  Then, HRIRs are shifted according to the relative onset times for the following interpolation process [see Fig. A.1(d)]. After the actual interpolation across azimuth angle is completed using whichever algorithm in time- or frequency-domain, the onset times are restored, and the HRIRs are downsampled to the original sampling frequency, where the onset times for the created part of the HRIRs are linearly approximated from the curve shown in Fig. A.1(c).

- **Domain**: Time-domain operation is relatively straightforward, where samples corresponding to each time instant are considered for the interpolation across azimuth angle. On the other hand, interpolation in the frequency domain is implemented separately for the magnitude and the phase. Similar to the scheme suggested by Langendijk and Bronkhorst [87], the magnitude, M and the phase, Φ are not handled in their original forms, but first converted to M' in dB scale and Φ' in the exponential form, respectively.

$$M' = 20\log_{10}M \tag{A.1}$$

$$\Phi' = \exp(i\Phi) \tag{A.2}$$

  After the actual interpolation in the azimuthal direction, M' and Φ' are combined to give HRTFs in frequency domain, which are then converted back to the time

domain by inverse FFT.

$$\mathrm{HRTF}_{\mathrm{interpolated}} = 10^{\frac{M'}{20}} \times \Phi' \tag{A.3}$$

- **Algorithm**: Linear and trigonometric interpolations have been implemented in Matlab 7.0 by built-in functions, *interp1* and *interpft*, respectively.

Those 8 schemes listed in table A.1 have been applied to the 6 HRTFs (distal-region) measured in chapter 3. Given the original spatial resolution of 5°, HRIRs at every 10° from 0° to 350° have been regarded as raw data, which were interpolated to recreate 5°-resolution HRIRs, and the percentage root mean square error (PRMSE) in the following form has been computed for each scheme.

$$\zeta(\theta) = \sqrt{\frac{1}{N}\sum_{k}^{N}\left(\frac{|H_\theta[k]| - |H'_\theta[k]|}{|H_\theta[k]|}\right)^2} \times 100(\%) \tag{A.4}$$

where $H_\theta[k]$ and $H'_\theta[k]$ indicate the original and the recreated HRTFs at θ, respectively.

Fig. A.2(a) shows the PRMSE obtained for each interpolation scheme applied to the left channel data, averaged for 6 subjects. Since the red and blue colour-coding indicates whether or not the onset time has been equalised, it is apparent that the recreation performance is generally better with onset-equalisation when the source is ipsilateral to the receiving side. However, in the narrow region of the contralateral side, say between 80° and 100°, the linear interpolation schemes without the onset-equalisation (LT and LF) show the less errors compared to others. This dependence of interpolation errors on the source location can be perhaps understood in relation to the performance of onset-equalisation. As shown in Fig. A.1(a), there is an apparent discontinuity found around 90° in the raw HRIRs, which results in the 'jump' in the equalised HRIRs in Fig. A.1(d). In other words, the realignment of HRIRs is inevitably incomplete in the contralateral side, which does not help to enhance the interpolation but to give greater errors.

The overall performance of the 8 interpolation schemes can be examined in Fig. A.2(b), where PRMSEs have been averaged across source location. From this figure, it is observed that (1) the influence of domain (time or frequency) is unclear, (2) linear interpolation is superior to the trigonometric method and (3) onset-equalisation generally enhances the interpolation performance. Relatively poor results produced by the trigonometric interpolation is closely related to the fact that 10°-resolution raw HRIRs had also to be recreated through the inverse Fourier transformation.

In order to further investigate the interpolation schemes, 1°-resolution distal-region HRTFs have been measured with the KEMAR. Ear-simulators (GRAS RA0045 with type 26AC preamplifier) have been placed in the ear-drum depth for signal recording, while other measurement arrangement has been identical to that described in section 3.2.

Fig. A.3 shows the PRMSEs averaged across source location and left-right channel for the recreation of 1° HRIRS from the HRIRs selected at every 5° from the original recordings. While each interpolation scheme is colour-coded as shown in the legend, PRMSEs corresponding to ipsilateral and contralateral source positions are shown separately along with the overall averages. (For the left ear, for example, 0° to 180° is contralateral, while 180° to 360° is ipsilateral.)

The contribution of each interpolation parameter is clearly depicted in Fig. A.3, where the following are observed.

- Interpolation performance is better on the ipsilateral side.
- The influence of computation domain - whether time or frequency - varies depending on the algorithm of interpolation. For example, time-domain performance is slightly better with linear interpolation, but frequency-domain interpolation gives less error when combined with the FFT algorithm. Nevertheless, the difference between the two cases is insignificant.
- Linear interpolation works better than the FFT algorithm, where the difference between the two algorithms is more prominent on the contralateral side. The recreation of already existing HRIRs can be attributed to the poor performance of FFT interpolation as discussed above.
- Onset-equalisation increases the interpolation errors, especially on the ipsilateral side.

While the first three observations agree with those made for the comparison of the schemes with the subjects' HRTFs, the increased error with onset-equalisation is completely contradictory (see Fig. A.2). It is possible that for the interpolation from 5°- to 1°-resolution HRIRs, the onset-equalisation may actually make the interpolation process deteriorate, which, however, cannot be confirmed using the subjects' HRTF databases due to the lack of data. (Obviously, recording 360 HRIRs for a human subject with the measurement arrangement given in section 3.2 is unsuitable.) On the other hand, it is also possible that the interpolation performance varies depending on individual HRTF databases.

Considering that the performance difference between LTeq and LT (or LF) is less significant in the overall mean errors shown in Fig. A.3 and that the subjects' HRTFs are those to be actually interpolated, a mixed approach is suggested, which is mainly based on the observations made for Fig. A.2. In the current study, all HRTFs have been interpolated in a way that LTeq was used for all the source locations except for particular ranges on the contralateral sides, $95^\circ \sim 120^\circ$ for the left ear and $240^\circ \sim 265^\circ$ for the right ear, in which LT was employed.

