Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing
Espoo 2001 Report 62

SPATIAL SOUND GENERATION AND PERCEPTION BY AMPLITUDE PANNING TECHNIQUES

Ville Pulkki

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing
Espoo 2001 Report 62

SPATIAL SOUND GENERATION AND PERCEPTION BY AMPLITUDE PANNING TECHNIQUES

Ville Pulkki

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission for public examination and debate in Chamber music hall, Sibelius Academy (Pohjoinen Rautatiekatu 9, Helsinki, Finland) on the 3rd of August, 2001, at 12 o'clock noon.

Helsinki University of Technology
Department of Electrical and Communications Engineering
Laboratory of Acoustics and Audio Signal Processing

Teknillinen korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto
Akustiikan ja äänenkäsittelytekniikan laboratorio

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
P.O.Box 3000
FIN-02015 HUT
Tel. +358 9 4511
Fax +358 9 460 224
E-mail lea.soderman@hut.fi

ISBN 951- 22-5531-6
ISSN 1456-6303

Otamedia Oy
Espoo, Finland 2001

Abstract

Spatial audio aims to recreate or synthesize spatial attributes when reproducing audio over loudspeakers or headphones. Such spatial attributes include, for example, locations of perceived sound sources and an auditory sense of space. This thesis focuses on new methods of spatial audio for loudspeaker listening and on measuring the quality of spatial audio by subjective and objective tests.

In this thesis the vector base amplitude panning (VBAP) method, which is an amplitude panning method to position virtual sources in arbitrary 2-D or 3-D loudspeaker setups, is introduced. In amplitude panning the same sound signal is applied to a number of loudspeakers with appropriate non-zero amplitudes. With 2-D setups VBAP is a reformulation of the existing pair-wise panning method. However, differing from earlier solutions it can be generalized for 3-D loudspeaker setups as a triplet-wise panning method. A sound signal is then applied to one, two, or three loudspeakers simultaneously. VBAP has certain advantages compared to earlier virtual source positioning methods in arbitrary layouts. Previous methods either used all loudspeakers to produce virtual sources, which results in some artefacts, or they used loudspeaker triplets with a non-generalizable 2-D user interface.

The virtual sources generated with VBAP are investigated. The human directional hearing is simulated with a binaural auditory model adapted from the literature. The interaural time difference (ITD) cue and the interaural level difference (ILD) cue which are the main localization cues are simulated for amplitude-panned virtual sources and for real sources. Psychoacoustic listening tests are conducted to study the subjective quality of virtual sources. Statistically significant phenomena found in listening test data are explained by auditory model simulation results. To obtain a generic view of directional quality in arbitrary loudspeaker setups, directional cues are simulated for virtual sources with loudspeaker pairs and triplets in various setups.

The directional qualities of virtual sources generated with VBAP can be stated as follows. Directional coordinates used for this purpose are the angle between a position vector and the median plane (θ_{cc}), and the angle between a projection of a position vector to the median plane and frontal direction (ϕ_{cc}). The perceived θ_{cc} direction of a virtual source coincides well with the VBAP panning direction when a loudspeaker set is near the median plane. When the loudspeaker set is moved towards a side of a listener, the perceived θ_{cc} direction is biased towards the median plane. The perceived ϕ_{cc} direction of an amplitude-panned virtual source is individual and cannot be predicted with any panning law.

Keywords: 3-D audio, multichannel audio, amplitude panning, binaural auditory model

This publication is copyrighted. You may download, display and print it for your own personal use. Commercial use is prohibited.

© 2001 Helsinki University of Technology

Preface

This work has been carried out in the Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, from 1995-2001 and during a visiting scholarship at the Center for New Music and Audio Technologies (CNMAT), University of California, Berkeley, USA, during the period August to December 1999.

I am especially thankful to Professor Matti Karjalainen for supporting my work in different ways. Without his deep knowledge in this field this thesis could not have been completed. The co-operation with Mr. Tapio Lokki, Dr. Jyri Huopaniemi, Mr. Tommi Huotilainen, and Docent Vesa Välimäki has been fruitful; I would like to thank them also. I also thank warmly all listening test attendees for their valuable input to this work. The acoustics lab staff has made this work enjoyable. The relaxed and crazy spirit of akulab has made everyday life and especially all Christmas and other parties very boisterous and noisy. Thank you, Aki, Cumhur, Henkka, Hynde, Jari, Juha, Mairas, Mara, Miikka, Prof. Paavo, Poju, Sami, Tero, Tom, Toomas, Unski, without forgetting the others. Special thanks go to Ms. Lea Söderman for management, Hanna and Riitta (Ladies' room members) for providing a mad working environment, and to the staff at the department canteen for filling my stomach.

I was a member of the Graduate School in Electronics, Telecommunications and Automation (GETA) between May 1997 and April 2001. I thank Prof. Iiro Hartimo, the chairman of GETA, for his patience with me, since I applied two times to postpone the dissertation date. I also thank Marja Leppäharju, the secretary of GETA, for making many practical arrangements easier.

The initiative to the study behind this thesis came from Dr. Andrew Bentley and Dr. Pauli Laine from the Sibelius Academy Computer Music Studio (nowadays the Centre for Music & Technology). My warm thanks go to them for the initiative and for believing in my embryonic ideas. I would like to thank Dr. David Wessel and all CNMAT people for hosting a fruitful visit in California. I have had interesting discussions with Dr. Richard Duda of San Jose State University, USA and Dr. Ervin Hafter of UC Berkeley, USA. I also thank the pre-examiners of this thesis, Dr. Bill Gardner and Prof. Lauri Savioja for providing some critical and constructive comments of this thesis.

I also thank my loving wife Sari and my smiling one-year-old boy Veikko for everything. I also acknowledge my parents and parents-in-law, who have supported me in various ways during this work.

I am grateful to the organizations that have found my work important to support enough it financially. I thank the GETA graduate school of the Academy of Finland, Tekniikan edistämissäätiö and the Research Foundation of Helsinki University of Technology.

Table of Contents

List of Symbols

C	gain factor of audio signal
f_m	pinna mode frequency
φ	azimuth angle of virtual source
φ_0	azimuth angle of loudspeakers in stereophonic setup
γ	elevation angle (only in [P1])
γ	a boundary value (only in Sec. 2.2.4)
g	gain factor of a loudspeaker, a subscript may be attached
$\mathbf{g}$	gain factor vector of a loudspeaker set denoted with a subscript
χ^2	chi-square distribution
l	Cartesian loudspeaker direction vector component
$\mathbf{l}$	Cartesian loudspeaker direction vector
$\mathbf{L}$	matrix containing Cartesian direction vectors of a loudspeaker set
L	loudness (in [P7])
$L(f_m)$	loudness at pinna mode frequency (in [P7])
$L_0(f_m)$, $L_0^*(f_m)$	loudness without pinna effect, an estimate of $L_0(f_m)$
L, R, F, B	Ambisonics decoder output Left, Right, Front or Back (only in [P1])
p	Cartesian panning direction vector component
$\mathbf{p}$	Cartesian panning direction vector of a virtual source
ϕ	elevation angle
ϕ_{cc}	angle within a cone of confusion
τ	time lag between ear signals in IACC calculation
θ	azimuth angle
θ_{cc}	angle between a position vector and the median plane
W	soundfield microphone output W, omnidirectional microphone
$x(t)$	sound signal as function of time parameter
$x_i(t)$	sound signal applied to loudspeaker i
X, Y, Z	soundfield microphone output X, Y or Z, directional pattern is figure of eight faced towards respective coordinate axis

List of Abbreviations

AES	Audio Engineering Society
ANOVA	analysis of variance
CLL	composite loudness level
DIVA	Digital Interactive Virtual Acoustics
DLL	directional loudness level
DSP	digital signal processing
ERB	equivalent rectangular bandwidth
FIR	finite impulse response
GETA	Graduate School for Electronics, Telecommunications and Automation
GTFB	gammatone filterbank
HRTF	head-related transfer function
HUT	Helsinki University of Technology
IACC	interaural cross-correlation
ILD	interaural level difference
ILDA	interaural level difference angle
ITD	interaural time difference
ITDA	interaural time difference angle
JND	just noticable difference
KEMAR	Knowles Electronics Mannequin for Acoustics Research
MDAP	multiple-direction amplitude panning
MOA	method of adjustment
VBAP	vector base amplitude panning
VLAL	very large array of loudspeakers

Chapter 1

Introduction

In many theaters and auditoria there exist sound reproduction systems that include a large number of loudspeakers. Audio systems with multiple loudspeakers are also becoming common in domestic use. Typically loudspeakers are in different positions in domestic use, in concert halls, and in studios. Generally a sound recording that is produced for one loudspeaker setup creates different spatial attributes if played back with another loudspeaker setup. The ideal case is a system that would be able to produce sound with identical spatial attributes using different loudspeaker configurations, and the sound would be perceived similarly in different positions of a large listening area. There exists some attempts to create such systems. However, they may be constrained to certain types of loudspeaker setups, or they might have timbral or spatial artefacts.

Methods for multi-loudspeaker setups are typically based applying amplitude panning [1] in which the same sound signal is applied to two or more loudspeakers equidistant from a listener with appropriate non-zero amplitudes. The listener perceives a virtual source in a location that does not coincide with any of the physical sound sources. In perfect reproduction the direction and spreading of a virtual source should be reproduced exactly as targeted. However, some imperfections occur in practice. Typically virtual sources cannot be produced in certain directions, and they are perceived to be spatially spread although point-like sources were desired. The directional quality of a virtual source is then degraded.

Traditionally the directional quality of virtual sources has been studied by listening tests. The listeners have described the virtual source position in some way, which requires extensive testing to obtain reliable and generalized results. To avoid making listening tests and to obtain a in-depth view of the perception of virtual sources, it is desired to create an objective tool for spatial sound quality evaluation. To create such a tool, the understanding of spatial hearing mechanisms is essential. Certain spatial attributes, such as basic cues of directional hearing, are well enough known.

In spatial hearing research the perception of virtual sources is of interest because different directional cues of a virtual source are not necessarily consistent with each other. The way humans assess weights to these cues, which suggest multiple directions, reflects something about the functioning of human spatial hearing. An important feature is the determination of which auditory cues listeners rely on the most when inconsistent cues are present. Also, when amplitude panning is applied to loudspeakers that are in the median plane, the functioning of elevation perception mechanisms can be explored. These mechanisms are partly unknown.

A field in which the study of localization of virtual sources is of particular interest is virtual

acoustic environments. Sound signals reaching a listener directly or via reflections are produced by a virtual source positioning method through loudspeakers or headphones. The evaluation of virtual source quality is of crucial importance in this approach.

1.1 Aims of the thesis

An aim of this thesis is to propose a generic virtual source positioning method for arbitrary loudspeaker setups. The method should be able to produce virtual source directions as similar as possible in different loudspeaker setups for large listening areas. Another aim is to study the objective evaluation of virtual sources, created using different spatial sound systems, with a binaural auditory model. It also aims to measure the directional quality of the proposed spatial sound method by psychoacoustic listening tests and by simulations using a binaural auditory model.

1.2 Organization of this thesis

The thesis is organized in the following manner. Chapter 2 provides an overview of human spatial hearing. Different spatial sound reproduction methods are reviewed in chapter 3. Chapter 4 discusses subjective and objective methods to examine the directional quality of virtual sources and in Chapter 5 the directional quality of amplitude-panned virtual sources is researched. The study is concluded in Chapter 6.

Chapter 2

Spatial hearing

Humans associate spatial attributes, such as direction and distance, to auditory objects. An auditory object as an internal percept is caused by a sound source that emits sound signals (or sound waves). The perception of such physical phenomena is denoted by the adjective "auditory", such as auditory object. Humans are also able to perceive attributes of the space they are in from reflections and reverberation to some degree. In this chapter the main issues of spatial hearing are considered.

