2011

Free field auditory localization and perception

Butcher, Andrew

Lethbridge, Alta. : University of Lethbridge, Dept. of Mathematics and Computer Sciencce, c2011

FREE FIELD AUDITORY LOCALIZATION AND PERCEPTION

ANDREW BUTCHER
Bachelor of Science, University of Alberta, 2004

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Mathematics and Computer Science
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

Abstract

We have designed a system suitable for auditory electroencephalographic (EEG) experiments, with the objective of enabling studies of auditory motion. This thesis details the perceptual cues involved in spatial auditory experiments, and compares a number of spatial panning algorithms while examining their suitability to this purpose. A behavioural experiment involving perception of static auditory objects was used in an attempt to differentiate these panning algorithms. This study was used to inform the panner choice used in an auditory EEG experiment. This auditory EEG experiment involved the effects of discontinuity in velocity and position, and their affects on object perception. A new event related potential (ERP) component – the lateralized object related negativity (LORN) – was identified, and we consider its significance. *libnetstation*, a library for connecting with the *NetStation* (EEG) system has been developed, and released as open source software.

Acknowledgments

I would like to thank my supervisors, Matthew Tata and Stephen Wismath for their guidance during my program. Finishing the thesis would not have been possible without their support. Additionally, I would like to thank my committee members Albert Cross and Howard Cheng for their advice throughout the program. Leah Hackman has been a source of unending encouragement, and I appreciate her waiting patiently as I finished my thesis. I'd like to thank Nolan Bard, Jeff Woodcroft, and Hailey Markowski for being good and patient friends, and Geoff Ryan for providing an endless number of discussions on physics and mathematics. My mother has provided support throughout my education, and I would not have succeeded without her. Finally I would like to thank BioWare for the opportunity to finish my thesis and return to a job when I was finished.

Contents

List of Figures

Chapter 1

Introduction

The human visual system has been carefully studied, leading to a wealth of information and successful models for computer vision. The human auditory system is relatively less understood, but interest in the field is growing rapidly. To this end we have developed an audio system suitable for auditory experiments involving electroencephalography. The system produces sound from an arbitrary number of speakers, and supports a number of spatial panning algorithms. Out-of-the-box spatial panning systems are generally limited to home-theatre layouts, which are not suitable for auditory experiments. Stimulus delivery required accurate timing, and robust, well tested software.

Chapter 2 of this thesis discusses perceptual issues involved in spatial audio reproduction. The primary auditory cues are the interaural time difference (ITD) and interaural level difference (ILD). These perceptual cues have been used as the basis of a variety of panning algorithms, and understanding their properties leads to a deeper understanding of the panning algorithms presented in Chapter 3.

In Chapter 3, two vector base panning techniques, and three ambisonic panning techniques are reviewed. Vector base panning techniques use two (or three) speakers at a time, and can be thought of as variations on “classical” stereo panning, expanded beyond two speakers. Ambisonic techniques are based on the work of Michael Gerzon, and were designed to improve on early attempts at surround sound. A large number of speakers are used to create a stable sound field, relying on interference effects to produce the perception of panned audio. Variants of vector base and ambisonic techniques are discussed, and related to the perceptual cues discussed in Chapter 2.

Chapter 4 reviews the requirements and implementation details of the stimulus presentation system used in our auditory electroencephalographic (EEG) experiments. Stimulus onset must be matched to a time-stamped label in the EEG data stream. The timing issues involved are discussed in depth, along with a description of our implementation and validation procedure. The open source library *libnetstation* was developed to interface with the EEG recording system. This library allows customization of experiments otherwise not allowed by the commercial experimental design software, *E-Prime*.

The implementation of each of the panning techniques was used in a behavioural experiment to both verify functionality, and to inform decisions about panner types in subsequent studies. In chapter 5 we define a perceptual "error metric", and evaluate the panning algorithms based on this metric and the perceived width of the rendered auditory image.

libnetstation and the stimulus presentation software were employed in a study of sound in motion, published as "A Lateralized Auditory Evoked Potential Elicited When Auditory Objects Are Defined By Spatial Motion" in Hearing Research [12]. The lateralized auditory evoked potential, and its relationship to other components of the auditory evoked potential are discussed.

Chapter 2

Perception

2.1 Localization

The perceived direction of a sound source depends on a large number of factors, of which we cover the most pertinent. A more complete review of the psychophysics involved can be found in [9], with computational models in [52]. For a sound located outside of the head, the perception of a sound being at a position in space can be expressed as two independent quantities - the angle relative to the listener and the distance to the listener. Sounds presented over headphones often appear to come from a location inside or slightly behind the head. We refer to the location of such stimuli with the term “lateralization”, reserving localization for sounds perceived outside of the head.

2.1.1 Sound Direction

The earliest investigations into the cues affecting the perceived direction of an auditory event focused on the sound pressure level difference between the ears. These investigations began with those of Lord Rayleigh in 1875. Rayleigh noted that level differences at the ears appeared to be responsible for some aspects of perceived sound direction, but a number of issues were unresolvable using the model. Thirty years later Rayleigh proposed a model of directional hearing that used both interaural level differences (ILD) and interaural time difference (ITD) [45], now known as “Duplex Theory”.

Mathematical models of directional hearing often assume a spherical model of the head. With such a model, a sound presented on the medial plane has equal pressure levels at the ears, inviting the interpretation that listeners should have difficulty localizing such stimuli.

The spherical head model also introduces the “cone of confusion”, a set of locations at the side of the head for which the ILD and ITD cues are constant. These locations lie on the surface of right circular cones, with the apex in the ear, extending perpendicular to the head. In both circumstances, head rotations may be used to disambiguate the location of the stimulus.

The interaural level difference varies as a function of the frequency being presented. Using a spherical head model, Rayleigh proposed that the head would not be an effective acoustic barrier for frequencies below 128 Hz [45]. Middlebrooks found a 20 dB ILD for tones at 4 kHz presented perpendicular to the head, which increased to a 35 dB ILD for tones at 10 kHz [33]. Blauert reports on findings in Kietz’ paper “Das Ramliche Horen”, that a 15 to 20 decibel difference is enough to lateralize a source to one side of the head or the other. The perceived width of an acoustic object in space is influenced by the loudness of the source [9]. A source may appear more diffuse as the loudness increases, resulting in perception of a wider sound. ILD cues have been found to be most effective above 1500 Hz, where the wavelength is short enough that the head is able to produce a large acoustic shadow.

Interaural time difference is an overloaded term that requires disambiguation; a sound presented binaurally may contain multiple time-difference related cues. Periodic tones presented binaurally offer timing information in terms of an interaural phase delay (IPD), which is known to be periodic over $[-\pi,\pi]$. An IPD of approximately π creates the perception of two auditory objects, lateralized to either side of the head. In general, with small phase angle differences a single object is perceived lateralized towards the ear considered to be leading in phase. This phase information is ambiguous at high frequencies, and the IPD has been found to be valid for frequencies below 1500-1600 Hz [9] [52].

Blauert notes that lateralization can be induced by modifying the phase difference between the envelopes of stimuli presented at the ears. When the frequency content in the two envelopes is sufficiently similar a single object is perceived, lateralized by the phase delay between the envelopes. Beyond some threshold of similarity, two events are perceived, one at either ear. In free field listening conditions, phase delay of the envelopes and phase delay of frequency content are congruous. It is possible to create conflicting IPD and envelope lateralization cues which may result in the perception of multiple auditory objects, or the perception of motion of auditory objects [9].

The directional cues of reflected sound may differ greatly from those present in the original sound. Early wavefronts have been found to contribute to the perceived direction of the source [9][52]. Later reflections contain important information about the spatial configuration of the room, including size, placement of objects, and distance to the source, but do not influence perception of direction. The “Franssen Effect” is a particularly striking demonstration of the influence of early wavefronts [30]. A narrow band signal is presented from a speaker with a sudden onset, and a gradual offset, while from a second location, the same narrow band signal is gradually introduced. The perception of the sound remains from the direction of the initial wavefront.

Interaural level difference and interaural time difference cues impart lateralization effects on audio, but the source is perceived to reside inside of the observer’s head. The head, chest, and pinnae impart spatial information into a signal, acting as frequency dependent filters which affect perceived location. Wightman and Kistler recorded the response to a wide band stimulus at the eardrum, as presented by various spatial locations around the head, and from headphones. This allowed them to construct a linear filter now known as the Head Related Transfer Function (HRTF) which can be used to model the effects of the torso, head and pinnae, and create free-field listening effects for sound presented over headphones [54]. The pinnae, captured in the HRTF, aid in determining elevation changes, and reduce front back confusion, but the approach is not without problems. Head rotation is an important factor in localization, but the HRTF approach requires some form of head tracking to facilitate it. HRTFs are usually not created for individual listeners; instead a pre-sampled filter is used. The mismatch between the listener's own HRTF and the sampled HRTF may result in localization problems, though subjects appear to learn to use a new HRTF given enough time.