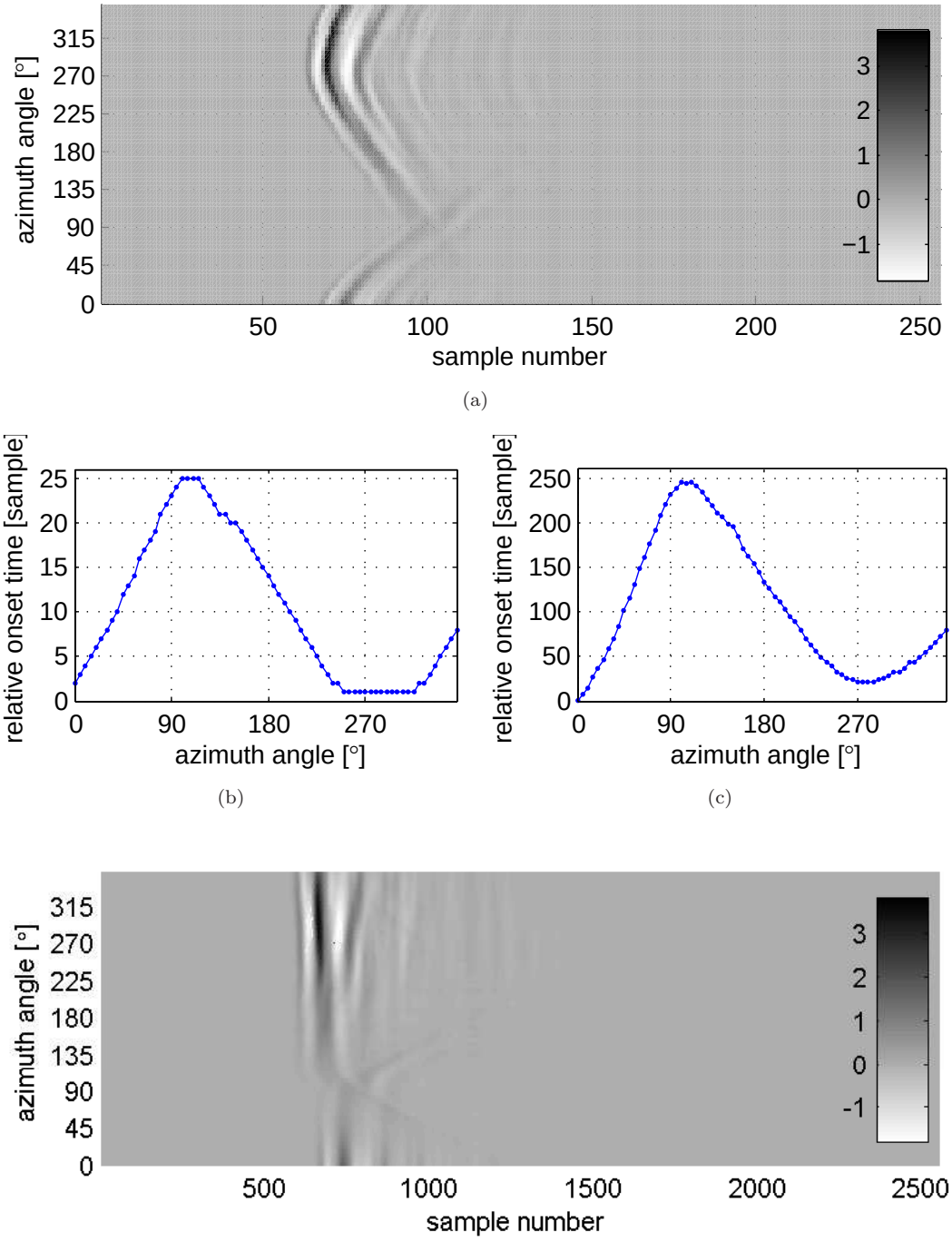
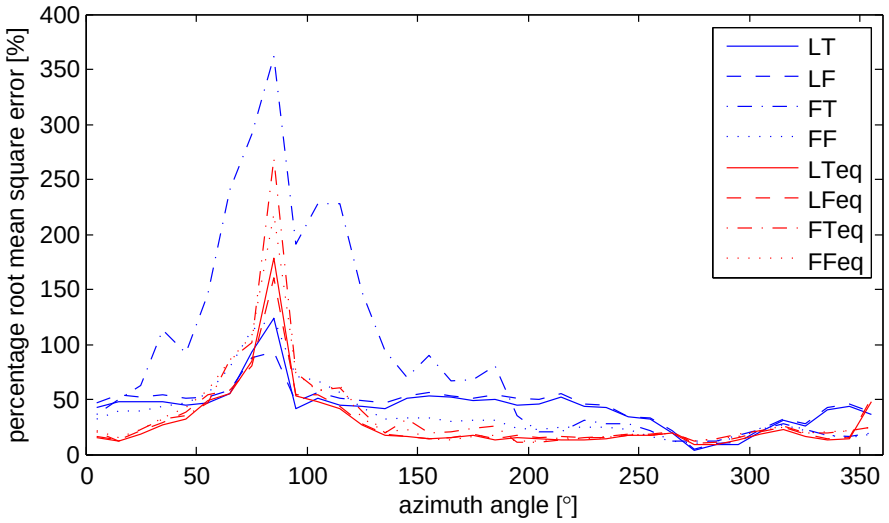
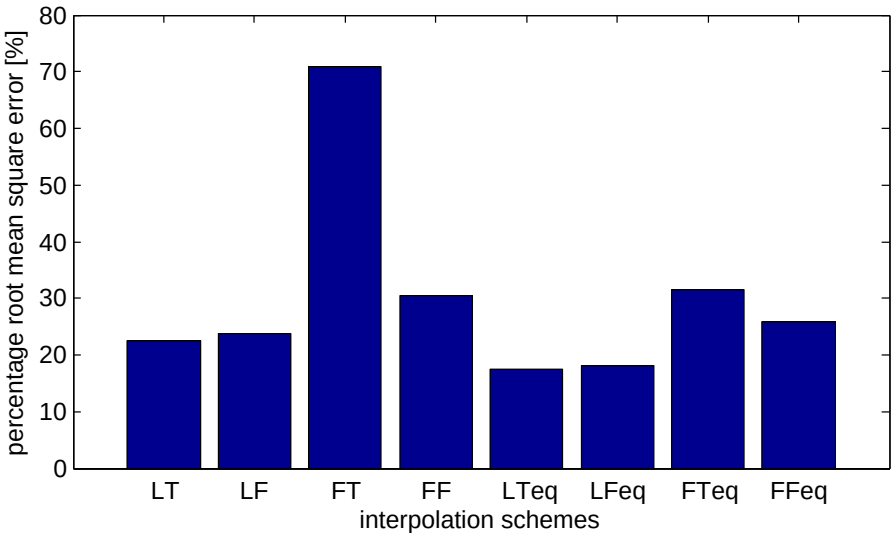