2.1 Ear canal signals

In [2] it is stated that the only input that listeners use in the perception of spatial sound is the sound pressures in ear canals captured by ear drums. It is therefore of interest to study what appears in a listener's ear canals in a room in which a sound source is present. The first approximation would be that the sound in the ears is equal to the sound signal emanated by a sound source. However, in nature it exists very seldom that a sound wave propagates only via one path from a sound source to a listener. The ground reflects sound, and if walls and a ceiling are present, there will be more reflections, which increase the number of propagation paths between the sound source and the listener. Typically, some diffraction also occurs in rooms. The reflections, reflections of reflections, and diffraction are fused to diffuse reverberation after a time interval [3]. The acoustic energy of reverberation decreases with time due to air absorption and losses in boundary reflections. All these sound signals arrive at the ear canals at different times from different directions, thus yielding signals that differ significantly from the sound source signal because of the room effect.

In addition to the room effect, sound pressures at ear canals also change as a function of sound source location relative to a listener. When a sound signal travels through space and reaches a listener's ears, it changes due to complex wave acoustics near the listener's torso, head, and pinna. The free-field transfer functions from a sound source to the ear canal are called head-related transfer functions (HRTF) [2]. They are heavily dependent on the direction of a sound source. Since the ears are located on different sides of the skull, the arrival times of a sound signal vary with direction. Also, the skull casts an acoustic shadow that causes the contralateral ear signal to have less magnitude. Shadowing is most prominent at high frequencies, and does not exist when the frequency is very low. The pinna and other parts of the body also change the sound signal.

HRTFs are dependent on distance as well. However, the dependence occurs when the source is

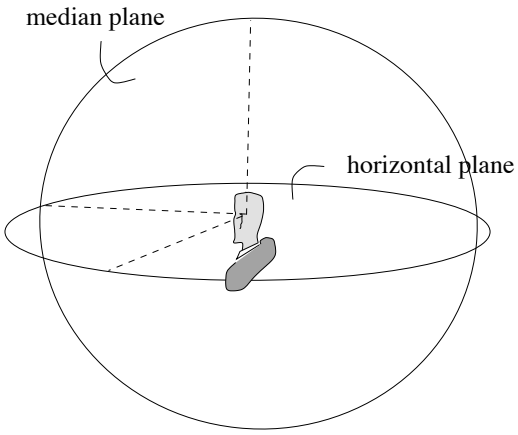

Figure 2.1: Median and horizontal planes.

close to the listener's head [4]. Far away from the listener the distance-dependence factor can be omitted. In this work sound sources are always located at least 2 meters distance from the listener, therefore this phenomenon is not considered further.

2.2 Human spatial hearing

Human listeners decode spatial information from different types of cues [5]: spectral content of ear canal signals, spectral or temporal differences between ear canal signals, and effect of head rotation to perceived binaural differences. These cues and their decoding mechanisms are considered in the following section.

2.2.1 Coordinate systems

Before perceptual issues are studied, some definitions are in order. Two planes that are important in spatial hearing and in spatial sound reproduction are presented in Fig. 2.1. The plane that divides symmetrically the space related to a listener into left and right parts is called the median plane. Each point in the median plane is equidistant from both the listener's ears, and if the listener's head is assumed to be symmetrical, it is symmetrical with respect to the median plane. The plane that divides space into upper and lower parts is called the horizontal plane. All points in the horizontal plane share the same height with both ears.

In spatial hearing an important concept is the cone of confusion. The cone is defined as a set of points which all satisfy following condition, the difference of distances from both ears to any point on the cone is constant. A cone of confusion can be approximated by a cone having axis of symmetry along a line passing through the listener's ears and having the apex in center point between the listener's ears, as in Fig. 2.2.

Spherical coordinates are often used to denote sound source directions in research of spatial hearing. Conventionally, they are denoted by azimuth (θ) and elevation (ϕ). This is, however, not a convenient coordinate system, since a cone of confusion can not be specified easily. An alternative spherical coordinate system has been used by Duda [6]. The sound source location is specified by defining the cone in which it lies, and further by the direction within the cone. The cone is defined

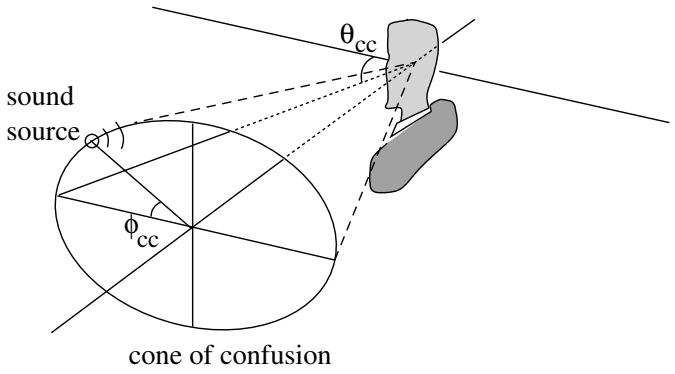

Figure 2.2: Cone of confusion. Spherical coordinate system that is convenient in directional hearing. The direction of a sound source is denoted as (θ_{cc},ϕ_{cc}) pair.

by the angle θ_{cc} between the median plane and the cone, Fig. 2.2. Variable θ_{cc} may have values between $-90°$ and $90°$. The direction within the cone of confusion is denoted as ϕ_{cc}, and it varies between $-180°$ and $180°$. These coordinates are used with conventional azimuth and elevation coordinates in this study. Standard nomenclature for these variables is lacking, though it would be beneficial. In some studies θ_{cc} is referred to as the left/right (L/R) direction.

2.2.2 Transforming ear canal signals to neural impulses

There are multiple organs involved in decoding ear canal signals to neural information. A good review of the functioning of different parts can be found, e.g., in [7]. The sound wave that arrives from a sound source to ear canal is turned to the mechanical motion of the ear drum, which is a membrane at the bottom of the ear canal. This motion is transmitted via ossicles to the cochlea which transforms the mechanical motion to neural impulses. Mechanical vibrations cause a traveling wave to the basilar membrane. The stiffness and other attributes of the membrane vary with position, which causes the displacement of vibration to reach its maximum at a position that is related to the frequency of the wave. The membrane movement stimulates receptor cells that are attached to it. The pulses of receptor cells are then transmitted to the cochlear nucleus, which acts as a preamplifier and preprocessor. The cochlear nucleus has a number of different responses to the same stimulus and the responses are passed to different parts of the brain [7].

In this study we are particularly interested in simulating the functioning of the hearing system in a way that is computationally achievable. The physiology of the auditory system is far too complicated to be modeled in detail. A simple way to approximate the propagation of sound from the external to the inner ear is to use a digital filter that simulates the minimum audible pressure (MAP) [8]. It yields the frequency response of the middle ear and to some extent also the frequency response of the cochlea.

The cochlea is often modeled using a filterbank that consists of filters modeling its frequency selectivity, for example by using a gammatone filterbank [9]. The hair cell behaviour can be simulated functionally by half-wave rectification and low-pass filtering. More advanced hair cell simulations exist [10] but they require too much computational facilities from the viewpoint of this study.

2.2.3 Monaural directional cues

Monaural cues are decoded from sound arriving at one ear, no binaural interaction is involved. Monaural temporal cues, monaural spectral cues, and monaural overall level are potential candidates to be spatial cues. However, in [5] it is stated that monaural temporal cues do not take part in the localization process. The overall level of sound affects distance perception mostly [2], which is beyond the scope of this thesis. However, monaural spectral cues have proven to be quite strong directional cues [5].

Monaural spectral cues

It is commonly accepted that monaural spectral cues mostly carry information on ϕ_{cc} direction of a sound source. In pinna-occlusion experiments it has been demonstrated that the localization ability in the median plane deteriorates when the pinna cavities are filled [11, 12]. The effect of the pinna is therefore an important cue. However, a debate on the spectral features that are decoded is still continuing.

Blauert [13] suggested that the frequencies of spectral peaks in ear canal signals are a prominent cue in ϕ_{cc} direction decoding. In some studies the frequencies of only spectral notches were found to be prominent cues [14], while other studies state that the frequencies of both notches and peaks are important [15]. Middlebrooks [16] calculated the correlations between shapes of HRTFs and shapes of sound spectra appearing at the ears. Zakarauskas & Cynander [17] suggested that the second finite difference of the spectrum could be used as a cue. It has been suggested that the spectral modifications could be decoded from the ILD spectrum [6] when the sound source is outside the median plane. It has also been shown that interaural cross-correlation conveys elevation and front-back discrimination ability in the median plane [18] when neural nets are used in elevation decoding.

The author studied ϕ_{cc} direction perception of amplitude-panned virtual sources in [P7]. In listening tests the subjects perceived the elevation of virtual source changing with panning angle. The auditory spectra of virtual sources were monitored, and it was found that the notches or peaks of the spectrum did not change with panning angle. This contradicts Blauert's [13] hypothesis. An alternative directional cue is hypothesized in [P7], based on decoding the loudness level at modal frequencies of the pinna and comparing it with an estimate of loudness level without the pinna effect. The hypothesis could not, however, be proven.

Modeling of monaural spectral cues

Loudness is a perceptual attribute that decodes how loud an auditory object is perceived [8]. The monaural spectrum can be modeled simply by calculating loudnesses of sound signals appearing in the frequency channels of a filter bank model of the cochlea using an appropriate formula [19, 20]. The loudness as a function of the filter bank channel number can be treated as a loudness spectrum of a virtual or real source.

2.2.4 Binaural cues

Binaural cues are derived from temporal or spectral differences of ear canal signals. They are the strongest directional cues. Temporal differences are called the interaural time differences (ITD) and spectral differences are called the interaural level differences (ILD). These differences are caused respectively by the wave propagation time difference (primarily below 1.5 kHz) and the shadowing effect by the head (primarily above 1.5 kHz). When a sound source is shifted within a cone of confusion, ITD and ILD remain constant, although, there might be some frequency-dependent changes in ILD [6]. These cues thus provide information in which cone of confusion a sound source is. In (θ_{cc},ϕ_{cc}) spherical coordinate system this means that θ_{cc} can be decoded with ILD or ITD. The auditory system decodes the cues in a frequency-dependent manner. These cues have been studied extensively, and a comprehensive review can be found in [2, 21].

In psychoacoustic tests it has been found that mechanisms that decode ITD are sensitive to the phase of signal at low frequencies below roughly 1.6 kHz and to envelope time shifts at high frequencies [2]. ITD is a quite stable cue; its value is almost constant with frequency. The absolute value, however, is higher at low frequencies. Also, the dependence of ITD on θ_{cc} is monotonic.

It has been found that ILD mechanisms are sensitive to differences of sound pressures at ear canals. Due to complex wave acoustics around the head, ILD is largely dependent on frequency: it is negligible at low frequencies and increases nonmonotonically with frequency. The behaviour of ILD with θ_{cc} is also problematic. It behaves monotonically only within some region $-\gamma < \theta_{cc} < \gamma$ [2] depending on frequency, where the value of γ is typically $40\,^{\circ} - 80^{\circ}$.

If ILD and ITD suggest conflicting sound source directions, it is somewhat unclear which cue will win. Traditionally it has been stated that ITD is dominant at low frequencies and ILD at high frequencies. There might also exist multiple images, or the perceived source may be spread to a larger image [2]. Wightman & Kistler [5] have proposed that when the cues are distorted, the auditory system applies the cue which is the most consistent one. A cue is consistent if it suggests the same direction in a broad frequency band.

Physiology and modeling of binaural cues

Scientists agree quite widely that the organ that decodes ITD is the medial superior olive [22] which is located in the brain stem. The medial superior olive has cells that fire when a pulse arrives from both ears simultaneously. The lengths of neural paths from each ear to the medial superior olive also change systematically at different parts of the organ, which enables decoding of ITD values. The location of ILD decoding is not known as thoroughly, but it is hypothesized that it occurs in multiple organs, including the lateral superior olive and the inferior colliculus [7].