2.1.2 Distance

At distances greater than 15m air begins to have an audible affect, acting as a low-pass filter. More important than physical distance is the reverberation quality of materials which reflect sound energy. Nielsen investigated perceived distance of sound presented over a loudspeaker in both echoic and anechoic rooms. He found that "... in normal rooms the sound is perceived at about the same distance as the physical distance, regardless of the playback level. In the anechoic room there is no correspondence between physical and perceived distance." [37]

The precedence effect, or "law of first wavefront", states that wavefronts arriving within a very short window of the initial wavefront are integrated to produce directional information. Subsequent reflections provide information about the spatial structure of the environment, and can be used to reinforce the initial auditory event without substantial loss of directional information [21].

Distance and reverberation are remarkably complex, but are not modelled in any of the acoustic experiments presented in this thesis, and we will not discuss them further.

Chapter 3

Panning

3.1 Amplitude Panning

Free field positional audio systems rely on the robustness of the brain's ability to localize sound. Such systems assume that approximations of a natural sound field will invoke a meaningful localization response. This assumption is reasonable if cues used to localize natural sound are also present in the approximation. The validity of the approximation can be understood by examining which cues are present, and to what extent they provide conflicting information. Amplitude panning techniques create an approximation of a natural sound field as the sum of wavefronts produced by speakers. The multitude of factors involved in localization has lead to a number of panning techniques and "panning laws" which model sound field propagation around the head. These panning laws are derived from physical properties of sound wave propagation, and perceptual models of spatial hearing. The wave equation is used to model sound wave propagation, and its solutions are important in developing physically plausible panning algorithms.

$$\frac{\partial^2}{\partial t^2} p(x,t) + c^2 \bigtriangledown^2 p(x,t) = 0 \tag{3.1}$$

Figure 3.1: The Wave Equation

$p(x,t)$ represents the position x of a particle at time t. c is proportional to the speed the wave through space (343 m/s) [52].

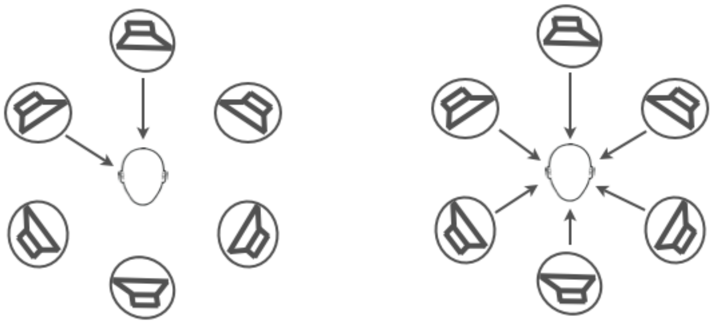
Panning techniques generally employ plane-wave solutions to the wave equation in an attempt to recreate an accurate sound field. The Fourier transform of the plane wave solution gives rise to the formula ($S_k = Pe^{i\vec{k}\cdot\vec{x}}$), with spectral component P, wave vector

$\vec{k}$ and listening position $\vec{x}$, which separates directional information and spectral content. The Fourier transform removes time dependence and separates the frequency (spectral) information from the directional information.

Amplitude panning techniques approximate the sound field around the head as the sum of a set of (Fourier transformed) plane waves $S_k = \sum P_n e^{i\vec{k}_n \cdot \vec{x}}$. Time and time-amplitude panning systems are generally unsuitable for loudspeaker reproduction due to conflicts arising from frequency dependent timing cues [42].

The two most commonly encountered panning laws are the "Sine Law" and the "Tangent Law", both of which arise from geometric analysis of the auditory scene. These laws are used to derive amplitude coefficients for a pair of loudspeakers, with intent of creating a virtual or "phantom" image accurately positioned in space. The sine and tangent panning laws state a relationship between the angles of the "phantom image", the speaker positions and the gains applied. The sine law $\frac{\sin(\theta_{virtual})}{\sin(\theta_{speaker})} = \frac{Gain_{Left} - Gain_{Right}}{Gain_{Left} + Gain_{Right}}$ makes the assumption that the head does not provide an obstacle to sound waves, making it applicable at low frequencies [7]. The tangent law, $\frac{\tan(\theta_{virtual})}{\tan(\theta_{speaker})} = \frac{Gain_{Left} - Gain_{Right}}{Gain_{Left} + Gain_{Right}}$, is derived by minimizing the difference between a plane wave from direction $\theta_{virtual}$ and the linear combination of two plane waves generated by loudspeakers [7]. The differences between the sine and tangent laws give rise to variations on panning techniques, optimized for different frequencies. Construction of a very accurate panning system requires frequency dependent panning, with very careful consideration of the transition region. The system implemented in the current study uses each of the following techniques independently.

The techniques presented here are capable of handling three-dimensional environments, but we restrict our discussion to sound being emitted from a horizontal ring of speakers in the plane of the listener. Ambisonic systems use a number of speakers to create a sound field that closely matches the sound field created by the natural event, where vector base panning systems employ fewer speakers at a time, at the expense of a less accurate but



(a) Vector-base amplitude panning (b) Ambisonic panning

Figure 3.2: Vector-base panning uses a small number of speakers at a time to create the illusion of sound direction. Ambisonic panning uses all available speakers to produce directional information

more stable model (Figure 3.2).

3.2 Ambisonics

The wave equation provides a model of acoustic phenomena, allowing a mathematical basis for the development and analysis of amplitude panning techniques. The plane wave solution is used in Ambisonic reproductions, controlling gain and phase at a number of speakers in an attempt to recreate a plane wave from an arbitrary direction. Modelling a plane wave as a finite sum of other plane waves results in a band-limited approximation of the original; a larger number of speakers implies a higher quality reproduction. Ambisonics uses this band-limited approximation to produce plane waves from a set of speakers to approximate the original plane wave function.

Plane wave solutions to the wave equation introduce a particular physical requirement; listeners must be far enough from speakers that the local curvature of the sound wave at the listening position is effectively zero. Jens Blauert reports that roughly 3 meters is required to meet this requirement for a point source emitter [9].

Modelling a plane wave as bessel functions and spherical harmonics is similar to a model used in physics [46]. In particular, the representation of the plane wave as a sum of Bessel functions and spherical harmonics was developed to model diffraction and scattering of particles interacting with the hydrogen atom. In video games, lighting functions greatly influence the atmosphere and visual quality of the final product. “Spherical harmonic lighting” has gained popularity in the field in an attempt to bring an impression of global illumination to real time lighting [43][49].

Michael Gerzon applied this technique to spatial audio, introducing an audio encoding, storage, and decoding technique he termed Ambisonics. Ambisonic recordings employ a “SoundField microphone”, a composite microphone which is equivalent to projecting the incoming directional information onto a spherical harmonic basis. Separate feeds for the directional information can be stored for later playback, an encoding known as a “B-Format”. A decoding matrix can be derived from a given speaker layout, allowing for reproduction over varying geometries, including mono and two speaker stereo compatibility. Non-symmetric geometries such as Dolby 5.1 require tuning of the analytical results [36]. Solutions corresponding to a symmetric speaker layout are often adopted even when the speaker layout is non-symmetric [35].

Synthetic directional information can be created for a monaural source, creating an Ambisonic feed, allowing Ambisonic decoding. The encoding and decoding can be combined into a single step, providing Ambisonic panning for a monaural source.

3.2.1 Plane Wave Expansion

When dealing with Ambisonics, the conventional cartesian coordinate system is rotated $90\circ$ counter clockwise, with the $+Z$ axis representing elevation. A plane wave at angle θ

to the x axis, with elevation ϕ can be written in terms of Bessel functions of the first kind and spherical harmonics as [13]

$$S = \sum_{m=0}^{\infty} i^m J_m(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \phi) \tag{3.2}$$

Y_{mn}^{σ} represents a (degree m, order n) spherical harmonic term with coefficents B_{mn}^{σ}, and $J_m(kr)$ the is the Bessel series.