FIGURE A.1: (a) Raw HRIRs for the subject SA (distal-region; left channel). (b) Relative onset time of the raw HRIRs at the original sampling frequency, 48 kHz. (c) Relative onset time obtained with the prior oversampling of raw HRIRs at 480 kHz (d) Alignment by shifting the HRIRs according to the onset times given in panel (c). Note that in (c) and (d), the sample numbers are increased by 10.



(a)



(b)

FIGURE A.2: (a) Percentage root mean square errors (PRMSE) are shown for the 8 interpolation schemes averaged across the 6 subjects' HRTFs measured in chapter 3 (10°-to-5 ° recreation; left channel only). (b) Overall PRMSEs for the subjects' HRTFs are shown for each interpolation scheme.

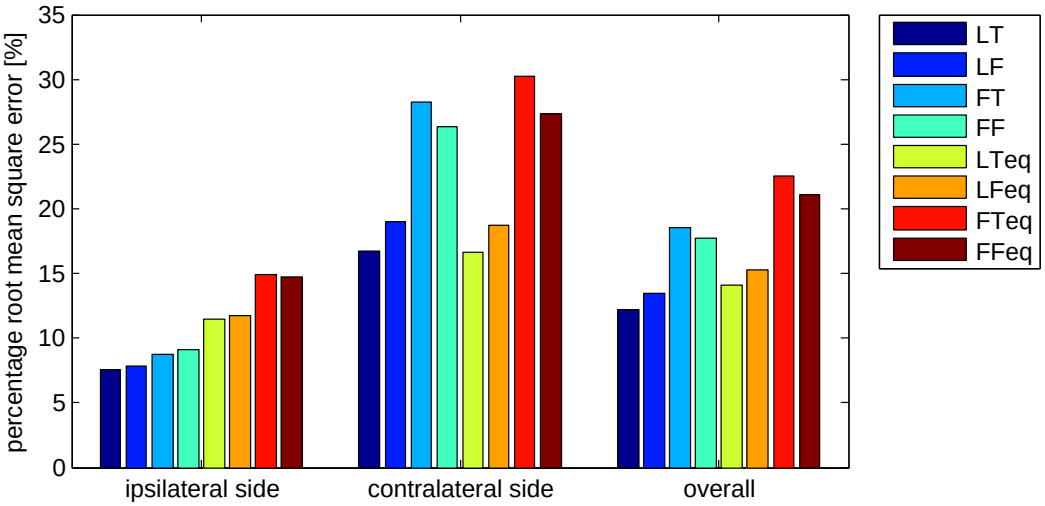


FIGURE A.3: The PRMSEs of the 8 interpolation schemes are shown, where KEMAR HRTFs measured with 1° spatial resolution have been used for the recreation from 5° to 1°. Bar graphs colour-coded for each interpolation scheme are grouped to show the PRMSEs on the ipsilateral side (receiver on the same side of source with respect to the median plane), the contralateral side (receiver on the opposite side of source) and the overall averages.

Appendix B

Normalisation of the EI-pattern

In this section, a brief description of EI-pattern normalisation is given, based on a couple of assumptions. First, it is assumed that the binaural signals on the left and right channels are related to each other by a simple delay τ_0 sec. and attenuation α_0 dB, which can be an approximation of the interaural relation found in the HRTFs:

$$L_i(t) = R_i(t - \tau_0)10^{-\frac{\alpha_0}{20}} \tag{B.1}$$

The double-sided exponential window $w(t)$ in Eq. (5.3) is further assumed to be a rectangular window:

$$w(t') = \begin{cases} 1, & t - \delta t < t' < t + \delta t \\ 0, & \text{elsewhere} \end{cases} \tag{B.2}$$

Then, Eqs. (5.1) and (5.2) can be combined so that

$$\begin{aligned} EI'(i,t,\tau,\alpha) &= \int_{t-\delta t}^{t+\delta t} \big[10^{\frac{\alpha-2\alpha_0}{40}} R_i(t' + \frac{\tau}{2} - \tau_0) - 10^{-\frac{\alpha}{40}} R_i(t' - \frac{\tau}{2})\big]^2 dt' \\ &= 10^{-\frac{\alpha}{20}} \int_{t-\delta t}^{t+\delta t} R_i^2(t' - \frac{\tau}{2})dt' \\ &\quad +10^{\frac{\alpha-2\alpha_0}{20}} \int_{t-\delta t}^{t+\delta t} R_i^2(t' + \frac{\tau}{2} - \tau_0)dt' \\ &\quad +2 \cdot 10^{-\frac{\alpha_0}{20}} \int_{t-\delta t}^{t+\delta t} R_i(t' - \frac{\tau}{2}) R_i(t' + \frac{\tau}{2} - \tau_0)dt' \end{aligned} \tag{B.3}$$

Since $\tau \ll \delta t$, Eq. (B.3) can be reduced to be

$$EI'(i,t,\tau,\alpha) = 2\Psi(0)10^{-\frac{\alpha_0}{20}} \big[\frac{10^{\frac{\Lambda}{20}} + 10^{-\frac{\Lambda}{20}}}{2} - \widehat{\Psi}(T)\big] \tag{B.4}$$

where $T = \tau - \tau_0$, $\Lambda = \alpha - \alpha_0$, and Ψ is the auto-correlation for $R_i(t)$, and $\widehat{\Psi}$ is its normalised version.