The ITD between ear signals can be simulated with models that are based on the coincidence-counting model by Jeffress [23]. Cross-correlations with different time lags between ear signals are calculated. The time lag that produces the highest value for cross-correlation is considered as the ITD value. ITD is calculated at different frequency bands, which produces values as a function of frequency. It has also been shown that if ITD has the same value at adjacent frequency bands, it is considered more relevant in localization. To simulate this, a second-level coincidence counting unit can be added [24]. The cross-correlation function at a frequency band is multiplied with cross-correlation functions at the adjacent bands.

The ILD can be modeled by calculating the loudness level of each frequency band and by subtract-

ing the values from the corresponding value at the contralateral ear. The difference of loudness levels between the ears at each frequency band is treated as the ILD spectrum.

2.2.5 Precedence effect

The precedence effect [2, 25] is an assisting mechanism of spatial hearing. It is a suppression of early delayed versions of the direct sound in source direction perception. This helps in reverberant rooms to localize sound sources. In a reverberant room when the reflections reach the listener, they are summed up to the direct sound, which changes the binaural cues prominently. Therefore, the only reliable cues are cues generated by direct sound without reflections. The direct sound can be heard only when a sound appears after silence, or when the level of a new sound is high enough for a short time, i.e., a transient occurs. The precedence effect takes advantage of this, and it searches actively for transients and uses binaural cues only for a short time ($\approx$ 1 ms) after a transient has occurred. It is a complex mechanism that has remained partly unexplored. There seem to be both low- and high-level mechanisms involved. The latest knowledge about precedence effect is reviewed in [26].

In the analysis of virtual sources the precedence effect model is often excluded. When a model lacks the precedence effect, it gives reliable results only if all incidents of a sound signal arrive to the ears of a listener within about a one-millisecond time window. This can be achieved only in anechoic conditions, since in all normal rooms there exist reflections and reverberation that violate the 1 ms window.

The precedence effect thus ensures that the localization mechanisms use only direct sound. Direct sound is not changed by room acoustics, thus it is equal in different rooms. This is a significant fact relating to this study; the listening test and simulation results that have been achieved in anechoic conditions can be applied at least qualitatively also in moderately reverberant conditions.

2.2.6 Effect of head rotation on binaural cues

When a listener rotates his/her head, the binaural cues change depending on the direction of the sound source [2]. The information coded in the changes could be used in localization. However, recently it has been found that this effect is used as a coarse cue that suggests if the sound source is in front, back, or above the listener [27].

2.3 Virtual sources

A virtual source denotes an auditory object that is perceived in a location that does not correspond to any physical sound source. Different methods to create virtual sources are reviewed in the following chapter. Typically the auditory cues of virtual sources do not correspond to cues of any real sources. In auditory research they are of interest because the way listeners perceive such distorted cues reflects human mechanisms for spatial sound perception.

If the auditory cues of a virtual source were equal to a small-sized real source, the virtual source could then be described as “point-like” or “sharp”. Typically there exist some deviations in virtual source cues in different frequency bands and between different cues. A virtual source can then be perceived to be “spread”, i.e., it is no longer point-like, and the perceived size of a virtual source

is bigger. In some cases the virtual source can be perceived to be “diffuse”, the direction of which is then undefined. In some cases it appears inside the listener’s head.

Chapter 3

Spatial sound reproduction

The term "spatial sound reproduction" denotes methods to reproduce or synthesize also the spatial attributes of sound to listeners. In this thesis the scope is in synthesizing spatial impressions; the aim is not to recreate spatial sound that existed and was recorded on some occasion. Different methods to position virtual sources are reviewed which aim to produce point-like virtual sources to defined directions. The quality that can be achieved with them is discussed as well.

3.1 Amplitude panning

Amplitude panning is the most frequently used panning technique. In it a sound signal is applied to loudspeakers with different amplitudes, which can be formulated as

$$x_i(t) = g_i x(t), \quad i = 1, \ldots, N, \tag{3.1}$$

where $x_i(t)$ is the signal to be applied to loudspeaker i, g_i is the gain factor of the corresponding channel, N is the number of loudspeakers, and t is the time parameter. The listener perceives a virtual source the direction of which is dependent on the gain factors.

3.1.1 Stereophony

Stereophonic listening configuration is most used listening setup. In it there are two loudspeakers placed in front of a listener, as illustrated in Fig. 3.1. The aperture of loudspeakers is typically $60\ ^\circ$. In the figure variable θ denotes the perceived azimuth of the virtual source. There are a number of panning laws that estimate θ from the gain factors of loudspeakers. The estimated direction is called panning direction or panning angle. In the panning laws that are presented in this section the panning directions θ_S, θ_T or θ_G are estimates of θ.

In the sine law [28] of perceived direction the wave propagation path from loudspeakers to ears was modeled with single straight lines. The path from the contralateral loudspeaker therefore penetrates the listener's head; that is highly unnatural. It is presented as

$$\frac{\sin\theta_S}{\sin\theta_0} = \frac{g_1 - g_2}{g_1 + g_2}, \tag{3.2}$$

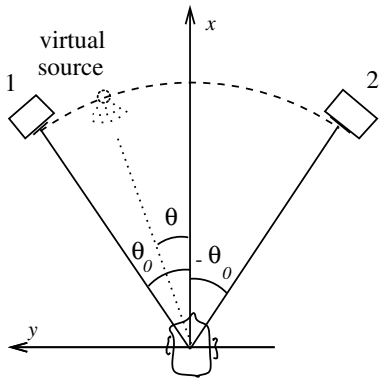

Figure 3.1: Standard stereophonic listening configuration.

where θ_S is an estimate of perceived azimuth angle (panning angle) of a virtual source. θ_0 is the loudspeaker base angle, as in Fig. 3.1. The equation is valid only when the frequency is below 500 Hz and when the listener's head is pointing directly forward. In the equation it is also assumed that the elevation is 0°. The equation does not set limitations to θ_S, but in most cases its value is set to satisfy $|\theta_S| \leq \theta_0$. If $|\theta_S| > \theta_0$ the amplitude panning will produce antiphase loudspeaker signals which may distort the virtual source [2], as is shown in [P5].

Bennett et al. [29] derived the law by improving the head model used in derivation of sine law by approximating the propagation path from contralateral loudspeaker to ear with a curved line around the head. He ended up with a law that was earlier proposed for different listening conditions in [30, 31]

$$\frac{\tan\theta_T}{\tan\theta_0} = \frac{g_1 - g_2}{g_1 + g_2}. \tag{3.3}$$

This equation is called the tangent law, and it has the same limitations as the sine law. The tangent law is equivalent to the law that is often used in computer music [32]

$$\begin{cases} g_n = \cos\theta_x \\ g_m = \sin\theta_x, \end{cases} \tag{3.4}$$

where n and m denote two adjacent loudspeaker, and θ_x is the angle between panning direction and loudspeaker n, and the loudspeaker aperture is 90°. The equivalency of Eqs. 3.3 and 3.4 can be proved with a simple trigonometric manipulation. In [P7] the author formulated sine and tangent laws with θ_{cc} directions, which also allows one to use the laws when the ϕ_{cc} directions of loudspeakers are not the same.

The gain factors cannot be resolved as such with sine and tangent laws. They only state the relation between gain factors. To be able to solve the gain factors, an equation can be stated that keeps the perceived virtual source loudness constant

$$\sqrt[p]{\sum_{n=1}^{n=N} g_n^p} = 1. \tag{3.5}$$

Here p can be chosen differently, depending on listening room acoustics, and its value affects the loudness of the virtual source [32]. In anechoic listening the virtual source loudness is roughly

equal when $p = 1$, and in real rooms with some reverberation the value is often set to $p = 2$. The first case preserves roughly the amplitude of virtual source signals in the ear canals, and the latter case, the energy of them.

There exists one more panning law that has been proposed by Chowning [33]

$$\begin{cases} g_n = \sqrt{\frac{\theta_m - \theta_G}{\theta_m - \theta_n}} \\ g_m = \sqrt{\frac{\theta_n - \theta_G}{\theta_n - \theta_m}}, \end{cases} \tag{3.6}$$

where θ_n and θ_m are the azimuth angles of an adjacent loudspeaker pair, θ_G is the panning angle, and g_n and g_m are the gain factors of loudspeaker channels n and m. In the derivation of this law the perceived direction of a virtual source was not estimated with any theory or listening tests. In [34] a standard nomenclature was proposed for tangent law with loudness normalization parameters $p = 1$ and $p = 2$, and for Chowning's law. In the same publication an alternative panning law was proposed that is equivalent to Chowning's law (Eq. 3.6) without square roots. The perceptual differences between different panning laws are simulated in Sec. 5.2.3.

3.1.2 2-D loudspeaker setups

In 2-D loudspeaker setups all loudspeakers are in the same plane with a listener. Typically the loudspeakers are in the horizontal plane. In quadraphonic setups four loudspeakers are placed evenly around the listener in azimuth angles $\pm 45\,^\circ$ and $\pm 135^\circ$, and in 5.1 setups five loudspeakers are in directions 0°,$\pm 30^\circ$, and $\pm 110^\circ$ as is standardized in [35].

Pair-wise amplitude panning [33] methods can be used in such loudspeaker systems. The sound signal is applied to two loudspeakers between which the panning direction lies. If a virtual source is panned coincident with a loudspeaker, only that particular loudspeaker emanates the sound signal. This yields the best possible directional quality that is achievable with amplitude panning. However, a drawback is that the directional spread varies with panning direction. When there is a loudspeaker in the panning direction, the virtual source is sharp, but when panned between loudspeakers, some spreading occurs.

This is avoided by applying the sound at least to two loudspeakers each time. In some methods the sound is applied to all loudspeakers, as in Ambisonics (Sec. 3.1.4). The directional spread is not then dependent on panning angle. However, the directional quality degrades prominently in such systems outside the best listening position, since the virtual sources are localized to the nearest loudspeaker that produces the virtual source signal. The virtual source sound signal arrive from that loudpeaker first to the listener, and the virtual source is localized to it due to the precedence effect. However, if the level of nearmost loudspeaker is significantly ($\approx$ 15 dB) lower than other loudspeakers that produce the same signal, the sound is localized elsewhere [2]. This effect is less disturbing with pair-wise panning than with systems that apply a sound to all loudspeakers. A virtual source is localized in all listening positions between the two loudspeakers that are used to generate it, or to either of them. The directional quality degrades less therefore outside the best listening position. Also, if the number of loudspeakers is increased, the virtual source directions perceived by different listeners are more similar, which does not occur with systems in which the sound is applied to all loudspeakers.

3.1.3 3-D loudspeaker setups

A three-dimensional loudspeaker setup denotes here a setup in which all loudspeakers are not in the same plane with the listener. Typically this means that there are some elevated and/or lowered loudspeakers added to a horizontal loudspeaker setup. Triplet-wise panning can be used in such setups. In it, a sound signal is applied to a maximum of three loudspeakers at one time that form a triangle from the listener's view point. If more than three loudspeakers are available, the setup is divided into triangles, one of which is used in the panning of a single virtual source at one time. The number of active loudspeakers is then one, two, or three at a time. Virtual sources can be also created by using 3-D setups with methods in which the sound is applied to all of the loudspeakers or to some subset of them. The discussion of drawbacks and advantages of pair-wise panning that was presented in the previous section is also valid here.

Triplet-wise panning is used commercially [36]. Ellison invented a SpaceNodes algorithm in 1992 [37]. The loudspeakers and virtual source locations are manually marked on a two-dimensional plane, and loudspeaker triplets are also formed manually. Gain factors are computed as barycentric coordinates of a virtual source location inside a loudspeaker triangle. The method is straightforward and can be used on many occasions succesfully. However, the relation of the plane coordinates of loudspeakers and virtual sources to physical or perceived directions of them is vague. For many uses a more generic method is desirable. Vector base amplitude panning proposed in [P1] is a method to formulate triplet-wise panning with loudspeakers in arbitrary setups, and it will be discussed in Sec. 3.6.1.