Restricting ourselves to two dimensions, the expression for the plane wave can be simplified [4]

$$S = PJ_0(kr) + P \sum_{m=0}^{\infty} 2i^m J_m(kr)[cos(m\theta)] \tag{3.3}$$

A linear combination of solutions to the plane wave equation is also a solution to the wave equation. Using this fact, and noting that sound can only be reproduced from discrete speaker directions, θ_n, Bamford suggested the following alternative equation [4].

$$\hat{S} = \sum_{n=1}^{N} P_n J_0(kr) + \sum_{m=0}^{\infty} 2i^m J_m(kr)[\sum_{n=1}^{N} P_n cos(m\theta_n)] \tag{3.4}$$

S and $\hat{S}$ (the original, and reconstructed plane-wave) are equivalent when the original sound pressure is preserved in the reconstruction of $\hat{S}$, and the sum of the reconstructed plane wave direction vectors is equal to the direction of the original source.

3.2.2 Encoding and Decoding

Encoding a plane wave from a general direction can be expressed as the projection of the plane wave onto the spherical harmonic basis. With the B-Format coefficients, the gains for each speaker can be expressed as $S = C^{-1}B$, where C is a matrix of spherical harmonics evaluated at the speaker positions. When C is non-square, the pseudo-inverse of

C is used[13].

3.2.3 Equivalent Panning

Daniel et al. [13] noted that for circularly symmetric layouts with n speakers and band-limit M, encoding and decoding can be combined into a single panning function.

$$G(\gamma) = \frac{1}{n}(g_0 + 2\sum_{m=1}^{M} g_m cos(m\gamma)) \tag{3.5}$$

This equation can be calculated for each speaker, where γ is the shortest angle between the intended sound direction and the speaker direction.

Weights g_m are used to define variants of the ambisonic panner. Due to differences in how sounds at various frequencies are localized, Gerzon suggests frequency dependent spatialization. Gerzon translated ITD cues and ILD cues into a mathematical model he called the “General Theory of Auditory Localization” [18]. The theory proposes a pair of vectors, known as the “velocity vector” and the “energy vector”, which are quality index of spatial reproduction, roughly corresponding to ITD and ILD cues, respectively. The velocity vector can be written as $\vec{r_v} = \frac{S.g}{\sum g_i}$ and the energy vector as $\vec{r_e} = \frac{S.g^2}{\sum g_i^2}$ Reproduction is most accurate when both vectors point in the same direction, that direction is the intended direction, and both vectors have magnitude 1. [28]. In general, it is not possible to satisfy the conditions over the velocity vector and the energy vector simultaneously; consequently, there are multiple decoder types for ambisonics, with ideal spatial reproduction requiring frequency dependent decoding, which was not performed in the current work.

Neukom [35] simplified the panning function of Daniel et al., substantially reducing the computational cost for high-order approximations.

$$f_{basic}(\theta, n, m) = \frac{sin(\frac{2m+1}{2}\theta)}{nsin(\frac{1}{2}\theta)} \tag{3.6}$$

$$f_{in-phase}(\theta, n, m) = cos(m\frac{\theta}{2})^{2m} \tag{3.7}$$

These equations provide simpler panning functions than provided by Daniel et al. Unfortunately the basic decoder is numerically unstable when θ is close to zero (corresponding to the source direction lying in the direction of a speaker), making it less practical than the in-phase equivalent panning function.

The implemented solution uses the equivalent functions from Daniel et al. [13] for basic and max $\vec{r}_e$ panning, and Neukom's [35] equivalent function for in-phase panning.

3.3 Vector Base Amplitude Panning

Vector base amplitude panning (VBAP) attempts to recreate the sound field for a single source by superposition of a small number of plane waves [39]. The theory is appropriate for speakers arranged equidistant from a central listening position, allowing for both two- and three-dimensional reproduction. Speakers do not need to be evenly spaced, allowing a designer to match positional fidelity to perceptual resolution and requirements.

For a virtual source with azimuth θ and elevation ϕ, we require gains over a set of loudspeakers that image the source in the intended direction. Gains can be viewed as weights over speakers, with the restriction that negative gains are set to zero. This view of the gains treats them as a set of positive barycentric coordinates over a space triangulated by speakers. For horizontal-only reproduction, this is equivalent to choosing speakers i, i+1, such that, $\theta_i \leq \theta_v < \theta_{i+1}$.

A source imaged by two, or three speakers sounds wider than a source imaged by a single speaker. This results in spatial blurring of sound as the angle between the virtual direction and the closest speaker increases. Pulkki suggests panning the source to multiple near-by directions to achieve a uniform spread. At speakers in front of a listener, separated by thirty degrees, spread was measured to be less than 3.5 degrees [40]. Spreading increases as a source moves away from a speaker, suggesting an alternative solution: Increasing the number of speakers used in the reproduction.

With the appropriate loudspeakers determined we now determine their gains. Writing the desired direction as $\vec{v}$, a conical combination of speaker vectors $\vec{B}_i$, $\vec{B}_j$ and $\vec{B}_k$, we are able to express direction as a set of weights.

$$\vec{v} = \begin{pmatrix} \vec{B}_i & \vec{B}_j & \vec{B}_k \end{pmatrix} g_{\vec{vbap}} \tag{3.8}$$

B is invertible where speaker directions $\vec{B_{ijk}}$ are not co-linear. Solving for $g_{\vec{vbap}}$ provides gains for speakers i, j and k.

$$g_{\vec{vbap}} = \begin{pmatrix} \vec{B}_i & \vec{B}_j & \vec{B}_k \end{pmatrix}^{-1} \vec{v} \tag{3.9}$$

To maintain constant amplitude as source moves through space, gains are normalized as $\frac{g_{\vec{vbap}}}{|g_{\vec{vbap}}|}$ before they are applied. A constant amplitude implies that the source moves on a sphere surrounding the listener. Distance information is not encoded, and must be modelled separately.

Vector base amplitude panning is a generalization of the tangent law, which can be seen by considering equation 3.9 in two dimensions [6].

3.3.1 Vector Base Intensity Panning

Pernaux noted that in the three channel B-Format used in Ambisonics, directional information was equivalent to that produced by VBAP panning, up to a constant scaling factor. Applying Gerzon's theory of localization, Pernaux suggested that VBAP was appropriate for modelling the velocity vector, and suggested Vector Base Intensity Panning to model the energy vector.

$$g_{\vec{i}_{vbip}} = \sqrt{\frac{g_{\vec{i}_{vbap}}}{\sum_j g_{\vec{i}_{vbap_j}}}} \tag{3.10}$$

Chapter 4

Software

4.1 Design Considerations

We required a system suitable for spatial audio EEG experiments. Pre-packaged software platforms exist for use in EEG research (e.g. E-Prime, Psychology Software Tools), but experience had shown that extending their functionality to spatial audio experiments was difficult. Rather than provide a complete "out-of-the-box" solution for spatial audio experiments, software was developed for specific experiments, with reusable components for playback, spatialization and EEG recording. Synchronization of auditory onset with EEG was required for data analysis, making low-, fixed- latency control a high priority.

Spatialization of an audio source can be performed using a number of different algorithms, each with their own characteristics. These characteristics are the result of compromises in the algorithm, often based on some mathematical or psycho-acoustic principle which may preclude the ability to perform certain types of experiments. We required a framework that exposed these details, and allowed for the substitution of panning algorithms.

We wanted to create a multi-speaker auditory display, exceeding the typical 5.1 and 7.1 speaker limits provided by most hardware. We wanted to allow for various speaker configurations, and the ability to expand the number of speakers at will.

4.1.1 Auditory Presentation

Audio presentation systems intended for psycho-acoustic research must provide stable, artifact– and distortion–free playback. Consumer grade speaker systems are engineered to provide a subjectively enjoyable sound for music and movies. The power output across the reproduced frequency range may not accurately represent the source signal. The "frequency response profile" of consumer grade systems is rarely published, making them unsuitable for stimulus presentation. Professional studio monitors used in recording engineering are designed with a flat frequency response profile which provides a neutral reproduction, which is ideal for mastering or mixing. These monitors have published frequency response profiles, and are carefully calibrated by the manufacturer. Flat response studio monitors make them ideal for our purposes.

An auditory stimulus presentation system should provide low, known-latency playback. Computer audio systems ensure stable reproduction by filling buffers with sample data before the data is required for playback. The size of the buffer directly determines the stability and the minimum latency in the system. Large buffers are more stable; smaller buffers are serviced more frequently, reducing the time between a playback request and realized onset. Scheduling and system load can result in variability between the requested onset time and the realized onset time. Accurate and precise timing is important for perceptually correct playback. The expected onset time for a buffer is provided by many audio programming APIs. A auditory event can be scheduled for an onset time with high precision by calculating a sample offset from the currently-in-service buffer onset time, filling buffers with silence until the auditory event should occur. We require the ability to determine if a source started on schedule, and to provide an alert if a scheduled time elapsed before playback could begin.