The auto-correlation at time 0 is related to the signal energy by

$$\Psi(0)10^{-\frac{\alpha_0}{20}} = \sqrt{e_L e_R} \tag{B.5}$$

where e_L and e_R are the signal energy at the left and the right channel, respectively.

From Eqs. (B.4) and (B.4),

$$\frac{E'(i,t,\tau,\alpha)}{\sqrt{e_L e_R}} = \frac{10^{\frac{\Lambda}{20}} + 10^{-\frac{\Lambda}{20}}}{2} - \widehat{\Psi}(T) \tag{B.6}$$

It is obvious that the normalised EI-pattern (with no noise mask yet applied) retains only the information relating to ITD and ILD, and its value is not affected by the signal amplitude and duration.

Appendix C

IPD Model

In this section, an analytical method is revisited and extended to estimate the location of virtual acoustic image created by the amplitude-panning method. Based on the assumption of free-field sound propagation, the phase difference between two receiver locations can be obtained for a single sound source positioned in the far field. Then, the **IPD model** relates this phase difference to that produced by the 2-channel amplitude panning scheme to give an estimate of source angular location in the horizontal plane, which was the very idea in Blumlein's stereophony [78]. Although there are some reliability issues associated with the free-field assumption and the phase ambiguity, the IPD model can be a good starting point for estimating the actual subjective perception of virtual image positions, and in this section, it will be extended to the asymmetric configuration of the 2-channel loudspeaker system.

First, the phase difference associated with a single acoustic source is approximated between two receiver points. An acoustic field created by a real sound source S is shown in Fig. C.1 along with the locations of two receivers (ears) marked as w_1 and w_2. The sound wave reaching w_2 will travel a longer distance by $h \sin\theta$ compared to the field at w_1, and this difference in travelling distance creates a phase difference ϕ_r. When the source is relatively far from the receivers, ϕ_r can be represented as

$$\phi_r = \frac{\omega h \sin\theta}{c} \tag{C.1}$$

where c is the speed of sound in air, and ω is the angular frequency of the sound wave. Stereophony assumes that the information required for a listener to appreciate a sound location can be provided by this phase difference [79] at low frequencies (approximately below 700 Hz), above which $h \sin\theta$ becomes less than half an acoustic wavelength, thus giving an ambiguous ϕ_r.

Fig. C.2(a) shows a stereophony system configured symmetrically with respect to the median plane, where the amplitude gains to the left and the right channels are **L** and **R**, respectively. Since each loudspeaker subtends an angle ψ with respect to the midline, the sound pressure w_1 and w_2 generated by either transducer will have a phase difference $2\omega\mu = (\omega h \sin\psi)/c$ when 2μ is the arrival time difference [79]. Then, w_1 and w_2 can be represented by

$$w_1 = \mathbf{L}\sin\omega(t-\mu) + \mathbf{R}\sin\omega(t+\mu) \tag{C.2a}$$

$$w_2 = \mathbf{L}\sin\omega(t+\mu) + \mathbf{R}\sin\omega(t-\mu) \tag{C.2b}$$

Eq. (C.2) can be rearranged to be

$$w_1 = \sqrt{2\kappa}\sin\omega(t-\frac{\phi_a}{2}) \tag{C.3a}$$

$$w_2 = \sqrt{2\kappa}\sin\omega(t+\frac{\phi_a}{2}) \tag{C.3b}$$

where

$$\kappa = \mathbf{L}^2 + \mathbf{R}^2 + 2\mathbf{LR}\cos\omega\mu \tag{C.4a}$$

$$\phi_a = 2\tan^{-1}\left(\frac{\mathbf{R}+\mathbf{L}}{\mathbf{R}-\mathbf{L}}\tan\omega\mu\right) \tag{C.4b}$$

The value of ϕ_a is the phase difference delivered by the two in-phase loudspeakers, and if ϕ_a can be made equal to ϕ_r in Eq. (C.1), a sound field can be created to provide a virtual acoustic image at θ_a in azimuth angle as shown in Fig. C.2(a). In this case, the amplitude **L** and **R** follow 'the *sine* law [10]' which is stated as

$$\frac{\sin\theta_a}{\sin\psi} = \frac{\mathbf{L}-\mathbf{R}}{\mathbf{L}+\mathbf{R}} \qquad (\omega\mu \ll 1) \tag{C.5}$$