3.1.4 Ambisonics

Ambisonics is basically a microphoning technique [38]. However, it can also be simulated to perform a synthesis of spatial audio [39]. In this case it is an amplitude panning method in which a sound signal is applied to all loudspeakers placed evenly around the listener with gain factors

$$g_i = \frac{1}{N}(1 + 2\cos\alpha_i), \tag{3.7}$$

where g_i is the gain of i:th speaker, N is the number of loudspeakers, and α is the angle between loudspeaker and panning direction [40]. The sound signal therefore emanates from all loudspeakers, which causes spatial artefacts described in Sec. 3.1.2. Second-order Ambisonics applies the sound with gain factors

$$g_i = \frac{1}{N}(1 + 2\cos\alpha_i + 2\cos 2\alpha_i) \tag{3.8}$$

to a similar loudspeaker system [41]. The sound is still applied to all of the loudspeakers, but the gains have prominently lower absolute values on the opposite side of a panning direction. This creates fewer artefacts. However, to get an optimal result, the loudspeakers should be in a symmetric layout, and increasing the number of them would not enhance the directional quality beyond a certain amount of loudspeakers. In Fig. 3.2 the gain factors for a quadraphonic loudspeaker setup are plotted as a function of panning angle for first- and second-order Ambisonics.

Ambisonics can be used with three-dimensional loudspeaker setups also. Typically it is applied to eight loudspeakers in a cubical array, or to twelve loudspeakers as two hexagons on top of each other.

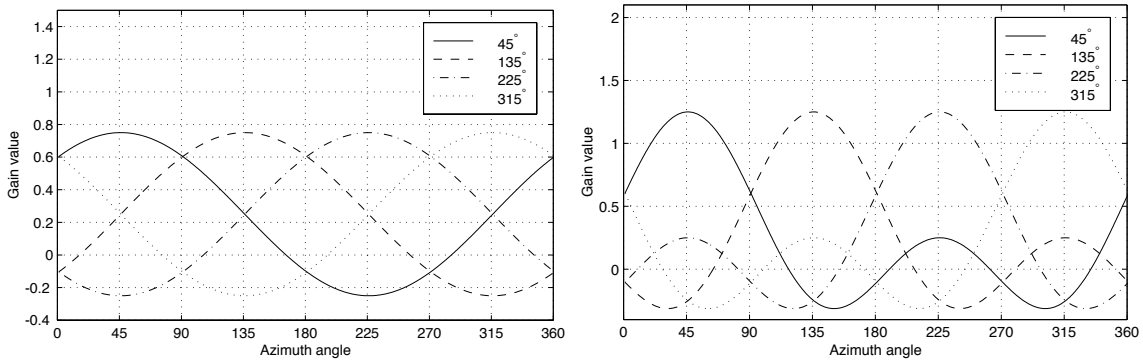

Figure 3.2: Gain factors for a quadraphonic loudspeaker setup calculated with a) first-order Ambisonics b) second-order Ambisonics.

3.2 Time panning

When a constant delay is applied to one loudspeaker in stereophonic listening, the virtual source is perceived to migrate towards the loudspeaker that radiates the earlier sound signal [2]. Maximal effect is achieved when the delay is approximately 1.0 ms.

In tests with different signals it has been found, however, that the perceived direction of virtual sources is dependent on frequency [42, 43]. When the author simulated time panning in [P5], it was found that the resulting cues varied with frequency, and different cues suggested different directions for virtual sources. Time panning is not widely used to position sources to desired directions, rather it is used when some special effects are created.

3.3 Time and amplitude panning

A method has been proposed in which the listening room is thought to be inside a bigger virtual room, and the loudspeakers are considered open windows to the virtual room [44]. A position is selected for the virtual sound source, the delays and amplitude changes from that position to each loudspeaker are calculated and are applied to each loudspeaker signal. The delays are calculated based on propagation delay, and amplitude change on distance attenuation. This method can thus be considered as a hybrid of time and amplitude panning.

The method is unfortunately problematic. If the number of loudspeakers is relatively small and if the virtual sources are far away, the method is equal to time panning, which does not generate stable virtual sources. Also, the precedence effect causes the location of virtual sources to collapse to the nearest loudspeaker of each listener outside the best listening area.

3.4 Wave field synthesis

When the number of loudspeakers is very large, Wave Field Synthesis [45] can be used. It reconstructs a whole sound field to a listening room. Theoretically it is superior as a technique, but unfortunately it is impractical in most situations. The most restricting boundary condition is

that the system produces the sound field accurately only if the loudspeakers are at a distance of maximally a half wavelength from each other. The centroids of loudspeakers should thus be a few centimeters from each other to be able to produce high frequencies correctly also, which cannot be achieved without a very large number of loudspeakers.

If the loudspeakers are positioned in a horizontal array, the wave field is then produced in the horizontal plane. Such systems have been constructed using roughly 100 loudspeakers. Accurate spatial reproduction is typically limited to about 1000 Hz. This can be regarded as good enough quality, since low-frequency ITD cues which are important cues in spatial hearing, are then produced accurately.

3.5 HRTF processing

3.5.1 Headphone reproduction

In headphone listening a monophonic sound signal can be positioned virtually to any direction, if HRTFs for both ears are available, for a desired virtual source direction [46, 47]. A sound signal is filtered with a digital filter modeling the measured HRTF. The method simulates the ear canal signals that would have been produced if a sound source existed in a desired direction. If a listener moves his/her head during listening, then the movements of his/her should also be taken into account in processing, otherwise the sound stage will be moving along when the listener rotates his/her head that may cause inside-head localization. These methods are not considered in this thesis.

3.5.2 Cross-talk cancelled loudspeaker listening

If, in stereophonic setup, the cross-talk between a loudspeaker and the contralateral ear is cancelled with some method, it is possible to control precisely the sound signal in a listener's ear canals [46, 47, 48]. However, the best listening area (sweet spot) is very small with cross-talk cancelling systems. Cross-talk cancelled reproduction is otherwise similar to HRTF techniques over headphones. These methods are neither considered in this thesis.

3.6 Contribution of this work to spatial sound reproduction

This work was initially started in order to create spatial sound reproduction methods for 3-D loudspeaker arrays. The main innovation is the vector base amplitude panning method that is discussed below. The author has suggested also a method to triangulate the loudspeaker setup and a method to enhance virtual sources.

3.6.1 Vector base amplitude panning

Vector base amplitude panning (VBAP) is a method to calculate gain factors for pair-wise or triplet-wise amplitude panning [P1]. In pair-wise panning it is a vector reformulation of the tangent law (Eq. 3.3). Differing from the tangent law, it can be generalized easily for triplet-wise panning.

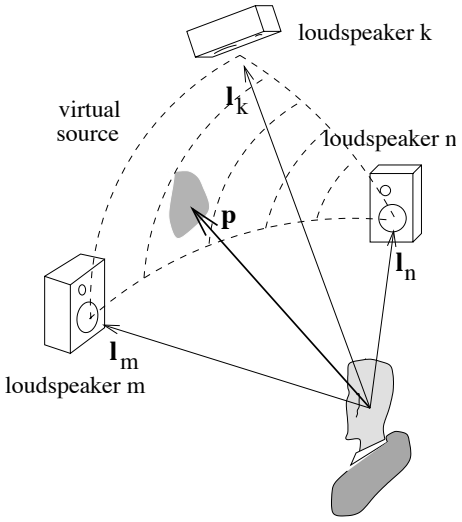

Figure 3.3: A loudspeaker triplet forming a triangle formulated for three-dimensional vector base amplitude panning (VBAP).

In VBAP the listening configuration is formulated with vectors; a Cartesian unit vector $\mathbf{l}_n = [l_{n1}\ l_{n2}\ l_{n3}]^T$ points to the direction of loudspeaker n, from the listening position. In triplet-wise panning unit vectors $\mathbf{l}_n$, $\mathbf{l}_m$, and $\mathbf{l}_k$ then define the directions of loudspeakers n, m, and k, respectively. The panning direction of a virtual source is defined as a 3-D unit vector $\mathbf{p} = [p_n\ p_m\ p_k]^T$. A sample configuration is presented in Fig. 3.3.

The panning direction vector $\mathbf{p}$ is expressed as a linear combination of three loudspeaker vectors $\mathbf{l}_n, \mathbf{l}_m$, and $\mathbf{l}_k$, and in matrix form:

$$\mathbf{p} = g_n\mathbf{l}_n + g_m\mathbf{l}_m + g_k\mathbf{l}_k, \tag{3.9}$$

$$\mathbf{p}^T = \mathbf{g}\mathbf{L}_{nmk}. \tag{3.10}$$

Here g_n, g_m and g_k are gain factors, $\mathbf{g} = [g_n\ g_m\ g_k]$ and $\mathbf{L}_{nmk} = [\mathbf{l}_n\ \mathbf{l}_m\ \mathbf{l}_k]^T$. Vector $\mathbf{g}$ can be solved

$$\mathbf{g} = \mathbf{p}^T\mathbf{L}_{nmk}^{-1} = [p_n\ p_m\ p_k] \begin{bmatrix} l_{n1} & l_{n2} & l_{n3} \\ l_{m1} & l_{m2} & l_{m3} \\ l_{k1} & l_{k2} & l_{k3} \end{bmatrix}^{-1}, \tag{3.11}$$

if $\mathbf{L}_{nmk}^{-1}$ exists, which is true if the vector base defined by $\mathbf{L}_{nmk}$ spans a 3-D space. Eq. 3.11 calculates barycentric coordinates of vector $\mathbf{p}$ in a vector base defined by $\mathbf{L}_{nmk}$. The components of vector $\mathbf{g}$ can be used as gain factors; a scaling of them may be desired according to Eq. 3.5. Barycentric coordinates were originally proposed by Moebius [49], they are used in various fields, e.g. in algebraic topology [50] and in computational geometry [51].

3.6.2 Triangulation of loudspeaker setup

When the number of loudspeakers is greater than three, the panning is performed triplet-wise. Such a system is shown in Fig. 3.4. The loudspeaker triangles that can be used with VBAP have to be specified before using it. In first implementations of VBAP this was conducted manually.

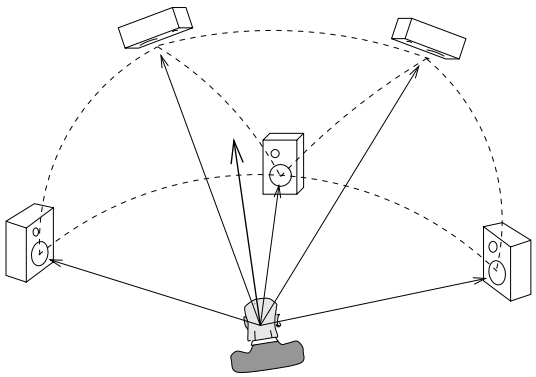

Figure 3.4: Triplet-wise amplitude panning with five loudspeakers.

An automated method was formulated in [P2]. It forms a set of loudspeaker triplets that meet the following conditions that are thought to be beneficial:

1 the loudspeakers of a triplet are not all in the same plane with the listener

2 triangles are not overlapping

3 triangles have as short sides as possible (however, not proven)

The first condition is stated because the gain factors can not be calculated with 3-D VBAP if the loudspeakers are in same plane with the listener. The vector base does not span a 3-D space in that case. In the second condition the overlapping of triangles is prohibited. If moving virtual sources are applied, this is of vital importance. When a virtual source crosses a side of the triangle, the triplet is changed. When triangles do not overlap, the triangle to which the virtual source enters shares the same side, therefore the louspeakers that produce mostly the sound are not changed. A rapid transition of gains could happen if the triangles were overlapping.

The third condition states that the lengths of the triangle sides are minimized. It is widely known that when loudspeakers are farther away from each other, virtual source quality degrades. By minimizing the length of the triangle sides, therefore, as good virtual sources as possible with current loudspeaker setup are produced.