Most audio hardware supports audio sampled at 44100 or 48000 Hz, requiring audio at other sampling rates to be resampled. Resampling during playback is processor intensive and may induce latency in the system. Choosing natively supported audio encodings reduces system load, and improves overall performance. Playback, real-time mixing, and digital signal processing require a significant amount of processing power. To meet these demands, audio samples are buffered and pre-processed on a real-time or high-priority thread. Any operation performed on this thread taking an extended amount of time will result in gaps of unexpected silence during playback. Allocating memory, or causing the thread to block by locking a shared mutex is unadvised. Consequently, developers writing audio code are advised to use lockless multi-threading techniques and carefully manage access to shared data.

Audio presentation systems intended for psycho-acoustic research must meet a set of criteria which ensure that the data produced are valid. In particular, it is important to provide low, known-latency audio playback. Time elapsed between a playback request and realized onset is not constant. As a result, initiating playback and immediately taking a time-stamp from the calling thread is not a reliable way to determine onset. A simple alternative is to calculate the onset time stamp from a supplied buffer-onset time, the offset into the buffer, and the sampling rate. This provides a simple mechanism to determine onset of a single source, but does not allow sources to be scheduled to have the same onset. A viable alternative is to begin playback, filling buffers with silence until an auditory event should occur. The event can then be scheduled for playback slightly in the future, giving sample-accurate resolution. Scheduling an event for immediate playback results in inaccurate timing because scheduling itself takes time. By the time the event has been scheduled, the time it was scheduled for has already passed. The reason for scheduling slightly in the future is now clear, but it is not clear how far into the future the schedule should be set. We require the ability to determine if a source started on schedule, and to provide an alert if a scheduled time elapsed before playback could begin. Scheduling also allows multiple sources to synchronize onset.

The system must be able to address any number of speakers, with the freedom to choose among a set of panning algorithms with known, and controllable properties. Most audio hardware supports no more than eight speakers at a time through a single interface type.

To support a larger number of speakers, multiple audio devices are required. Each audio device has an internal clock, which must be synchronized with the host CPU.

Support for streaming audio is required, with the ability to queue and transition between buffers without transient effects.

4.2 Audio Formats

A number of proprietary and open standard audio compression schemes exist to reduce audio file size. These formats often employ a lossy encoding scheme, modifying spectral information of a signal based on various perceptual characteristics. Lossy encoding is also prone to a particular artifact known as “pre-echo”, which blurs the temporal onset of a signal. Decompressing the audio format into a standard pulse code modulated (PCM) signal requires additional processing which increases latency and may increase latency variability. PCM signals with a sample rate that differs from the native sample rate of the operating system or audio hardware may undergo a sample rate conversion. Additional latency and overhead resulting from sample rate conversion can be avoided if the native audio format is used. Any experiment dealing explicitly with timing or spectral information should not use a compressed audio format.

4.2.1 Electroencephalography

In addition to interfacing audio hardware, we require the ability to send time-stamped labels to NetStation – proprietary EEG recording software developed by Electrical Geodesics Incorporated. NetStation exposes a TCP/IP interface, which allows interaction from within custom experiment software. The time stamps sent to NetStation must accurately reflect the onset times of stimuli, as they are used to mark segments of EEG for further study.

EEG data is augmented with timestamped markers indicating the onset of a stimulus or event. For each type of event an ERP is formed by averaging data following the marker. This averaging process acts as a filter, dampening noise (and high frequency components), while showing trends in low frequency data. The stimulus event is thought to cause phase-locking of electrical potentials of a particular frequency, and the following voltage series is averaged between subjects in an attempt to remove noise and isolate activity strictly related to the stimulus event. Variability between the onset time and the recorded timestamp acts as a low pass filter, potentially invalidating the data.

4.2.2 Acoustic Environment

Speakers were placed at a distance of 1.27 m to the central listening position. Rubberized sound attenuating material was attached to the walls of the room, dampening acoustic reflection. To further reduce acoustic reflection in the room, foam baffles were attached to walls and placed in corners. Curtains were hung from the walls, and the ceiling panel was backed by insulation to reduce ambient noise. The ventilation system in the room could not be covered, and was responsible for a background ambient noise. Acoustically reflective surfaces included a metal arm that connected to the EEG net, the chair and table at which subjects worked, as well as the monitor, keyboard, mouse, and the ventilation housing. Ambient noise in the room was measured to be 60 db.

4.3 Software Framework

OpenAL, a handful of Linux APIs, Microsoft's DirectSound, XAudio, and Core Audio frameworks, and Apple's Core Audio framework were considered for their suitability in acoustic experiments.

OpenAL is a high level auditory spatialization API which is appropriate for use in games and virtual environments. Most implementations provide a number of spatialization algorithms, making it suitable for use in free-field or with headphones. The spatialization algorithm used can not be explicitly chosen. API settings allow the developer to request panning quality, without specific knowledge of the algorithm selected or those available. Examining the reference implementation is useful in this regard, but does not provide enough information to determine the ecological validity of the implementation, and may not reflect distributed implementations. The API does not provide a mechanism to address individual speakers directly, and applies the same spatialization algorithm to all sources. There is no way to schedule sources for playback, making it difficult to accurately control onset variability, and multiple source onset synchronization.

Linux based operating systems offer a variety of audio interfaces, each independent of the others. This has resulted in a number of problems in audio driver stability on the platform. Real-time Linux kernels are available for popular distributions such as Ubuntu. Time-sensitive neuro-imaging experiments stand to benefit from such systems, and as drivers and software interfaces mature, Linux based operating systems will become more attractive.

Microsoft Windows offers a number of audio interfaces with various characteristics. Audio interfaces on Windows XP are generally considered high latency. Many professional recording engineers use the Steinberg ASIO interface. This interface is not native to the Windows XP operating system, but allows software to bypass most of the Windows audio stack to directly interface with hardware, reducing latency and improving performance. Microsoft DirectSound was widely used to provide 3D audio for games and for a number of experiments, but is now deprecated and no longer supported. DirectSound offered similar functionality to OpenAL, with similar drawbacks.

In Vista the Microsoft Windows audio framework was redesigned to improve performance. There are a number of audio interfaces at various levels of abstraction. The XAudio2 interface provides functionality similar to DirectSound. Source synchronization is supported, but scheduling is not. Sources will begin playback immediately once the playback request has been handled. Microsoft's Core Audio is a lower level audio framework with interfaces that expose output buffers. Direct access to output buffers allows for implementation of the scheduling strategy described above.

OS X provides a variety of audio APIs collectively termed Core Audio. (This framework is not related to, and is more established than Microsoft's Core Audio framework.) Core Audio supports audio processing by connecting audio processing units together into a processing graph. One end of the graph represents the connection to the physical output (or input) device. Each node in the graph, starting with the end point, requests sample data from the previous node in the graph. Nodes may have callbacks registered against them that provide the requested data directly. These callbacks allow the scheduling mechanism above to be implemented. OS X provides a unit capable of providing 3D audio support. The unit supports a variety of panning algorithms but does not allow for arbitrary speaker arrangements, nor for an arbitrary number of speakers. Current support is limited to no greater than Dolby 5.1 home theatre systems, which is not sufficient for research use. The matrix mixer unit allows inputs to be mapped arbitrarily to outputs, allowing customized spatialization algorithms to be implemented. The mixer requires data buffering which increases audio latency, but the latency is a constant function of mixer parameters. Timestamps for audio onset are available, allowing for correct timing and registration with EEG recording.

4.4 System Development

Core Audio (OS X) was chosen as the foundation to build the audio experiments. The choice was based on the requirement specification, the advice of recording engineers and hobbyists, available hardware, and personal experience. The author has more experience developing high-performance applications for OS X and for Linux-based operating systems than for Microsoft Windows. NetStation is not available for Linux-based operating systems or for Windows. Concern over non-local network connection latency motivated us to target OS X, so that experiment programs and NetStation could run on the same machine, if necessary.

4.4.1 Audio Devices

OS X provides audio device aggregation, allowing software to interface with a single virtual audio device. Device aggregation automatically maintains clock synchronization between the devices and the host machine. Two M-Audio FireWire 410 interfaces were aggregated to provide 16 channel 3/4 inch analogue outputs. Fourteen Mackie HR624 MK2 studio monitors were connected, offering high-quality sound reproduction with flat frequency response profiles across the audible frequency domain. Speakers were equally spaced on a ring of 1.27 m around a central listening position. Azimuthal position and height were calibrated using a laser level situated at the desired listening position. The system does not require a ring configuration in general, but it reflects the configuration required for the majority of our intended experiments.

4.4.2 Scheduling

Accurate timing of stimulus onset was a primary goal of this research system. Time on OS X is represented as a processor-dependent quantity known as “mach time”, which shares a linear relationship with “wall-clock” time. Mach time provides nanosecond accuracy, a resolution greater than the 44100 Hz required for sample-accurate scheduling. Sources are scheduled for playback directly in mach time, allowing for tight integration with other systems.