The above derivation of the relationship between the perceived image position and the input gains of the two loudspeakers can be generalised for an asymmetric stereophony system shown in Fig. C.2(b) where the centre line connecting the midpoint of loudspeakers to the listener has been tilted to one side by θ_c. If this loudspeaker configuration is to be considered, Eq. C.2 has to be modified to be

$$w_1' = \mathbf{L}\sin(\omega t + \phi_1) + \mathbf{R}\sin(\omega t + \phi_2) \tag{C.6a}$$

$$w_2' = \mathbf{L}\sin(\omega t - \phi_1) + \mathbf{R}\sin(\omega t - \phi_2) \tag{C.6b}$$

where ϕ_1 and ϕ_2 are given by

$$\phi_1 = \frac{\omega h \sin(\theta_c - \psi)}{2c} \tag{C.7a}$$

$$\phi_2 = \frac{\omega h \sin(\theta_c + \psi)}{2c} \tag{C.7b}$$

Similar to Eq. (C.3), Eq. (C.6) can be rearranged to be

$$w_1' = \sqrt{A^2 + B^2} \sin(\omega t + \tan^{-1} \frac{B}{A}) \tag{C.8a}$$

$$w_2' = \sqrt{A^2 + B^2} \sin(\omega t - \tan^{-1} \frac{B}{A}) \tag{C.8b}$$

where A and B are given by

$$A = \mathbf{L} \cos \phi_1 + \mathbf{R} \cos \phi_2 \tag{C.9a}$$

$$B = \mathbf{L} \sin \phi_1 + \mathbf{R} \sin \phi_2 \tag{C.9b}$$

From Eq. (C.8), the phase difference ϕ_a between w_1 and w_2 is finally given by

$$\phi_a = 2 \tan^{-1} \frac{B}{A} \tag{C.10}$$

Obviously, equating this ϕ_a to ϕ_r in Eq. (C.1) can give the generalised relationship between the perceived image position and the input gains of the stereophony system:

$$\sin \theta_a = \frac{2c}{\omega h} \tan^{-1} \left(\frac{\sin \phi_1 + \frac{\mathbf{R}}{\mathbf{L}} \sin \phi_2}{\cos \phi_1 + \frac{\mathbf{R}}{\mathbf{L}} \cos \phi_2} \right) \tag{C.11}$$

which is slightly more complex than Eq. (C.5), but readily computable.

It is noticeable that, within the validity of the assumptions discussed in the beginning of this section, Eqs. (C.5) and (C.11) are usable for any pair of input gains regardless of a specific amplitude panning scheme. For example, the input gains obtained by the constant-power panning method [see Eq. (6.2)] can be substituted to Eq. (C.11) to make predictions of the associated image positions.

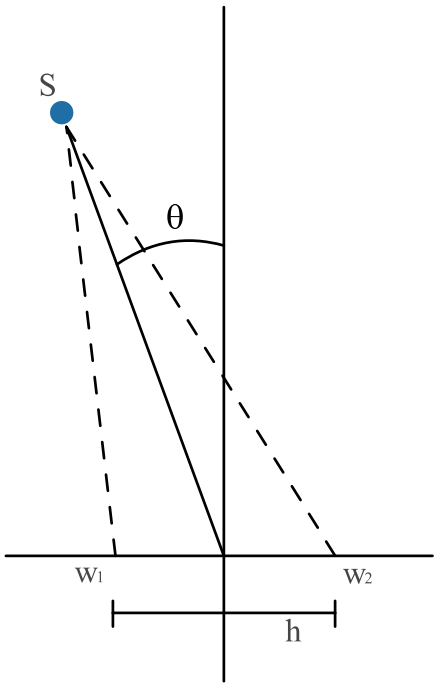


FIGURE C.1: Relative positions of a single sound source and two receivers (ears).