This triangulation method is almost equivalent with an existing triangulation method, the greedy triangulation [52]. The only difference is that in greedy triangulation the first condition does not necessarily hold. The parts of the presented method that implement conditions 2 and 3 are equivalent with corresponding parts of greedy triangulation. It has been proved for the greedy triangulation that the sides of the triangles are as short as possible [52].

3.6.3 Enhancing virtual sources

A drawback of pair- and triplet-wise panning is reported in section 3.1.2 in which the spread of a virtual source is dependent on panning direction due to different numbers of loudspeakers producing the same signal. This can be avoided by using multiple-direction amplitude panning (MDAP), proposed by the author in [P3]. In it gain factors are calculated for multiple panning directions around the desired panning direction. The gain factors of each loudspeaker are summed up and normalized with Eq. 3.5.

With a proper set of panning directions it is possible to create virtual sources the quality of which is not dependent on panning direction and the spreading of virtual sources is then homogeneous. Furthermore, the virtual signal is still not applied to all loudspeakers, but a subset of them. The directional quality therefore does not degrade as much as it degrades with systems that apply a same sound signal to all loudspeakers.

Chapter 4

Methodology of spatial sound evaluation

Subjective and objective evaluation of spatial sound has been studied recently in increasing detail. The perceptual attribute palette for the evaluation of spatial sound reproduction systems has not been formed yet, but there has been a lot of research towards it [53, 54]. Such attributes might include, e.g., envelopment, naturalness, sense of space, directional quality, timbre etc. However, in the present study only directional quality, which is defined as how well a targeted spatial distribution is produced to a virtual source, is investigated. Typically it is most difficult to produce point-like virtual sources, which is investigated in this thesis.

The need for point-like virtual sources can be argued. According to Huyghen's principle, any sound fields can be produced if large numbers of small-sized loudspeakers are mounted on the walls of a listening room. We may assume that any sound field perception can be created for a listener if point-like virtual sources can be formed in any direction. This high quality can not be reached easily, but it may stand as an ultimate goal.

Subjective and objective questions of virtual source directional quality evaluation and measurement are considered in this chapter. In subjective tests different ways to conduct listening tests to explore virtual source quality are reviewed, and in objective tests different methods to measure objectively the quality are considered. The literature on subjective and objective evaluation of spatial sound is reviewed, and the approach used in this work is proposed.

4.1 Subjective evaluation of virtual source direction

In listening tests the experimenter cannot naturally have direct access to the subject's perceptions. The subject should be able to transmit some attributes (in this study the perceived direction) of auditory objects to the experimenter. A test subject can be thought of as consisting of two processes, a part including the mechanisms for decoding and perception for the auditory object (perception segment) and another part for describing the auditory object (description segment) [2].

A perceived direction can be described by different methods. In verbal description subjects report the perceived direction in directional angles [5] or with some other coordinate system. The sound direction can also be described by pointing towards an auditory object with eyes [55], nose [16], or with any pointing device. A problem with description methods is that the mechanisms involved

easily generate errors in the results.

In some cases method-of-adjustment (MOA) [56] can be applied. In it a subject adjusts an attribute of an auditory object to match as well as possible the same attribute of a reference auditory object. In this method, perceptions need not be described, and there may be less error in results. Sandel et al. [57] and Theile [58] performed tests in which the test subject moved a real source to the same direction as an amplitude-panned virtual source.

The direction of a virtual source can also be studied by asking a subject to compare the direction of a virtual source with the direction of a real source. For example, the subjects can be asked if the virtual source is to the left or right of a real source, as was done in [30].

4.2 Objective evaluation of virtual source direction

Conducting listening tests is time-consuming and expensive. A way to avoid them would be to form a model of directional hearing that predicts what the subjects would perceive in listening tests. The results of listening tests would also be more understandable if the performance of test attendees could be explained with a model.

4.2.1 Shadowless head model with ITD-based approach

In early studies the localization of the θ_{cc} direction of amplitude-panned virtual sources was often estimated based on low-frequency ITD, an important cue in sound localization. In the simplest model the listener is approximated to two spatially separated ears with no acoustic shadow from the head [59]. The model is valid at frequencies roughly below 500 Hz. ILD or high-frequency ITD cannot be estimated with this model. Using this model and phasor analysis [28] some stereophonic phenomena have been partly explained.

In [29] the shadowless model was improved slightly. The propagation path from a sound source was not assumed to go through the head. The sound is assumed to propagate via the shortest path between sound source around the head to the ear. Although this approximation is better than simple shadowless stereo for ITD estimation, it can be questioned if the complex wave acoustics near the head can be approximated with a single wave propagation path.

4.2.2 Simple head models

Cooper [42] used a spherical model of the head to study phase and group delay differences as well as level differences between sound signals appearing at ear positions. He claimed that the approximation is valid below 3.13 kHz. A BEM (boundary-element method) model of the head was used in [60] in a similar approach. The spherical model was slightly improved by elongating the sphere in front-rear axis to simulate the real shape better.

A problem with these models is that the acoustics of the pinna cannot be simulated with simple models. The acoustic effect of the pinna appears at frequencies higher than 4 kHz, and therefore the use of simple models is restricted to the low-frequency region.

4.2.3 Evaluation of generated wave field

Large area

A wave field created by a spatial sound system is monitored and compared with the wave field that was targeted to construct [45, 61]. If the wave fields are equal in a large listening area, there is no doubt that the directional quality of virtual sources is brilliant. However, in real life there are always some deviations from perfect reconstruction of a wave field. What is the perceived quality of spatial sound when some degradations from target sound field occur remains a question that cannot be answered by comparing the produced and targeted sound fields directly.

Single point

The wave field in a single point is often used in analysis of the Ambisonics [62]. The point in which the wave field is monitored is the free field point which corresponds to the centroid of a line connecting a subject's ears. Two vectors are calculated, one of which is considered to estimate ITD and be prominent at frequencies below 700 Hz, and the other is considered to estimate ILD and to be prominent at frequencies of 500 – 5000 Hz.

The above model is able to predict some phenomena in perception of spatial audio. However, the validity of this method can be criticized in general. It is not evident that binaural localization can be approximated with the wave field in one point. The facts that a human has a head and that he/she perceives spatial sound based on sound pressures occurring in the ear canals on different sides of the head are omitted. The model does not explain why virtual sources cannot be positioned between loudspeakers in azimuth angles 60° and 120° as found in [58]. Also, the division of frequency scale to two parts is far too rough an approximation and does not explain the frequency-dependent behaviour of perception of virtual sources in [P6]. In [63] it is stated that the model would also be valid with 3-D loudspeaker setups. However, in [P7] the present author claims that the ϕ_{cc} direction perception of a virtual source in such setups is individual and cannot be predicted.

4.2.4 Binaural auditory models

Modern knowledge of human auditory localization mechanisms and rapid development of computing facilities have made it possible to simulate directional perception relatively accurately. Sound pressure in ear canals can be recorded with a dummy or a real head or it can be simulated with measured HRTFs. The binaural cues are then computed with an auditory model from ear canal signals.

In [64] a simplified binaural model was used for this purpose. The ear canal signals were captured for tested spatial sound systems using a KEMAR dummy head. A single value for ITD was calculated using inter-aural cross correlation (IACC). The ILD was calculated simply as a power spectrum ratio. A more sophisticated model is presented in [65]. The model consisted of ear signal simulation with measured HRTFs, basilar membrane modeling with 16 filters, hair cell modeling with half-wave rectification as well as high-frequency smoothing, precedence effect modeling, ITD modeling with IACC, and ILD modeling. The higher level brain processing that forms a direction percept from decoded cues (high-level perceptual stages) are modeled with a database search. The database is constructed of cues of real sources in different directions. The best match between virtual source cue value and real source cue values is found at each frequency band, and

the direction of the best matched real source is considered as the direction that would be perceived.

4.3 Method to evaluate spatial sound used in this study

In this study both subjective and objective tests were conducted. In subjective tests the attendees adjusted the perceived direction of an amplitude-panned virtual source to match best with the perceived direction of a real source [P6, P7]. A listener was thus comparing auditory objects by matching one to another. As a result a set of control parameters was achieved that produced virtual sources the same direction with real sources.

In objective analysis a binaural auditory model was used to explain the results of listening tests and to measure the directional quality of virtual sources. The model is most evolved in [P6] and [P7], while in [P5] there are slightly fewer details. The ear canal signals were simulated with measured HRTFs, up to 20 individual HRTF sets were used. The middle ear was modeled with a filter that approximates a response function derived from the minimum audible pressure curve. The frequency resolution of the cochlea was simulated with a gammatone filter bank with 42 frequency bands. Hair cells were modeled by half-wave rectification and low-pass filtering. ITD was calculated with IACC, and ILD as a loudness level difference between ears in corresponding frequency bands. The cue values were translated with a database search to θ_{cc} angles that they suggested, and the final values were called the ITD angle (ITDA) and the ILD angle (ILDA).

The virtual sources that the listeners favored most in the listening tests were analyzed with the described auditory model. In most cases the performance of subjects could be explained with simulated ILDA and ITDA values [P6] and [P7].

Chapter 5

Directional quality of amplitude-panned virtual sources

5.1 Early studies with stereophonic listening

In standard stereophonic listening the amplitude-panned virtual source is localized fairly consistently to the same direction at most frequencies. The mechanisms producing localization of a virtual source have been explained only at low frequencies. The reasons why high-frequency virtual sources are localized consistently with low-frequency virtual sources have remained unexplained. Also the dependence of direction perception on frequency and temporal structure has not been studied.

Sandel et al. [57] performed tests in which the test subject moved a real source to the same direction as an amplitude-panned virtual source. The sound signal was a sinusoid at eight frequencies. The time and amplitude differences between a subject's ears were calculated analytically. The results were analyzed by recording the sound signal in the ear canals for real and virtual sources and comparing the amplitudes. The test results could be analyzed at frequencies below 1500 Hz using calculated interaural time differences of virtual sources. Some individual frequency-dependent performance was observed at higher frequencies, which remained unexplained.

Leakey [30] conducted listening tests in standard stereophonic listening with broad-band and narrow-band speech. A perceived virtual source direction was compared with perceived direction a real source, and the listener reported if the real source was to the right or left from the virtual source. The test was repeated for real sources in multiple directions. In the results all frequencies were localized consistently. He suggested that the virtual source formation at high frequencies was also based on interaural time difference cues. He used a simple shadowless head model in the analysis, which fails to estimate ITD at high frequencies, and does not estimate ILD at all.

A set of similar studies are reviewed in [2]. Frequency-dependence can be observed in the results. Also, the effect of different temporal structures is present. These phenomena are not explained.

In [58] the localization of virtual sources was studied with listening tests when the stereophonic pair was moved to the side of subjects. It was found that the direction of a virtual source changed inside a loudspeaker pair towards frontal direction when the pair was more on the side of a listener. When the centroid of a pair was in a lateral direction, it was not at all possible to create stable

virtual sources between loudspeakers.

Most evolved objective studies were presented in [42, 60] with spherical model or rough BEM model of the head as discussed in Sec. 4.2.2. In both studies level differences and time and group differences were monitored for an amplitude-panned virtual source. The values were not discussed or compared with listening test results, and no conclusions were made of virtual source quality. However, computed group differences and level differences seem to match at least qualitatively at results obtained in [P6, P7] with frequencies below 3 kHz.

5.2 Evaluations conducted in this work

5.2.1 Stereophonic panning

In [P6] the localization of virtual sources in stereophonic listening was revisited. The directions were evaluated with methods presented in chapter 4.3. In earlier studies it had been proposed that only the low-frequency ITD cue is relied upon when localizing the amplitude-panned virtual sources. In this study it was found that high-frequency ILD is also relatively consistent with low-frequency ITD cues. This answers the question why high frequencies are localized consistently with low frequencies. However, there are some deviations from this. Large discrepancies were found between ILD and ITD at frequencies near 1.7 kHz. Also neither of the cues is consistent with low-frequency ITD at that frequency band. A broad-band virtual source may therefore be spread out due to discrepant cues at this frequency band.