The developed system provides audio hardware with data by invoking a function pointer. This function is executed periodically for each source the audio system has registered against it, whether the source is currently playing or not. The function must determine the state of the source, whether data should be copied to the output buffers, perform the data copy, and update the source state. When invoked, the function is supplied with pointers to output buffers, the number of required samples for each buffer, and the expected onset time for the buffers being filled.

Initiating stream playback has a small amount of overhead. A commonly employed programming technique to reduce playback latency is to begin playback well before the stream is required. The stream may then be paused, or in our case, filled with silence until the auditory event should occur. The event can then be scheduled for playback slightly in the future, giving sample-accurate resolution. Scheduling an event for immediate playback results in inaccurate timing due to the buffering technique described earlier. To ensure that time stamps are accurate we require the ability to determine if a source started on schedule, and to provide an alert if a scheduled time elapsed before playback could begin. Scheduling also allows multiple sources to synchronize onset.

Schedule realization in the callback is performed by first determining the state of the audio source. Sources marked as being in a play state must determine the number of samples left in their buffer, accounting for looping properties. The offset into the output buffer is computed from the onset time and the sample rate of the output buffer. Data are copied to output buffers, and the source sample offset is incremented. When a non-looping source has reached the end of its buffer, it is marked as being in a stopped state, and the sample offset is set to zero. Onset scheduling is performed once each callback, meaning that a source scheduled for immediate playback must wait until the next audio IO cycle to realize onset. Buffer sizes can be adjusted to reduce onset latency, incurring an increase in the number of callbacks required to process the audio data. The system may not be fast enough to process callbacks for small buffer sizes, resulting in unintended gaps in the stimulus. Core Audio uses a default buffer size of 512 samples – approximately 11.61 ms of audio data sampled at 44100 Hz. This is significant for scheduled sources as it provides a lower bound on the time delta required to meet scheduling demands.

Regardless of scheduled onset, realized onset time is available for each source. Realized onset times should be used to inform electroencephalographic or behavioural systems. In addition to scheduled and realized onset times, offset times for sources are also available.

4.4.3 Multithreading

Core Audio guidelines suggest the use of lockless multithreading to ensure uninterrupted stimulus output. Consider a lock shared between the real-time data-providing thread, and an application thread. If the application thread acquires the lock it may prevent the data-providing thread from filling buffers within its time-slice. It is tempting to believe that a lock held for a short period of time would not result in stimulus presentation failure. The scheduler in the operating system may swap threads or processes, suspending the thread holding the lock, extending the intended duration of the lock. This may result in stimulus presentation failure.

Developers attempting to write lock free threaded code must be aware of subtleties of both the hardware architecture and the optimizing compiler under consideration. Section 8.2.2 of the “Intel 64 and IA-32 Architectures Software Developer’s Manual” [1] summarizes hardware re-ordering of memory stores and loads on modern x86 processors. Instruction re-ordering can cause otherwise algorithmically correct code to fail, as in the case of Peterson’s spin lock [23]. To order memory accesses without use of a shared lock, a memory barrier can be used. Modern processors offer a set of atomic operations that can be used in conjunction with memory barriers to build lock-free thread-safe code. Careful factoring of code, and a good knowledge of the memory ordering rules for the processor running the code can also be used for designing lock-free protocols.

The audio code in the stimulus presentation system was factored to allow lock-free multithreading. This was done using standard double- and ring-buffering techniques. Audio is streamed into a ring buffer from the application thread, which is then used to fill output buffers. Reading and writing operations on the buffer do not overlap, allowing each to work independently.

4.5 Electroencephalographic Interface

Electrical Geodesics Incorporated supports a TCP/IP interface to NetStation, allowing developers to write software to send time-stamped events. These events can be used to segment the EEG stream, for the creation of ERPs. The primary consumer of the interface is E-Prime. E-Prime is experimental control software developed by Electrical Geodesics Incorporated. The software allows non-programmers to create simple auditory and visual experiments, and offers basic scripting for more advanced control. The software is limited in scope, and advanced spatial auditory experiments are very difficult or not possible.

Our research required development of a library to interface with NetStation. Electrical Geodesics Inc. provides documentation on NetStation's communication protocol, but the documentation is erroneous and incomplete. In particular, the documentation does not accurately describe the required packet layout required by the recording machine.

libnetstation is an open-source implementation of the NetStation protocol developed by the author, and made available online. The library provides a simple interface to NetStation, abstracting network communication. The library has been tested extensively for use on OS

X and Windows, and has no known issues.

Keeping the clocks synchronized between the stimulus presentation system and the recording system is essential to capturing useful data. The interface allows this by sending a synchronize command, followed by a timestamp. This information is used to calculate the offset between the timestamp sent and a timestamp taken on the local machine, to account for clock skew. This approach relies on low latency transmission between the stimulus presentation system and the recording system. It is important to remove routers, switches, and hubs between the two systems, which can introduce variable latency. TCP/IP stacks implement algorithms designed to reduce the amount of data transmitted over a network. Nagle's algorithm introduces latency in packet transmission by waiting for an acknowledgement of data currently in flight before sending the next packet. The delayed acknowledgement algorithm attempts to reduce the number of acknowledgement packets sent by assuming that the TCP acknowledgement packet will be followed by response data from a program. The algorithm delays the TCP acknowledgement packet in an attempt to concatenate it with the response data. These algorithms, and the interaction of Nagle's algorithm and the delayed acknowledgement algorithm, increase latency on the order of hundreds of milliseconds. *libnetstation* reduces latency by disabling Nagle's algorithm using the TCP_NODELAY socket option.

This library has been used in a number of papers and posters presented by the cognitive neuroscience lab at the University of Lethbridge, and will continue to be used as a basic element of experiments moving forward. A list of the papers, posters, and presentations that make use of *libnetstation* can be found in the appendix.

The code has also been shared with scientists studying the brain using NetStation at the University of Alberta. The code has also been used as the basis of the python library *pynetstation*, developed by scientists at Riken - a Japanese national research institute `http://www.riken.go.jp/engn/`.

libnetstation is available at `http://code.google.com/p/libnetstation`.

4.6 Verification And Testing

Timing was verified using two independent techniques. Preliminary verification was conducted by running simple auditory event-related potential (ERP) experiments. Auditory onset ERPs have a characteristic shape, and timing variability results in spatial blurring or lack of formation of the ERP. The auditory ERP was clearly visible in the resulting analysis, demonstrating that the *libnetstation* library was not inducing significant timing variability.

The protocol specification of the *NetStation* experimental interface does not match the protocol implemented in the *NetStation* software package. For example, response codes do not match the documented format. The specification explicitly states that a number of fields are optional, but these fields are required. The message headers contain a size parameter to allow for optional field data, but the specification does not state whether this size includes command headers. These issues resulted in data corruption after transitioning from running *NetStation* from *OS 10.5* to *OS 10.6*.

A Python script was developed to allow real-time examination of both response codes and the corrupted data. Corrections to the protocol specification were reverse engineered from this data. The python script wrapped libnetstation, and allowed quickly iterating over message data being sent. By trying to control the corrupt data being received, we were able to determine the implemented message format. Using the data collected, the python code was used as a server (in place of *NetStation*) for development and debugging *libnetstation*. Using the python wrapper of *libnetstation* with the server code allowed verification of the client from the perspective of the server. Following this debugging phase the code was again tested against *NetStation*.

Rigorous testing was performed using a diagnostic timing tool available from Electrical

Geodesics Inc. The tool takes an auditory input, and connects to the EEG amplifier, generating a timestamped event when an auditory input is received. The timestamps generated by experiments were compared to timestamps generated by the diagnostic tool, and showed acceptable timing characteristics. Chaining together multiple auditory outputs could be expected to induce latency; we verified that timing characteristics met our criteria by comparing the observed latency for each of the auditory output devices.

Panning code was verified by choosing a number of random positions on the array, calculating the expected outputs by hand, and comparing against software output. This lead to the development of a Python script with an independent implementation of the panning algorithms and automated comparison of the algorithm outputs.

Chapter 5

System Validation

5.1 Introduction

Bamford's analysis and comparison of spatial auditory presentation systems indicates that second-order ambisonic systems have an advantage over discrete panning systems [4]. Pernaux compared vector base amplitude panning, vector base intensity panning and an ambisonic system in a 5.1 home theatre setup, and found that vector panning methods were favorable for stationary sound objects, while ambisonic methods were favorable for moving sound objects [38]. Basic, in-phase, and max $\vec{r}_e$ ambisonic panning systems, vector base amplitude panning and vector base intensity panning were implemented, and characterized in terms of their perceptual accuracy and width. This comparison will be used to inform future perceptual studies.