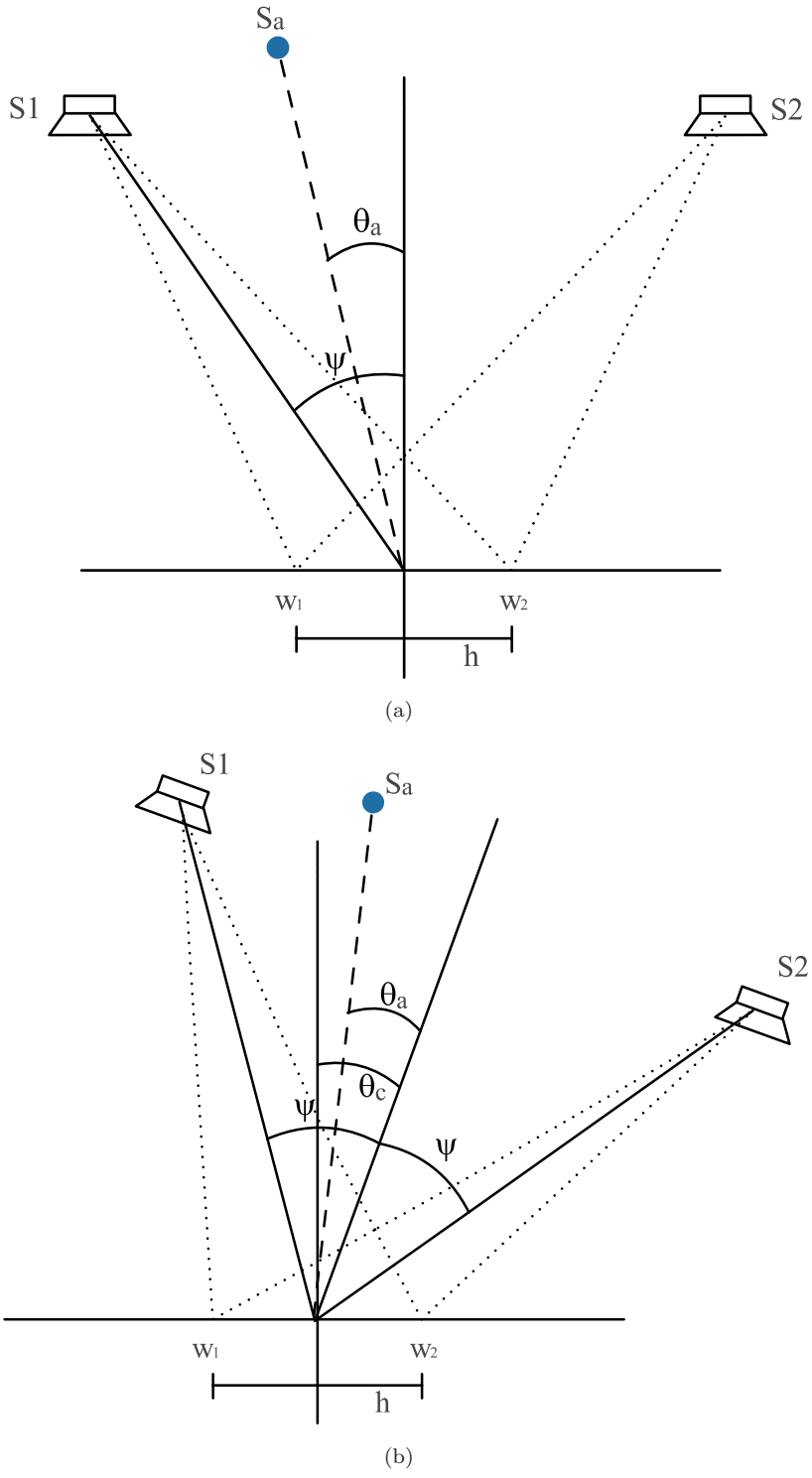


FIGURE C.2: (a) Configuration of a left-right symmetric stereophony system, where an image source is to be positioned at θ_a. (b) Configuration of an asymmetric stereophony system. θ_c represents the centre of the two loudspeakers, while 2ψ is the angular aperture.

Bibliography

[1] J. Breebaart, S. van de Par, and A. Kohlrausch. Binaural processing model based on contralateral inhibition. i. model structure. *Journal of the Acoustical Society of America*, 110(2):1074–1088, 2001.

[2] E. B. Goldstein. *Sensation and perception.* Wadsworth, 6th edition, 2002.

[3] A. J. King, J. W. H. Schnupp, and T. P. Doubell. The shape of ears to come: dynamic coding of auditory space. *Trends in Cognitive Sciences*, 5(6):261–270, 2001.

[4] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT Press, London, 2001.

[5] L. A. Jeffress. A place theory of sound localization. *J. Comp. Physiol. Psychol.*, 41:35–39, 1948.

[6] W. Gaik. Combined evaluation of interaural time and intensity differences - psychoacoustic results and computer modeling. *Journal of the Acoustical Society of America*, 94(1):98–110, 1993.

[7] C. Lim and R. O. Duda. Estimating the azimuth and elevation of a sound source from the output of a cochlear model. In *Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 399–403 vol.1, 1994.

[8] W. Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition .1. simulation of lateralization for stationary signals. *Journal of the Acoustical Society of America*, 80(6):1608–1622, 1986.

[9] E. A. Macpherson. A computer-model of binaural localization for stereo imaging measurement. *Journal of the Audio Engineering Society*, 39(9):604–622, 1991.

[10] V. Pulkki, M. Karjalainen, and J. Huopaniemi. Analyzing virtual sound source attributes using a binaural auditory model. *Journal of the Audio Engineering Society*, 47(4):203–217, 1999.

[11] R. M. Stern and H. S. Colburn. Theory of binaural interaction based on auditory-nerve data .4. model for subjective lateral position. *Journal of the Acoustical Society of America*, 64(1):127–140, 1978.

[12] C. Schauer, T. Zahn, P. Paschke, and H. M. Gross. Binaural sound localization in an artificial neural network. In *Proceedings of ICASSP*, volume 2, pages II865–II868 vol.2, 2000.

[13] T. Dau, D. Puschel, and A. Kohlrausch. A quantitative model of the "effective" signal processing in the auditory system .1. model structure. *Journal of the Acoustical Society of America*, 99(6):3615–3622, 1996.

[14] C. Jin, M. Schenkel, and S. Carlile. Neural system identification model of human sound localization. *Journal of the Acoustical Society of America*, 108(3):1215–1235, 2000.

[15] E. M. Overholt, E. W. Rubel, and R. L. Hyson. A circuit for coding interaural time differences in the chick brain-stem. *Journal of Neuroscience*, 12(5):1698–1708, 1992.

[16] C. E. Carr and M. Konishi. Axonal delay-lines for time measurement in the owls brain-stem. *Proceedings of the National Academy of Sciences of the United States of America*, 85(21):8311–8315, 1988.

[17] G. Plenge. On the differences between localization and lateralization. *The Journal of the Acoustical Society of America*, 56(3):944–951, 1974.

[18] E. R. Hafter. Quantitative evaluation of a lateralization model of masking-level differences. *Journal of the Acoustical Society of America*, 50(4):1116–1122, 1971.

[19] J. Braasch. Localization in the presence of a distracter and reverberation in the frontal horizontal plane: Iii. the role of interaural level differences. *Acta Acustica United with Acustica*, 89(4):674–692, 2003.

[20] C. Faller and J. Merimaa. Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *Journal of the Acoustical Society of America*, 116(5):3075–3089, 2004.

[21] K. D. Martin. Estimating azimuth and elevation from interaural differences. In *Proceedings of the IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pages 96–99, 1995.

[22] D. P. Phillips and S. E. Hall. Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level. *Hearing Research*, 202(1-2):188–199, 2005.

[23] G. H. Recanzone. Rapidly induced auditory plasticity: The ventriloquism aftereffect. *Proceedings of the National Academy of Sciences of the United States of America*, 95(3):869–875, 1998.

[24] S. Carlile, S. Hyams, and S. Delaney. Systematic distortions of auditory space perception following prolonged exposure to broadband noise. *Journal of the Acoustical Society of America*, 110(1):416–424, 2001.

[25] M. S. Brainard, E. I. Knudsen, and S. D. Esterly. Neural derivation of sound source location - resolution of spatial ambiguities in binaural cues. *Journal of the Acoustical Society of America*, 91(2):1015–1027, 1992.