Another interesting phenomenon was also found in these tests. At frequencies near 1.7 kHz, where the cues were discrepant, the attendees relied more on ITD when listening to click trains and relied more on ILD when listening to pink noise. The temporal structure of signals therefore changed the prominence of cues, which explains why virtual sources with different types of signals are localized differently.

5.2.2 Three-dimensional panning

In [P7] a more generic view to directional quality of amplitude-panned virtual sources was searched for. In earlier studies only virtual sources in the horizontal plane had been studied, while in this study arbitrary setups are investigated. Listening tests were conducted with a loudspeaker pair in the median plane, and with two loudspeaker triangles in front of the listener. In the tests the loudspeakers were placed within angle distance of approximately $60\,^\circ$ from each other. The results were interpreted with the auditory model. Furthermore, the auditory model was used to simulate virtual sources in various two- and three-loudspeaker configurations that were not applied in listening tests. The results are reported for θ_{cc} and ϕ_{cc} directions separately.

Perceived θ_{cc} direction

Some features were found on perceived θ_{cc} direction of amplitude-panned virtual sources of arbitrary loudspeaker pairs or triplets. When a pair or triplet is near the median plane the directional quality is relatively good: ITD cues follow quite precisely the values that VBAP proposes. High-frequency ILD is also relatively consistent with low-frequency ITD cues, although some discrep-

ancies in cue values are present. This is not a surprising result, since similar cues were achieved in stereophonic panning. However, it is a new result that the same features are present with pairs that have loudspeakers in different ϕ_{cc} directions, and with loudspeaker triangles. It can thus be concluded that near the median plane a θ_{cc} panning angle of 3-D VBAP describes relatively well the perceived θ_{cc} direction of a virtual source.

When a loudspeaker pair or triplet is moved farther away from the median plane, some artefacts emerge. The ILD cue may be distorted, and it does not generally coincide with the ITD cue. The low-frequency ITD cue behaves more consistently with panning direction than ILD cue. However, its value is biased from the panning direction value towards the median plane. If the centroid of a loudspeaker set is near lateral direction ($\pm 90°, 0°$), there are regions between loudspeakers where a virtual source cannot be positioned. The panning angle of VBAP predicts therefore worse the virtual source location in lateral directions than near the median plane. Fortunately, however, the accuracy of directional hearing is degraded in lateral directions anyhow [2].

Perceived ϕ_{cc} direction

The question of how well the ϕ_{cc} panning direction describes the perceived ϕ_{cc} direction had not been addressed before. In this study it was found that the listening test attendees could control the perceived ϕ_{cc} direction of a virtual source with ϕ_{cc} panning direction if the sound signal included frequencies near 6 kHz. A subject adjusted the ϕ_{cc} panning angle to similar values on different trials, but unfortunately the results varied prominently between individuals. From this information we may conclude that the perceived ϕ_{cc} direction cannot be predicted with any generic panning law.

5.2.3 Comparison of panning laws

In Sec. 3.1.1 different panning laws were proposed for stereophonic listening. The differences of the three panning laws are compared now. To investigate how virtual sources are localized with different gain factors, 16 different gain factor pairs were chosen and virtual sources were simulated using the auditory model presented in [P6] with 6 individual HRTF sets. The level ratio g_1/g_2 varied from 0 dB to ∞ dB, the virtual source should therefore appear at 16 different angles between $0°$ and $30°$. The ITDA and ILDA values were calculated with these gain factor pairs and a mean value and standard deviation were computed from the cues. The ITD cue is distorted at high frequencies, and therefore the values between 200 and 1000 Hz only were taken into account. The cue values are plotted in Fig. 5.1.

The panning directions that correspond to each gain factor pair were calculated with the sine law (Eq. 3.2), the tangent law (Eq. 3.3) and with Chowning's law (Eq. 3.6) for each gain factor pair. The panning direction should match with ITDA, ILDA, or an average of the cues. The cues and panning directions of different laws are plotted in Fig. 5.1. The tangent law panning angle matches best the ITDA when compared to other laws. Its values match with ITDA values with all gain factor pairs, while the sine law predicts slightly smaller values. The Chowning law prediction differs from other laws, it predicts prominently higher values. It seems to match well with the mean of ILDA cue. However, the standard deviation of ILDA is so large that the predictions of other laws are also inside it.

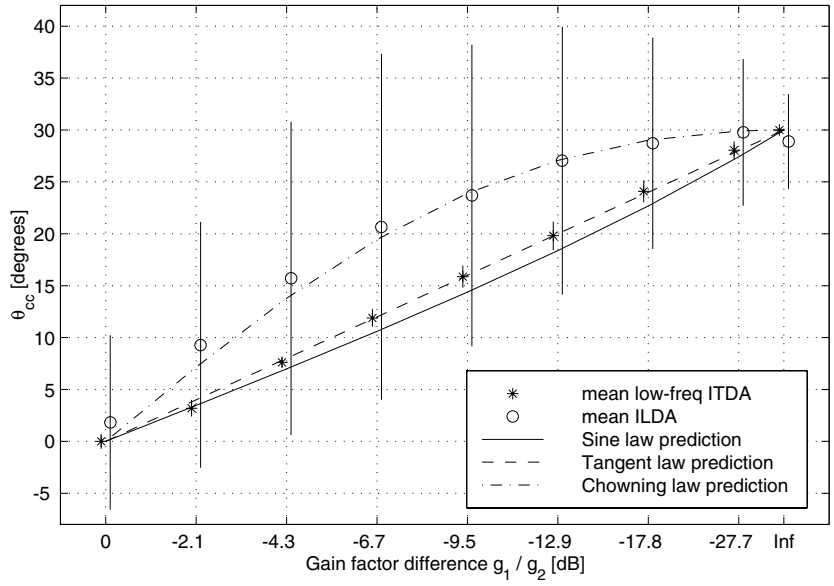

Figure 5.1: Auditory cues (ITDA and ILDA) and virtual source direction predictions as function of gain difference of loudspeakers in stereophonic configuration. Simulation was conducted with six individual HRTF sets to both sides of the listener. The whiskers denote $\pm$ standard deviation of ITDA or ILDA values.

5.3 Discussing multi-channel loudspeaker layouts

This study has been focusing to study the localization of virtual sources in different loudspeaker layouts. It is of interest to discuss with the results achieved in this study and earlier how the loudspeakers should be arranged to produce a desired virtual source localization quality using pair- or twiplet-wise panning. Typically the loudspeakers are positioned in the horizontal plane, although some setups are used in which also elevated loudspeakers exist. These cases will be discussed separately.

We may list some phenomena that affect choosing the loudspeaker directions in horizontal setups. The perception of θ_{cc} direction of an amplitude-panned virtual source is similar between individuals, and can be predicted relatively accurately in the best listening area. The panning angle is a fairly good estimate of perceived direction. Outside the best listening area the virtual source is localized to the nearmost loudspeaker producing the virtual source due to the precedence effect [2]. Furthermore, it is known that when the aperture between loudspeakers is wider, the localization accuracy is worse [59, 29]. It is also known that there should be loudspeakers at lateral directions to be able to create virtual sources there [58].

With louspeaker setups in which also elevated loudspeakers exist, the perception of θ_{cc} direction is similar as with horizontal setups. The perception of ϕ_{cc} is more problematic. In this study it was shown that the listeners perceived the ϕ_{cc} of a virtual source individually. There is thus no panning law that could be used in positioning of virtual sources in ϕ_{cc} direction. However, in many cases the perceived ϕ_{cc} direction coincides with ϕ_{cc} direction of one of the loudspeakers. Thus the error in virtual source ϕ_{cc} direction perception is lower if the loudspeakers that produce a virtual source are more close to each other. A good quality can be achieved if the number of loudspeakers is increased and triplet-wise panning is used. A single triplet can be made so small that a sufficiently low error of perceived direction can be achieved. It must be also taken into account that the ϕ_{cc}

direction perception ability of humans is worse than with θ_{cc} direction [2]. Thus larger errors may be tolerated. By taking these phenomena into account, the localization quality may be optimized in some directions, or an even localization quality may be created to all directions.

Some existing horizontal multi-channel loudspeaker layouts are now discussed. Quadraphonic setup has no loudspeakers in lateral directions, it fails to create stable images there, and the loudspeaker aperture of $90°$ may be too large to produce stable images in front or back of the listener. In the 5.1 loudspeaker setup [35] the rear loudspeakers are in directions $\pm 110°$, which provides better imaging for lateral directions. The frontal loudspeakers are in directions $\pm 30°$ and $0°$, that provides good quality for virtual sources in front of the listener. This is optimized for motion pictures, a good localization is desired in the direction of screen. However, the localization behind the listener may be unstable, since the $140°$ loudspeaker aperture is too large for stable virtual source creation.

The author has used in demonstrations in ICAD98 and AES 16th Int. Conf. a setup in which there was eight loudspeakers used for a 3-D loudspeaker setup. Three loudspeakers were positioned to elevated directions as will be explained later. There were thus five loudspeakers to be placed in the horizontal plane. Two of them were positioned to lateral directions $\pm 90°$, as requested by Theile's experiment [58]. Two of the remaining three loudspeakers were placed in front of the listener $(\pm 30°)$, evenly between the lateral loudspeakers, and one back, to $180°$. The frontal horizontal plane was favored, since perceptual quality was found to be less important in backward direction than in frontal direction. A similar setup has been used by Holman, as reported in [66]. If the number of loudspeakers is increased, the span of loudspeaker pairs decreases, which makes directional errors smaller in and out of the best listening area. Setups of six or eight loudspeakers evenly around the listener has been used in various setups. With eight loudspeakers the loudspeaker span is only $45°$, that produces a relatively good virtual source quality for a large listening area.

The author tested informally some different loudspeaker setups with elevated loudspeakers. If only one elevated loudspeaker $(0°, 90°)$ was added to a horizontal setup, the author perceived the virtual sources mostly to the horizontal plane. If the panning direction was near the zenith, the virtual source jumped abruptly to the elevated loudspeaker. Two elevated loudspeakers provided slightly better result, but there seemed to be many directions where virtual sources could not be positioned. Three elevated loudspeakers seemed to provide relatively good localization quality in all elevated directions. The demonstrations in ICAD98 and AES 16th Int. Conf. were performed with three elevated loudspeakers, in (θ, ϕ) directions $(\pm 45°, 45°)$ and $(180°, 45°)$. In AES 16th Conf. there was an additional demo in which three loudspeakers were added to lowered directions $(\pm 45°, -45°)$ and $(180°, -45°)$. This enabled creation of virtual sources also below the listener.

Chapter 6

Conclusions and future directions

In this work a virtual source positioning method for arbitrary multi-loudspeaker setups was presented and the directional qualities of virtual sources created with it were investigated.

Different topics concerning virtual source positioning have been presented in publications [P1-P4]. The method to position virtual sources was introduced in [P1], and it was named the vector base amplitude panning (VBAP). It is an amplitude panning method to calculate gain factors for loudspeaker pairs or for loudspeaker triplets forming triangles from the listener's view point. VBAP can be used with any number of loudspeakers in any positioning. The usability of the method was improved with an automatic triangularization method for 3-D loudspeaker setups proposed in [P2], and a method to control the spreading of virtual sources was suggested in [P3]. Implementation of these methods was presented in [P4].

The perception of virtual sources generated with presented methods was considered in publications [P5-P7]. Simulation of perception of virtual sources was carried out in [P5] using a binaural auditory model adapted from the literature. In [P6-P7] the directional qualities of amplitude-panned sources with different loudspeaker setups were investigated by listening tests. The listening test results were interpreted using an auditory model with quite good accuracy. Therefore the directional quality of virtual sources was also estimated using the binaural auditory model for setups in which listening tests were not conducted.

As a result it was found that VBAP predicts the θ_{cc} direction of a virtual source quite accurately, when a loudspeaker set is near the median plane. When the set is moved towards a lateral direction, the perceived direction is biased towards the median plane. It was also found that the perceived ϕ_{cc} direction is individual to each subject and cannot be predicted with any panning law.