5.2 Methods

14 participants from the University of Lethbridge participated in the study. 2 males ages 20 to 21, and 12 females, ages 18 to 26, gave informed consent, and were granted course credit for their participation. All participants reported normal hearing, and normal, or corrected to normal vision. Subjects were seated in the centre of a circular array of fourteen Mackie MK-2 studio monitors. The monitors were adjusted to 72 dBA sound pressure level (SPL), and the ambient noise floor was measured at 62 dBA (SPL). Speakers were evenly spaced, at a distance of 1.26 meters to the listening position. Participants were instructed to remain still during the experiment, were asked to refrain from turning their heads, but this was not strictly enforced. Participants were told that sound would be presented from around them,

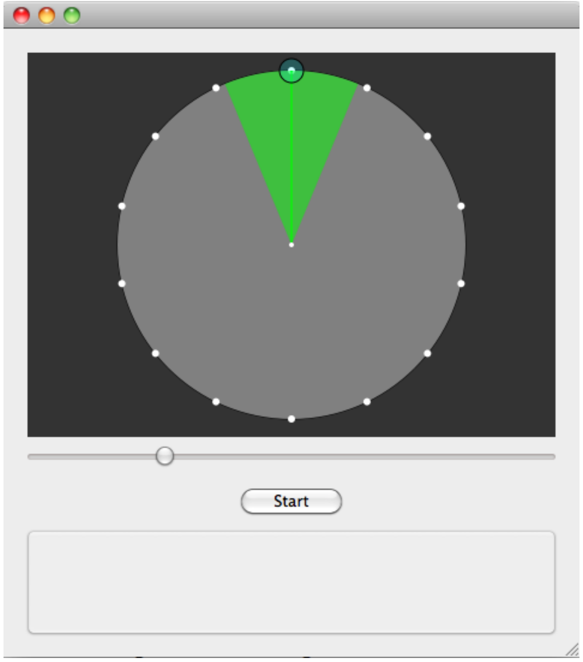


Figure 5.1: The radial control used in our validation experiment

and that sound may be coming from places in the room other than the speakers.

To evaluate vector base amplitude panner performance Pulkki [41] used a rotating loudspeaker that participants adjusted to match the direction of a virtually panned source. The azimuthal difference between the virtually panned source and the loudspeaker was interpreted as a measure of error in the panner. Other researchers have used motion tracked pointing devices [20], rotating pointing arms [19], panning of virtual sources [8], and radial controls displayed on a computer screen.[53]

We chose to use a radial control similar to that used by Wenzel [53] to collect subject response data (Figure 5.1). The direction of the radial arm indicates the azimuth of the source with respect to the listener. In addition to indicating direction, subjects used a slider control to describe the perceived width of the stimulus. Response times were collected, but there was no forced response period.

Mozart's Symphony Number 41 (Jupiter Symphony) was chosen as the stimulus for the localization experiment. The classical piece provides a wide frequency envelope, and was chosen to reduce bias towards panning algorithms optimized for particular frequency bands.

For each panning type, 56 azimuthal directions were pseudo-randomly selected around the listener, and the stimulus was panned to the selected direction. In addition, four presentations from each of the 14 speakers were presented in pseudo-random order. Subjects indicated perceived source direction and width on a graphical display. A break was offered to the subject between panning types allowing subjects an opportunity to stretch and move around, so as to keep their attention from dropping off quickly.

5.3 Results

Measurement of the difference between the perceived sound direction and the target sound direction required adoption of an error metric. This metric is not strictly a measure of "error" as there is no objective measure of direction, but we use the term to designate the difference between the perceived and intended direction. We defined this error metric as the (signed) minimum angle between the perceived direction and the target direction. The sign was dependent on the relative directions of the target and perceived sound direction; when the perceived direction was anterior to the target direction the error was positive, otherwise it was negative.

Figure 5.2 shows the normalized differences between the intended presentation angle and the perceived presentation angle, which was used as a measure of the accuracy of the panner. Cross-modal mapping between the free-field acoustic display to the visual user interface required by the response system likely induced a small amount of error, creating a floor effect. Error rates at the floor could not be reduced by improving panner accuracy. Error measures collected for direct speaker presentation were used as an indication of the error floor.

Levene's test for equality of variance over panner types produces a statistic of 5.302, $p < 0.01$, rejecting the hypothesis that the samples come from a distribution with the same variance. This is supported by the kurtosis of error distributions for each panner type (Basic = 14.5645, In Phase = 11.7139, Max r$\vec{e}$ = 15.2722, VBAP = 16.1057, VBIP = 17.3237, Speaker = 20.1804). The JarqueBera, Lillefors, and Kolmogorov-Smirnov tests were used to compare the sample distributions for deviations from normality. All tests rejected the hypothesis that the sampled error distributions were normal, likely due to their high kurtosis. A one-way, six-level, repeated measures ANOVA with factors of "Error Kurtosis" failed to find a main effect of panner type.

A one-way, six-level, repeated measures ANOVA with factors of perceived width revealed a significant main effect of panner type $F_{(5,60)} = 4.337$, $p = 0.0007$. (Note that the assumption of sphericity was violated – $\chi^2 = 29.765, p = 0.010$ – and the Greenhouse-Geisser correct significance and original degrees of freedom are reported.) Tukey's LSD tests revealed that the In-Phase panner performed significantly worse than other panners (Basic: p = 0.008; Max R$\vec{e}$: p = 0.003; VBAP: p = 0.073; VBIP: p = 0.002; Speaker: p = 0.005), but pairwise comparisons revealed no other significant differences.

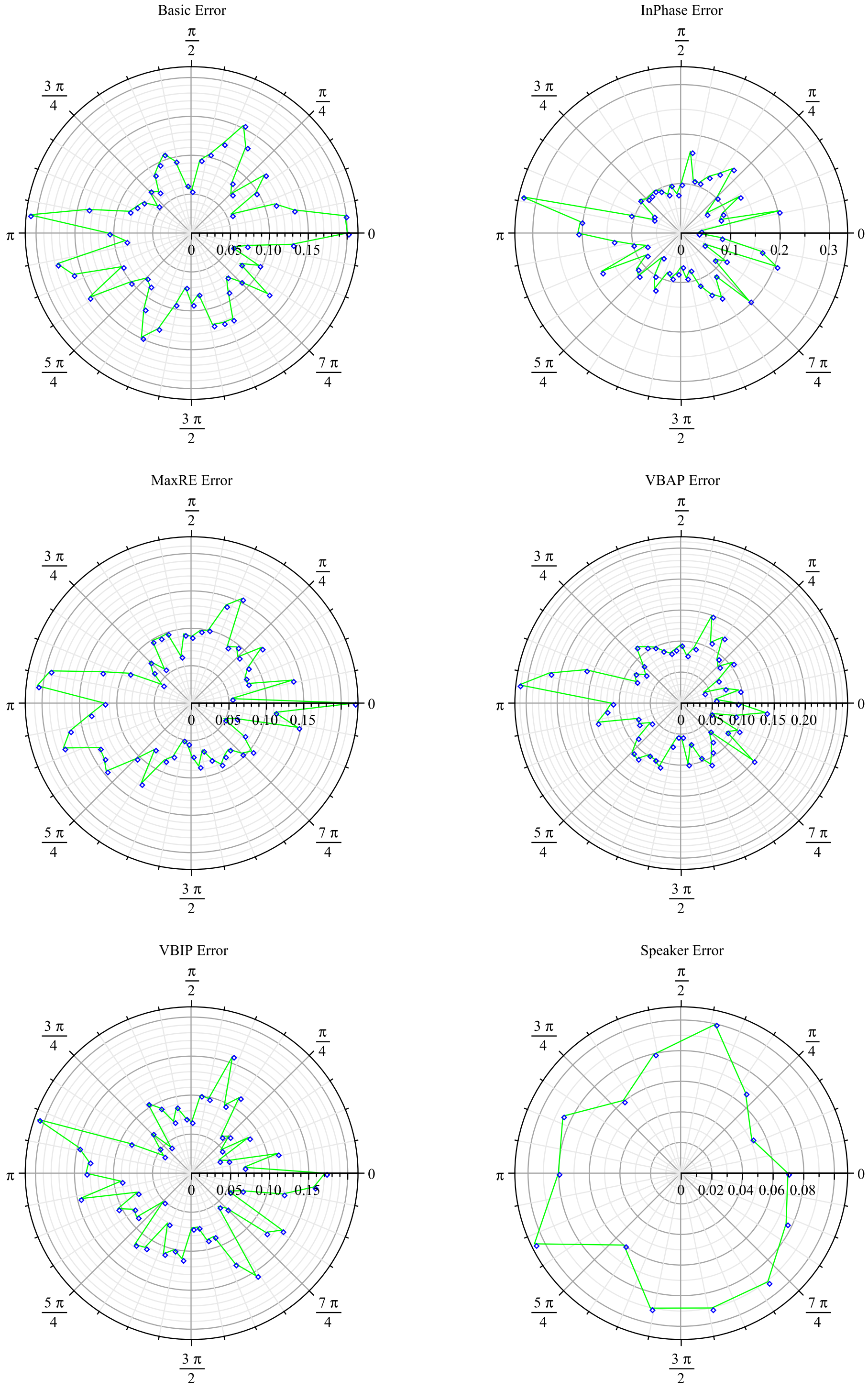
5.4 Discussion

Head rotations help resolve front-back reversals and spatial confusion for sounds presented on the midline. In the present study participants were asked to refrain from making such rotations. The same cues that cause front-back reversals cause spatial blurring, reducing accuracy even when front-back reversals do not occur. This is reflected by artifacts near 0○ and 180○ in figure 5.2, where error rates larger than in the surrounding regions.