[26] B. C. J. Moore. *An introduction to the psychology of hearing.* Academic Press, 5th edition, 2003.

[27] P. Mannerheim, M. Park, T. Papadopoulos, and P. A. Nelson. The measurement of a database of head related transfer functions. Contract Report 04/07, Institute of Sound and Vibration Research, 2004.

[28] R. M. Stern and C. Trahiotis. Models of binaural perception. In R. H. Gilkey and T. R. Anderson, editors, *Binaural and spatial hearing in real and virtual environments*, pages 499–531. Lawrence Erlbaum Associates, New Jersey, 1997.

[29] R. H. Domnitz and H. S. Colburn. Lateral position and interaural discrimination. *Journal of the Acoustical Society of America*, 61(6):1586–1598, 1977.

[30] M. Bear, B. Connors, and M. Paradiso. *Neuroscience: Exploring the brain.* Lippincott Williams and Wilkins, 3rd edition, 2007.

[31] B. McA. Sayers. Acoustic-image lateralization judgments with binaural tones. *Journal of the Acoustical Society of America*, 36(5):923–926, 1964.

[32] W. M. Hartmann and A. Wittenberg. On the externalization of sound images. *Journal of the Acoustical Society of America*, 99(6):3678–3688, 1996.

[33] P. M. Zurek, R. L. Freyman, and U. Balakrishnan. Auditory target detection in reverberation. *Journal of the Acoustical Society of America*, 115(4):1609–1620, 2004.

[34] W. A. Yost. Lateral position of sinusoids presented with inter-aural intensive and temporal differences. *Journal of the Acoustical Society of America*, 70(2):397–409, 1981.

[35] H. Babkoff, C. Muchnik, N. Ben-David, M. Furst, S. Even-Zohar, and M. Hildesheimer. Mapping lateralization of click trains in younger and older populations. *Hearing Research*, 165(1-2):117–127, 2002.

[36] J. L. Schiano, C. Trahiotis, and L. R. Bernstein. Lateralization of low-frequency tones and narrow bands of noise. In *Proceedings of the ASA conference*, volume 79, pages 1563–1570, 1986.

[37] T. M. Shackleton, R. Meddis, and M. J. Hewitt. Across frequency integration in a model of lateralization. *Journal of the Acoustical Society of America*, 91(4): 2276–2279, 1992.

[38] J. C. Makous and J. C. Middlebrooks. 2-dimensional sound localization by human listeners. *Journal of the Acoustical Society of America*, 87(5):2188–2200, 1990.

[39] S. Carlile, P. Leong, and S. Hyams. The nature and distribution of errors in sound localization by human listeners. *Hearing Research*, 114(1-2):179–196, 1997.

[40] G. F. Kuhn. Model for interaural time differences in azimuthal plane. *Journal of the Acoustical Society of America*, 62(1):157–167, 1977.

[41] J. W. Stutt. On our perception of sound direction. *Philosophical magazine*, 13: 214–232, 1907.

[42] D. Mcfadden and E. G. Pasanen. Lateralization at high-frequencies based on interaural time differences. *Journal of the Acoustical Society of America*, 59(3):634–639, 1976.

[43] H. Moller, M. F. Sorensen, D. Hammershoi, and C. B. Jensen. Head-related transfer-functions of human-subjects. *Journal of the Audio Engineering Society*, 43(5):300–321, 1995.

[44] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. H. Tang. Approximating the head-related transfer function using simple geometric models of the head and torso. *Journal of the Acoustical Society of America*, 112(5):2053–2064, 2002.

[45] B. F. G. Katz. Boundary element method calculation of individual head-related transfer function. i. rigid model calculation. *Journal of the Acoustical Society of America*, 110(5):2440–2448, 2001.

[46] M. Otani and S. Ise. Fast calculation system specialized for head-related transfer function based on boundary element method. *Journal of the Acoustical Society of America*, 119(5):2589–2598, 2006.

[47] H. Hasegawa, M. Kasuga, S. Matsumoto, and A. Koike. Simply realization of sound localization using hrtf approximated by iir filter. *Ieice Transactions on Fundamentals of Electronics Communications and Computer Sciences*, E83a(6):973–978, 2000.

[48] A. Kulkarni and H. S. Colburn. Infinite-impulse-response models of the head-related transfer function. *Journal of the Acoustical Society of America*, 115(4):1714–1728, 2004.

[49] J. Sodnik, R. Susnik, and S. Tomazic. Principal components of non-individualized head related transfer functions significant for azimuth perception. *Acta Acustica United with Acustica*, 92(2):312–319, 2006.

[50] M. J. Evans, J. A. S. Angus, and A. I. Tew. Analyzing head-related transfer function measurements using surface spherical harmonics. *Journal of the Acoustical Society of America*, 104(4):2400–2411, 1998.

[51] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The cipic hrtf database. In *Proceedings of the IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pages 99–102, 2001.

[52] W. G. Gardner and K. D. Martin. Hrtf measurements of a kemar. *Journal of the Acoustical Society of America*, 97(6):3907–3908, 1995.

[53] D. S. Brungart and W. M. Rabinowitz. Auditory localization of nearby sources. head-related transfer functions. *Journal of the Acoustical Society of America*, 106 (3):1465–1479, 1999.

[54] IRCAM. *LISTEN HRTF Database*. http://recherche.ircam.fr/equipes/salles/listen, 2002.

[55] Y. Cho, P. Mannerheim, and P. A. Nelson. Measurement of a near field head-related transfer function database. Contract Report 05/10, Institute of Sound and Vibration Research, 2005.

[56] H. J. Simon, C. C. Collins, A. Jampolsky, D. E. Morledge, and J. Yu. The measurement of the lateralization of narrow bands of noise using an acoustic pointing paradigm - the effect of sound-pressure level. *Journal of the Acoustical Society of America*, 95(3):1534–1547, 1994.