The directional quality of amplitude-panned virtual sources has been explored relatively carefully in this study. However, this work opens some directions for further studies. Directional quality is only one attribute of amplitude-panned virtual sources. The timbral effects in amplitude panning are left as a future task. Nor was amplitude panning studied in relation to higher level perceptual attributes of spatial sound, such as envelopment and sense of space.

When a loudspeaker set is moved towards the side of a listener, the ITDA cue of an amplitude-panned virtual source is biased towards the median plane, compared to the panning direction of VBAP. The quality of amplitude-panned virtual sources could be enhanced quite easily by modifying the panning law, dependent on θ_{cc} directions of loudspeakers, to reduce the bias of ITDA cues.

A more complicated topic would be to enhance the amplitude panning to be dependent on frequency. The gain factors would then be no more scalars, but they would vary with frequency. This can be implemented using digital filters. The inaccurate behaviour of amplitude-panned virtual source cues could then be equalized to some degree. Preliminary results have been achieved and reported in [67]. However, this approach is problematic since it may make the quality of virtual sources more dependent on the listening position. Also, the individuality of HRTFs should be taken into account in such enhancement.

The binaural auditory model used in this study has proven to be a reliable tool in the analysis of perceived direction of virtual sources. It could also be used in the analysis of other spatial sound systems. A question also remains if there is a relation from directional quality of virtual sources to higher level perceptual attributes.

In [P7] a hypothesis of a cue for ϕ_{cc} direction perception was drawn, which should be tested by further investigations.

List of Publications

[P1] V. Pulkki. “Virtual sound source positioning using vector base amplitude panning.” *Journal of the Audio Engineering Society*, 45(6) pp. 456-466, June 1997.

[P2] V. Pulkki and T. Lokki. “Creating auditory displays to multiple loudspeakers using VBAP: A case study with DIVA project”. In *International Conference on Auditory Display*, Glasgow, Scotland, 1998.

[P3] V. Pulkki. “Uniform spreading of amplitude panned virtual sources,” *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. Mohonk Mountain House, New Paltz, New York.

[P4] V. Pulkki. “Generic panning tools for MAX/MSP”. *Proceedings of International Computer Music Conference 2000*. pp. 304-307. Berlin, Germany, August, 2000.

[P5] 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) pp. 203-217 April 1999.

[P6] V. Pulkki and M. Karjalainen. “Localization of amplitude-panned virtual sources I: Stereophonic panning.” Accepted to *Journal of the Audio Engineering Society*.

[P7] V. Pulkki. “Localization of amplitude-panned virtual sources II: Two- and Three-dimensional panning.” Accepted to *Journal of the Audio Engineering Society*

Summary of articles and author's contribution

Publication [P1]

This article introduces the vector base amplitude panning (VBAP) method, which is a method to position virtual sources in arbitrary loudspeaker setups. VBAP is based on pair-wise panning in 2-D loudspeaker setups and on triplet-wise panning in 3-D setups. The panning direction vector is expressed as a linear weighted sum of loudspeaker direction vectors of a pair or triplet. The weights can be used as gain factors of loudspeaker signals after scaling. If VBAP is applied to pair-wise panning, the calculated gain factors follow an existing panning law, the tangent law. VBAP is the first triplet-wise panning method that can be used in arbitrary 3-D loudspeaker setups. Previous panning methods for 3-D sound systems either used all loudspeakers to produce virtual sources, which results in some artefacts, or they used loudspeaker triangles with a non-generalizable 2-D user interface.

Publication [P2]

VBAP is integrated as part of a large system creating digital interactive virtual acoustics (DIVA). Direct and reflected sound signals are simulated by a model of listening space, and are positioned using VBAP. A method to triangularize a 3-D loudspeaker setup is introduced. The author's contribution consists of proposing the triangularization method and implementing the VBAP method to the DIVA environment in co-operation with Mr. Lokki.

Publication [P3]

An idea is introduced on how pair- or triplet-wise panning can be enhanced in such a way that the spreading of virtual sources is more independent of panning direction. The introduced method is called multiple direction amplitude panning (MDAP). The gain factors are calculated for multiple directions around the panning direction and the resulting factors are summed up and normalized. This yields that the sound signal is never applied to a single loudspeaker, which spreads out the virtual sources also at the directions of loudspeakers. The spread of virtual sources is analyzed with an auditory model and it is found that the spread can be made independent of panning direction with a sufficiently large set of panning directions.

Publication [P4]

Implementations of VBAP, the triangularization method and MDAP are described in Max/MSP synthesis software. C-programming language sources and executables for Max/MSP have been made freely available. In the implementation current loudspeaker configuration is specified first, and the gain factors for different virtual sources are calculated based on that information. A specific 3-D sound stage can be produced as similar as possible in different arbitrary loudspeaker systems, which has not been possible earlier. A similar implementation of VBAP has been completed in the Institut de Reserche et Coordination Acoustique/Musique (IRCAM) in France to their

spatialization software Spat-1.2.2c, May 29, 2000. However, the implementation lacks automatic triangularization and spreading control of virtual sources.

Publication [P5]

A method to evaluate virtual sources with an auditory model is presented. The method includes simulation of ear canal signals in anechoic listening using measured head-related transfer functions (HRTFs) of listeners and simulation of decoded binaural auditory cues. The cues of real and virtual sources are compared, and the qualities of virtual sources are discussed. The author's contribution included adapting the auditory model from literature under Prof. Matti Karjalainen's supervision and performing all virtual source simulations except for the HRTF-processed virtual sources, which were conducted in co-operation with Dr. Jyri Huopaniemi.

Publication [P6]

Localization of amplitude-panned virtual sources is studied with listening tests in which the subjects adjusted the panning angle to match the perceived direction of a virtual source with a perceived direction of a reference real source. The tests were conducted with narrow- and broad-band pink noise and narrow-band impulse trains. It was found that the resulting panning angles are dependent on frequency and signal type. The auditory model is used to monitor the auditory cues of subjects. The phenomena existing in listening test data were explained with model-based analysis. The most consistent cues of virtual sources in stereophonic listening are low-frequency ITD cues and also, to some degree, high-frequency ILD cues. ITD cues are corrupted at frequencies above roughly 1.5 kHz, and ILD cues are unstable and suggest directions outside the loudspeaker span near frequencies 700 Hz – 2 kHz. It is also shown that the panning angle computed from the tangent law corresponds accurately to the direction that low-frequency ITD values suggest.

These results bring some new information on the localization of amplitude-panned virtual sources in stereophonic listening when compared with earlier studies. The tests were conducted with a more extensive set of stimulus signals, which revealed new phenomena in virtual source localization. The auditory model used in the study was more detailed and realistic than in the previous studies, which yielded a good match between listening test results and simulation results. The work was conducted by the author, except that Prof. Karjalainen helped to design the listening tests.

Publication [P7]

Listening tests are reported in which subjects adjusted panning angles to match the perceived virtual source direction with a perceived direction of a reference real source for a loudspeaker pair in the median plane and with two different loudspeaker triangles in front of the subject. In the results it was found that the subjects adjusted θ_{cc} (angle between a position vector and the median plane) panning direction consistently with each other, and that they adjusted ϕ_{cc} (direction angle inside a cone of confusion) individually but consistently with themselves. The performance of listeners in θ_{cc} panning angle adjustment was explained with the auditory model, but the performance of ϕ_{cc} panning angle judgments did not seem to match current theories of ϕ_{cc} direction decoding. A new type of cue was hypothesized based on decoding the loudness level at modal frequencies of the pinna and comparing it with an estimate of loudness level without the pinna effect.

To investigate the perception of amplitude-panned virtual sources in more general cases, the θ_{cc} localization of virtual sources with arbitrary loudspeaker pairs and triplets was modeled with the auditory model. It is shown that the VBAP panning angle predicts well the low-frequency ITDA and roughly the high-frequency ILDA cues with different loudspeaker pairs or triplets when they

are near the median plane. When a set is moved towards a side of the listener, ITDA is biased towards the median plane when compared with panning direction. Perception of the ϕ_{cc} direction cannot be predicted with any panning law, which is concluded from highly individual responses in the median plane listening tests.

Bibliography

[1] A. D. Blumlein. U.K. Patent 394,325, 1931. Reprinted in Stereophonic Techniques, Audio Eng. Soc., NY, 1986.

[2] J. Blauert, *Spatial Hearing, Revised edition.* Cambridge, MA, USA: The MIT Press, 1997.

[3] M. Barron, *Auditorium Acoustics and Architectural Design*. E & FN Spon, 1993.

[4] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1465–1479, 1999.

[5] F. L. Wightman and D. J. Kistler, "Factors affecting the relative salience of sound localization cues," in *Binaural and Spatial Hearing in Real and Virtual Environments* (R. H. Gilkey and T. R. Anderson, eds.), Mahwah, NJ, YSA: Lawrence Erlbaum Assoc., 1997.

[6] R. O. Duda, "Elevation dependence of the interaural transfer function," in *Binaural and Spatial Hearing in Real and Virtual Environments* (R. H. Gilkey and T. R. Anderson, eds.), pp. 49–75, Mahwah, New Jersey: Lawrence Erlbaum Associates, 1997.

[7] J. O. Pickles, *An Introduction to the physiology of hearing*. Academic Press, 1988.

[8] B. C. J. Moore, *An introduction to the psychology of hearing*. San Diego: Academic Press, fourth ed., 1997.

[9] R. Patterson, K. Robinson, J. Holdsworth, D. Mckeown, C. Zhang, and M. H. Allerhand, "Complex sounds and auditory images," in *Auditory Physiology and Perception* (L. D. Y. Cazals and K. Horner, eds.), pp. 429–446, Oxford: Pergamon, 1992.

[10] R. Meddis, "Simulation of auditory-neural transduction: Further studies," *J. Acoust. Soc. Am.*, vol. 83, pp. 1056–1064, 1988.

[11] M. B. Gardner and R. S. Gardner, "Problem of localization in the median plane: effect of pinnae cavity occlusion," *J. Acoust. Soc. Am.*, vol. 53, no. 2, pp. 400–408, 1973.

[12] K. Iida, M. Yairi, and M. Morimoto, "Role of pinna cavities in median plane localization," *J. Acoust. Soc. Am.*, vol. 103, p. 2844, May 1998.

[13] J. Blauert, "Sound localization in the median plane," *Acustica*, vol. 22, pp. 205–213, 1969/70.

[14] P. J. Bloom, "Creating source elevation illusions by spectral manipulation," *J. Audio Eng. Soc.*, vol. 25, no. 9, pp. 560–565, 1977.

[15] A. J. Watkins, "Psychoacoustical aspects of synthesized vertical locale cues," *J. Acoust. Soc. Am.*, vol. 63, pp. 1152–1165, 1978.

[16] J. C. Middlebrooks, "Narrow-band sound localization related to external ear acoustics," *J. Acoust. Soc. Am.*, vol. 92, pp. 2607–2624, November 1992.

[17] P. Zakarauskas and M. S. Cynader, "A computational theory of spectral cue localization," *J. Acoust. Soc. Am.*, vol. 94, pp. 1323–1331, September 1993.

[18] J. Backman and M. Karjalainen, "Modelling of human directional and spatial hearing using neural networks," in *Proc. ICASSP-93*, (Minneapolis), pp. I–125 – I–128, 1993.

[19] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. Heidelberg, Germany: Springer-Verlag, 1990.

[20] B. C. J. Moore, "A model for the prediction of thresholds, loudness, and partial loudness," *J. Audio Eng. Soc.*, vol. 45, no. 4, pp. 224–240, 1997.

[21] R. H. Gilkey and T. R. Anderson, eds., *Binaural and Spatial Hearing in Real and Virtual Environments*. Mahwah, NJ, US: Lawrence Erlbaum Assoc., 1997.