It is important to notice that the error and width diagrams above are each at their own scales. Error and width were normalized by dividing by π to make an error of 1 represent a complete reversal and a width of 1 represent an entire hemisphere. Figure 5.2 seems to indicate that sound present from a single speaker are more accurately located by participants than a sounds presented by any panning technique. VBAP and VBIP panning techniques were verified to present sound over a single speaker when the phantom image was in the direction of the speaker, suggesting that this difference is not directly inherent in the panners used. Each trial contained 56 spatial samples, which were presented from a pseudorandomly selected set of locations for each panner type. Presentation from 14 speaker locations meant that these speakers were repeated 4 times in a trial. It is possible that subjects learned these positions more accurately, resulting in the lower error rate.

Panners were not significantly different in their error rates, though VBAP and VBIP had higher kurtosis than the ambisonic panners. This suggests that some participants were able to use VBAP and VBIP more effectively than ambisonic panners, but that there was high variance. In-phase ambisonic panning produces gains that are all in-phase with each other at the expense of a wider sound. This was reflected in the ANOVA of perceived widths over panner types.

In sample sizes used for a typical neuroscience experiment, with non-moving sound, we have failed to find significant differences between VBAP, VBIP, basic and Max $r\vec{e}$ panners. With no strong reason to choose one panner over the other, we note VBAP's relation to the tangent law and its high kurtosis relative to Ambisonic panners. High kurtosis in our distribution of error frequency demonstrates that the perceived direction of sound and the presented direction of sound were found frequently to lie in the same direction. For this reason, we choose VBAP for future experiments.



Basic Error

InPhase Error

MaxRE Error

VBAP Error

VBIP Error

Speaker Error

Figure 5.2: Panning Angle v. Perceptual Angle: Normalized Error Rate

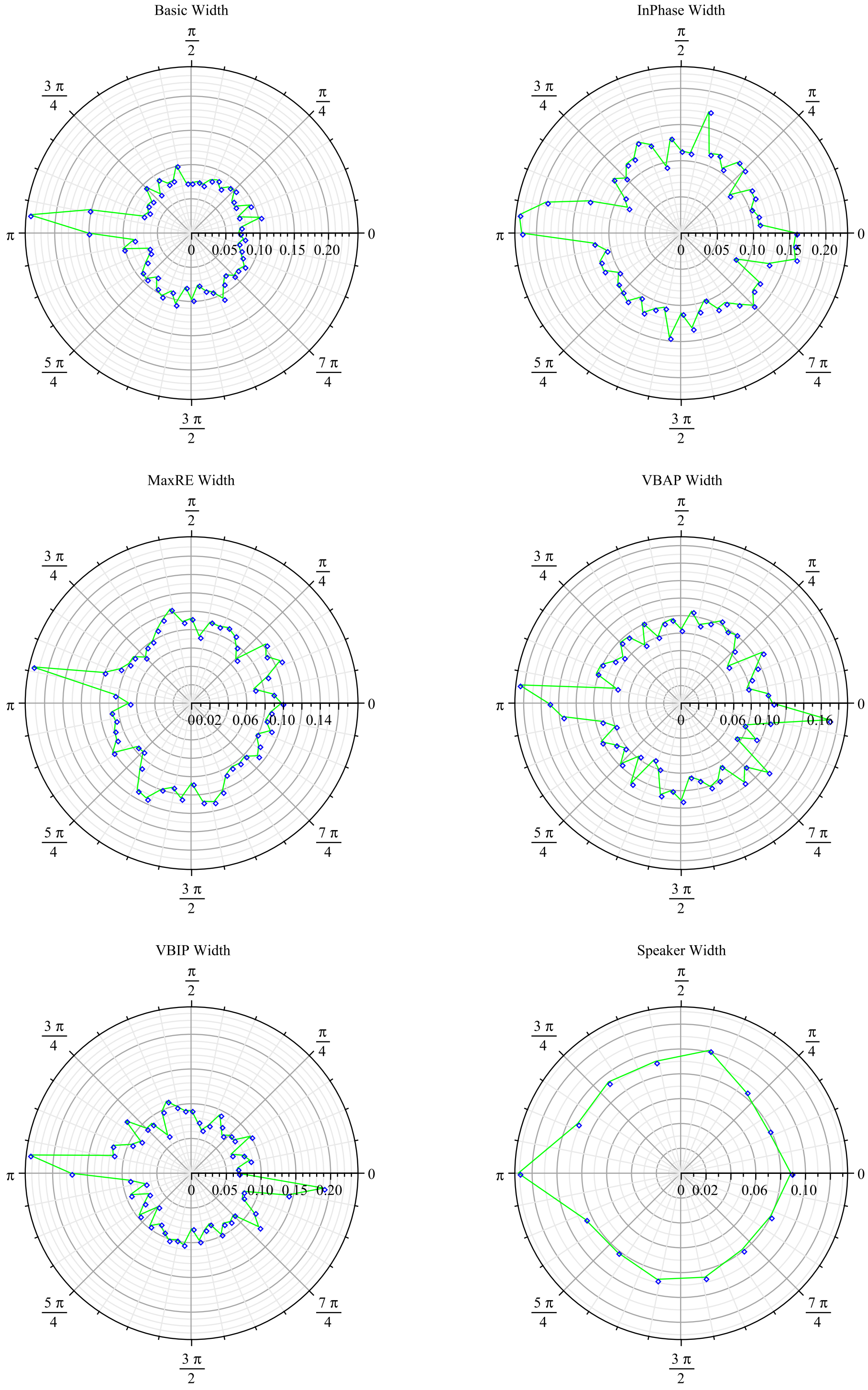
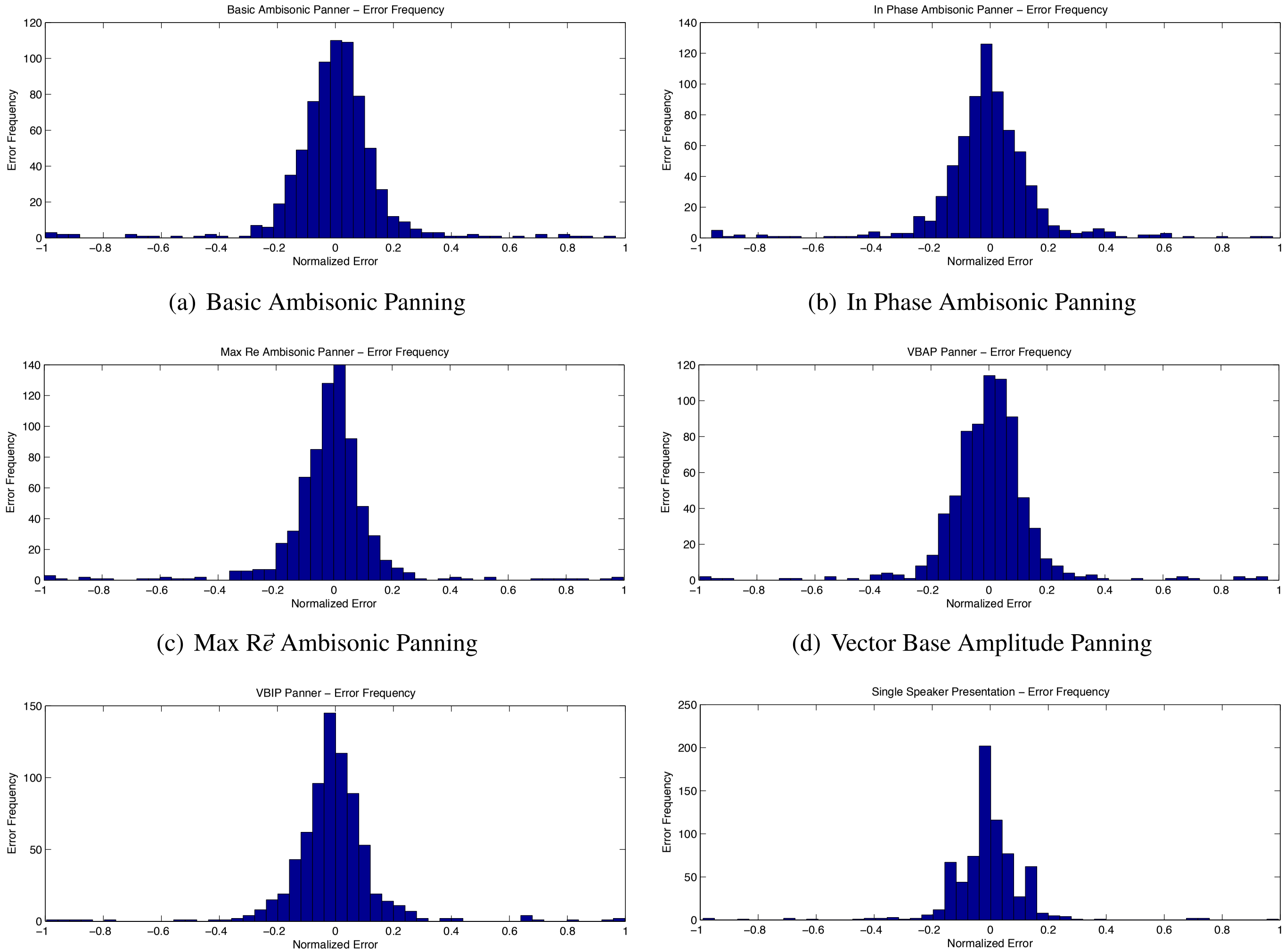