[57] L. R. Bernstein and C. Trahiotis. Measures of extents of laterality for high-frequency "ransposed" stimuli under conditions of binaural interference. *Journal of the Acoustical Society of America*, 118(3):1626–1635, 2005.

[58] E. H. A. Langendijk and A. W. Bronkhorst. Contribution of spectral cues to human sound localization. *Journal of the Acoustical Society of America*, 112(4):1583–1596, 2002.

[59] H. Moller. Fundamentals of binaural technology. *Applied Acoustics*, 36(3-4):171–218, 1992.

[60] D. Howitt and D. Cramer. *Introduction to Statistics in Psychology.* Pearson Education Limited, 3rd edition, 2005.

[61] W. J. Conover. *Practical Nonparametric Statistics.* John Wiley & Sons Inc., 1971.

[62] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten. Peaq - the itu standard for objective measurement of perceived audio quality. *Journal of the Audio Engineering Society*, 48(1-2):3–29, 2000.

[63] A. W. Rix, M. P. Hollier, A. P. Hekstra, and J. G. Beerends. Perceptual evaluation of speech quality (pesq) - the new itu standard for end-to-end speech quality assessment - part i - time-delay compensation. *Journal of the Audio Engineering Society*, 50(10):755–764, 2002.

[64] J. G. Beerends, A. R. Hekstra, A. W. Rix, and M. P. Hollier. Perceptual evaluation of speech quality (pesq) - the new itu standard for end-to-end speech quality assessment - part ii - psychoacoust model. *Journal of the Audio Engineering Society*, 50(10):765–778, 2002.

[65] K. Hartung and S. Sterbing. A computational model of sound localization based on neurophysiological data. In S. Greenberg and M. Slaney, editors, *Computational models of auditory function*, pages 113–126. IOA Press, 2001.

[66] N. I. Durlach. Binaural signal detection: Equalization and cancellation theory. In J. Tobias, editor, *Foundation of modern auditory theory*, volume 2, pages 371–462. Academic Press, New York, 1972.

[67] M. Bodden. Binaural modeling and auditory scene analysis. In *Proceedings of the IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pages 31–34, 1995.

[68] E. D. Boer. Synthetic whole-nerve action potentials for cat. *Journal of the Acoustical Society of America*, 58(5):1030–1045, 1975.

[69] T. Irino and R. D. Patterson. A time-domain, level-dependent auditory filter: The gammachirp. *Journal of the Acoustical Society of America*, 101(1):412–419, 1997.

[70] R. D. Patterson, M. H. Allerhand, and C. Giguere. Time-domain modeling of peripheral auditory processing - a modular architecture and a software platform. *Journal of the Acoustical Society of America*, 98(4):1890–1894, 1995.

[71] M. Slaney. *Auditory Toolbox, ver. 2.* http://rvl4.ecn.purdue.edu/ malcolm/interval/1998-010/, 1998.

[72] F. Palmieri, M. Datum, A. Shah, and A. Moiseff. Learning binaural sound localization through a neural network. In *Proceedings of the IEEE Seventeenth Annual Northeast Bioengineering Conference*, pages 13–14, 1991.

[73] T. R. Anderson, J. A. Janko, and R. H. Gilkey. Modeling human sound localization with hierarchical neural networks. In *Proceedings of the IEEE International Conference on Neural Networks*, volume 7, pages 4502–4507 vol.7, 1994.

[74] O. Abdel Alim and H. Farag. Modeling non-individualized binaural sound localization in the horizontal plane using artificial neural networks. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, volume 3, pages 642–647 vol.3, 2000.

[75] R. M. Stern, A. S. Zeiberg, and C. Trahiotis. Lateralization of complex binaural stimuli - a weighted-image model. *Journal of the Acoustical Society of America*, 84 (1):156–165, 1988.

[76] F. E. Toole and B. McA Sayers. Lateralization judgments and the nature of binaural acoustic images. *The Journal of the Acoustical Society of America*, 37(2):319–324, 1965.

[77] G. H. Recanzone, S. D. D. R. Makhamra, and D. C. Guard. Comparison of relative and absolute sound localization ability in humans. *Journal of the Acoustical Society of America*, 103(2):1085–1097, 1998.

[78] A. D. Blumlein. *Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems.* British Patent No. 34657, 1933.

[79] H. Clark, G. Dutton, and P. Vanderlyn. The stereophonic recording and reproducing system. *IRE. Trans. Audio*, 5(4):96–111, 1957.

[80] F. Rumsey. *Spatial Audio.* Focal Press, London, 2001.

[81] A. T. Sabin, E. A. Macpherson, and J. C. Middlebrooks. Human sound localization at near-threshold levels. *Hearing Research*, 199(1-2):124–134, 2005.

[82] K. Norbert, B. Virginia, and B. G. Shinn-Cunningham. Sound localization with a preceding distractor. In *Proceedings of the ASA conference*, volume 121, pages 420–432, 2007.

[83] G. Theile and G. Plenge. Localization of lateral phantom sources. *Journal of the Audio Engineering Society*, 25(4):196–200, 1977.

[84] G. Martin, W. Woszczyk, J. Corey, and R. Quesnel. Sound source localization in a five-channel surround sound reproduction system. In *Preprint in 107th AES Convention*, 1999.

[85] J. West. *Five-channel panning laws: an analytical and experimental comparison.* MSc thesis, University of Miami, 1998.

[86] F. R. Moore. *Elements of computer music.* Prentice Hall, Englewood Cliffs, NJ, 1990.

[87] E. H. A. Langendijk and A. W. Bronkhorst. Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *Journal of the Acoustical Society of America*, 107(1):528–537, 2000.

[88] T. Takeuchi. *Systems for virtual acoustic imaging using the binaural principle.* PhD thesis, University of Southampton, 2001.