[22] T. C. T. Yin, P. X. Joris, P. H. Smith, and J. C. K. Chan, "Neuronal processing for coding interaural time disparities," in *Binaural and Spatial Hearing in Real and Virtual Environments* (R. H. Gilkey and T. R. Anderson, eds.), pp. 399–425, Mahwah, New Jersey: Lawrence Erlbaum Associates, 1997.

[23] L. A. Jeffress, "A place theory of sound localization," *J. Comp. Physiol. Psych.*, vol. 61, pp. 468–486, 1948.

[24] R. M. Stern and C. Trahiotis, "Models of binaural perception," in *Binaural and Spatial Hearing in Real and Virtual Environments* (R. H. Gilkey and T. R. Anderson, eds.), pp. 499–531, Mahwah, NJ, YSA: Lawrence Erlbaum Assoc., 1997.

[25] P. M. Zurek, "The precedence effect," in *Directional Hearing* (W. A. Yost and G. Gourewitch, eds.), pp. 3–25, Springer-Verlag, 1987.

[26] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Gutman, "The precedence effect," *J. Acoust. Soc. Am.*, vol. 106, pp. 1633–1654, October 1999.

[27] E. M. Wentzel, "Effect of increasing system latency on localization of virtual sounds," in *The proceedings of the AES 16th international conference*, (Rovaniemi, Finland), pp. 42–50, AES, April 1999.

[28] B. B. Bauer, "Phasor analysis of some stereophonic phenomena," *J. Acoust. Soc. Am.*, vol. 33, pp. 1536–1539, November 1961.

[29] J. C. Bennett, K. Barker, and F. O. Edeko, "A new approach to the assessment of stereophonic sound system performance," *J. Audio Eng. Soc.*, vol. 33, pp. 314–321, May 1985.

[30] D. M. Leakey, "Some measurements on the effect of interchannel intensity and time difference in two channel sound systems," *J. Acoust. Soc. Am.*, vol. 31, pp. 977–986, July 1959.

[31] B. Bernfeld, "Attempts for better understanding of the directional stereophonic listening mechanism." 44th Convention of the Audio Engineering Society, Rotterdam, The Netherlands, 1973.

[32] F. R. Moore, *Elements of Computer Music*. Englewood Cliffs, New Jersey 07632: Prentice Hall, 1990.

[33] J. Chowning, "The simulation of moving sound sources," *J. Audio Eng. Soc.*, vol. 19, no. 1, pp. 2–6, 1971.

[34] J.-M. Jot, V. Larcher, and J.-M. Pernaux, "A comparative study of 3d audio encoding and rendering techniques," in *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction*, (Rovaniemi, Finland), AES, April 1999.

[35] I. R. BS.775-1, "Multichannel stereophonic sound system with and without accompanying picture," tech. rep., International Telecommunication Union, Geneva, Switzerland, 1992-1994.

[36] URL: *http://www.lcsaudio.com/*.

[37] S. Ellison personal communication, 1999.

[38] M. A. Gerzon, "Panpot laws for multispeaker stereo," in *The 92nd Convention 1992 March 24-27 Vienna*, Audio Engineering Society, Preprint No. 3309, 1992.

[39] D. G. Malham and A. Myatt, "3-d sound spatialization using ambisonic techniques," *Comp. Music J.*, vol. 19, no. 4, pp. 58–70, 1995.

[40] J. Daniel, J.-B. Rault, and J.-D. Polack, "Ambisonics encoding of other audio formats for multiple listening conditions," in *Proc 105th Audio Eng. Soc. Convention*, 1998. Preprint # 4795.

[41] G. Monro, "In-phase corrections for ambisonics," in *Proc. Int. Computer Music Conf.*, (Berlin, Germany), pp. 292–295, 2000.

[42] D. H. Cooper, "Problems with shadowless stereo theory: Asymptotic spectral status," *J. Audio Eng. Soc.*, vol. 35, pp. 629–642, September 1987.

[43] S. P. Lipshitz, "Stereophonic microphone techniques... are the purists wrong?," *J. Audio Eng. Soc.*, vol. 34, no. 9, pp. 716–744, 1986.

[44] F. R. Moore, "A general model for spatial processing of sounds," *Computer Music J.*, vol. 7, Fall 1983. reprinted in C Roads (ed.), "The music machine", The MIT Press.

[45] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am*, vol. 93, May 1993.

[46] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-related transfer functions of human subjects," *J. of Audio Eng. Soc.*, vol. 43, pp. 300–321, May 1995.

[47] D. R. Begault, *3-D Sound For Virtual Reality and Multimedia*. Cambridge, MA, USA: AP Professional, 1994.

[48] B. Gardner, *3-D Audio Using Loudspeakers*. PhD thesis, Massachusetts Institute of Technology, Massachusetts, USA, 1997.

[49] F. Moebius, *August Ferdinand Moebius, Gesammelte Werke*. Werlag von Suttirzel, 1885.

[50] J. R. Munkres, *Elements of Algebraic Topology*. Addison-Wesley, 1984.

[51] G. Farin, *Computer Aided Geometric Design*. Academic Press, 1984.

[52] F. Preparata and M. Shamos, *Computational Geometry – and introduction*. New York: Springer-Verlag, 1985.

[53] S. Bech, "Methods for subjective evaluation of spatial characteristics of sound," in *Proc. AES 16th int. conf. on Spatial Sound Reproduction*, (Rovaniemi, Finland), pp. 487–504, AES, April 1999.

[54] J. Berg and F. Rumsey, "Spatial attribute identification and scaling by repertory grid technique and other methods," in *Proc. 16th AES int. conf. on Spatial Sound Reproduction*, (Rovaniemi, Finland), pp. 51–77, AES, April 1999.

[55] P. M. Hofman, J. G. A. V. Riswick, and A. J. V. Opstal, "Relearning sound localization with new ears," *Nature neuroscience*, vol. 1, pp. 417–421, September 1998.

[56] B. L. Cardozo, "Adjusting the method of adjustment: SD vs DL," *J. Acoust. Soc. Am.*, vol. 37, pp. 768–792, May 1965.

[57] T. T. Sandel, D. C. Teas, W. E. Feddersen, and L. A. Jeffress, "Localization of sound from single and paired sources," *J. Acoust. Soc. Am.*, vol. 27, pp. 842–852, September 1955.

[58] G. Theile and G. Plenge, "Localization of lateral phantom sources," *J. Audio Eng. Soc.*, vol. 25, pp. 196–200, April 1977.

[59] H. A. M. Clark, G. F. Dutton, and P. B. Vanderlyn, "The 'stereosonic' recording and reproducing system," *J. Audio Eng. Soc,*, vol. 6, no. 2, 1958. Reprinted in Stereophonic Techniques, Audio Eng. Soc., NY, 1986.

[60] K. B. Rasmussen and P. M. Juhl, "The effect of head shape on spectral stereo theory," *J. Audio Eng. Soc.*, vol. 41, pp. 135–142, March 1993.

[61] C. Landone and M. Sandler, "Issues in performance prediction of surround systems in sound reinforement applications," in *Proc. 2nd COST G-6 Workshop on Digital Audio Effects*, (Trondheim, Norway), 1999.

[62] M. A. Gerzon, "Periphony: With-height sound reproduction," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10, 1972.

[63] M. A. Gerzon, "General metatheory of auditory localization." Preprint #3306 of 92nd AES Convention, Vienna, 1992.

[64] C. J. Mac Cabe and D. J. Furlong, "Virtual imaging capabilities of surround sound systems," *J. Audio Eng. Soc.*, vol. 42, pp. 38–49, January - February 1994.

[65] E. A. Macpherson, "A computer model for binaural localization for stereo imaging measurement," *J. Audio Eng. Soc.*, vol. 39, no. 9, pp. 604–622, 1991.

[66] M. Bosi, "High quality multichannel audio coding: Trends and challenges," in *Proc. AES 16th Int. Conf.*, April 1999.

[67] V. Pulkki, M. Karjalainen, and V. Välimäki, "Localization, coloration, and enhancement of amplitude-panned virtual sources," in *The proceedings of the AES 16th international conference*, (Rovaniemi, Finland), pp. 257–278, AES, April 1999.

HELSINKI UNIVERSITY OF TECHNOLOGY
LABORATORY OF ACOUSTICS AND AUDIO SIGNAL PROCESSING

34 V. Välimäki: Fractional Delay Waveguide Modeling of Acoustic Tubes. 1994

35 T. I. Laakso, V. Välimäki, M. Karjalainen, U. K. Laine: Crushing the Delay—Tools for Fractional Delay Filter Design. 1994

36 J. Backman, J. Huopaniemi, M. Rahkila (toim.): Tilakuuleminen ja auralisaatio. Akustiikan seminaari 1995

37 V. Välimäki: Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters. 1995

38 T. Lahti: Akustinen mittaustekniikka. 2. korjattu painos. 1997

39 M. Karjalainen, V. Välimäki (toim.): Akustisten järjestelmien diskreettiaikaiset mallit ja soittimien mallipohjainen äänisynteesi. Äänenkäsittelyn seminaari 1995

40 M. Karjalainen (toim.): Aktiivinen äänenhallinta. Akustiikan seminaari 1996

41 M. Karjalainen (toim.): Digitaaliaudion signaalinkäsittelymenetelmiä. Äänenkäsittelyn seminaari 1996

42 M. Huotilainen, J. Sinkkonen, H. Tiitinen, R. J. Ilmoniemi, E. Pekkonen, L. Parkkonen, R. Näätänen: Intensity Representation in the Human Auditory Cortex. 1997

43 M. Huotilainen: Magnetoencephalography in the Study of Cortical Auditory Processing. 1997

44 M. Karjalainen, J. Backman, L. Savioja (toim.): Akustiikan laskennallinen mallintaminen. Akustiikan seminaari 1997

45 V. Välimäki, M. Karjalainen (toim.): Aktiivisen melunvaimennuksen signaalinkäsittelyalgoritmit. Äänenkäsittelyn seminaari 1997

46 T. Tolonen: Model-Based Analysis and Resynthesis of Acoustic Guitar Tones. 1998

47 H. Järveläinen, M. Karjalainen, P. Maijala, K. Saarinen, J. Tanttari: Työkoneiden ohjaamomelun häiritsevyys ja sen vähentäminen. 1998

HELSINKI UNIVERSITY OF TECHNOLOGY
LABORATORY OF ACOUSTICS AND AUDIO SIGNAL PROCESSING

48 T. Tolonen, V. Välimäki, M. Karjalainen: Evaluation of Modern Sound Synthesis Methods. 1998

49 M. Karjalainen, V. Välimäki (toim.): Äänenlaatu. Akustiikan seminaari 1998

50 V. Välimäki , M. Karjalainen (toim.): Signaalinkäsittely audiotekniikassa, akustiikassa musiikissa. Äänenkäsittelyn seminaari 1998

51 M. Karjalainen: Kommunikaatioakustiikka. 1998

52 M. Karjalainen (toim.): Kuulon mallit ja niiden sovellutukset. Akustiikan seminaari 1999

53 Huopaniemi, Jyri: Virtual Acoustics And 3-D Sound In Multimedia Signal Processing. 1999

54 Bank, Balázs: Physics-Based Sound Synthesis of the Piano. 2000

55 Tolonen, Tero: Object-Based Sound Source Modeling. 2000

56 Hongisto, Valtteri: Airborne Sound Insulation of Wall Structures — Measurement And Prediction Methods. 2000

57 Zacharov, Nick: Perceptual Studies On Spatial Sound Reproduction Systems. 2000

58 Varho, Susanna: New Linear Predictive Methods For Digital Speech Processing. 2001

59 Pulkki, Ville; Karjalainen, Matti: Localization Of Amplitude-Panned Virtual Sources. 2001

60 Härmä, Aki: Linear Predictive Coding With Modified Filter Structures. 2001

61 Härmä, Aki: Frequency-Warped Autoregressive Modeling And Filtering. 2001

ISBN 951-22-5531-6
ISSN 1456-6303