Figure 5.3: Panning Angle v. Perceptual Width: Normalized Width



(a) Basic Ambisonic Panning

(b) In Phase Ambisonic Panning

(c) Max $R\vec{e}$ Ambisonic Panning

(d) Vector Base Amplitude Panning

(e) Vector Base Intensity Panning

(f) No Panning

Figure 5.4: Distribution Of Errors For Implemented Panner Types

Chapter 6

Experiment

6.1 Introduction

Segmentation of acoustic phenomena into “objects” and “streams” is a fundamental process in the perception of our auditory environment [11]. The auditory system is capable of responding to frequency, power and timing cues, and it is these cues that help define auditory objects. Onset of an acoustic event is defined by a spectro-temporal edge – a sudden increase in amplitude at some moment in time. Onset edges are known to elicit a characteristic set of deflections in ERP data, known as the “P1-N1-P2 complex”. This series of deflections in ERP data are thought to index initial processing of stimuli in the auditory cortex. These amplitude defined edges have gained attention in computational neuroscience with recent success modelling neural response to amplitude transients [15] [16]. In the present study we examined edges created by discontinuities in otherwise smooth motion of auditory objects.

6.2 Methods

6.2.1 Participants and stimulus presentation

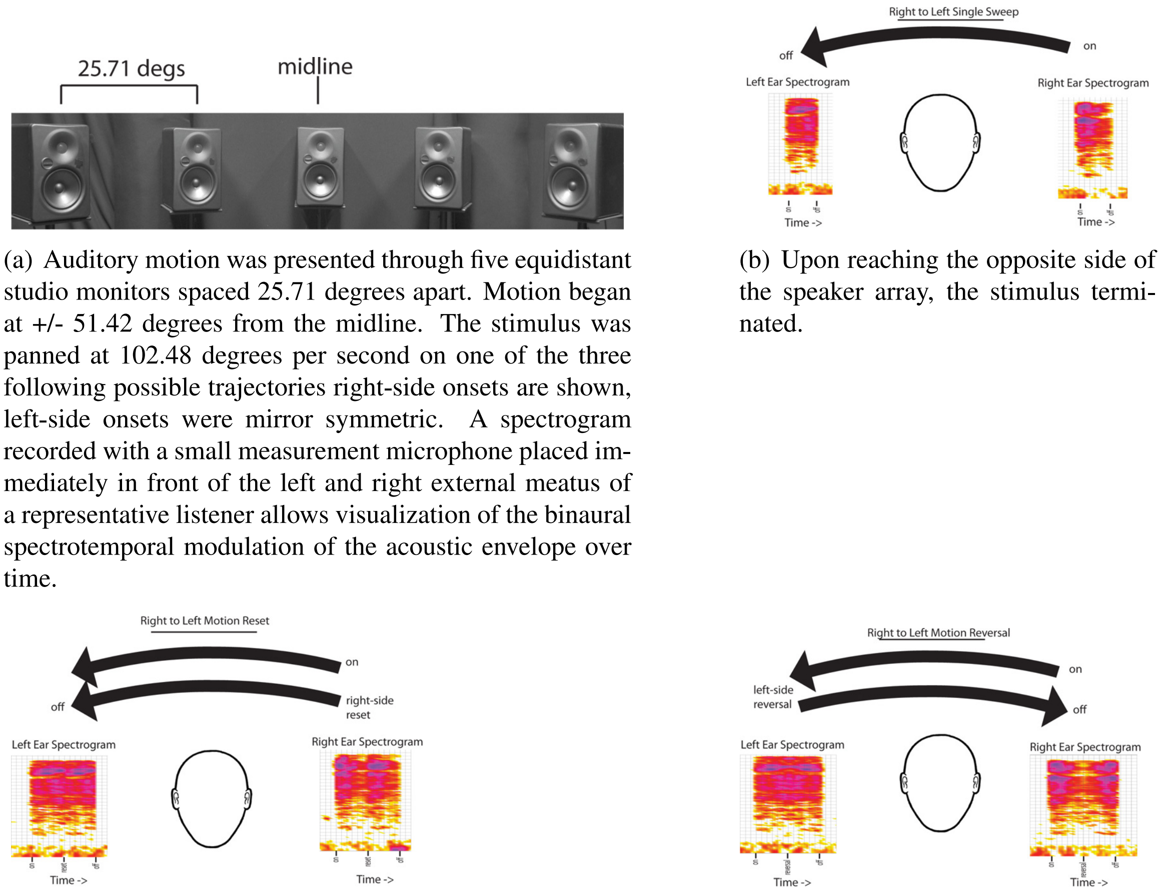
A number of students from the University of Lethbridge participated in the study in exchange for class credit. Participants were rejected on the basis of artifacts in data and for various medical conditions. Data from two male and sixteen female undergraduate students remained after the participant screening process. Participants were told that they were participating in an auditory attention experiment, and were instructed in the task, but were not told the purpose of the study until after completion of the experiment. Participants were situated in the centre of a ring of 14 Mackie HR624 MK-2 studio monitors. Studio monitors were spaced at intervals of 25.71◦, at distance of 1.27 m to the listening position. Five of the monitors were used for the present study, situated at 0 degrees (the midline), ± 25.71◦, and ± 51.42◦. Monitor gains were individually calibrated and matched to 70 dbA using Smaart 6.0 acoustical analysis software and a calibrated measurement microphone (Audix TR-40). Room reverberation was attenuated by custom fiberglass semicircular acoustic traps and sound isolating rubber wall covering. The noise floor was at 62 db SPL due almost entirely to constant ventilation background noise. The experiment was controlled by a custom Cocoa application running on a Mac Pro (Mac OS 10.6). The software framework for auditory stimulus presentation was built using Apple's "Core Audio" framework. Two M-Audio FireWire 410 devices were daisy-chained to provide an interface to the studio monitors. Participants responded to auditory stimulus on an LCD display located below the centre speaker, and were allowed to practice until they felt confident that they understood and were to perform the task.

During the experimental design, unfiltered uniform random noise presented at 70 db SPL was used as calibration stimuli. The measurement microphone was placed at the nominal location for the centre of a participants head. Pressure level and frequency dependent energy were examined under experimental conditions using Smaart 6.0, and a narrow band notch filter near 16 kHz was discovered. Consequently, stimuli consisted of uniform random noise that was generated and filtered at 14 kHz using Praat [10], and presented at 70 db. The phantom image was generated using Vector Base Amplitude Panning (VBAP)[39]. VBAP was chosen due to its low computational requirements, its relationship to the tangent panning law, and its ecological validity relative to other options.

At the beginning of each trial participants were presented with a screen displaying a fixation cross and a start button. Participants were instructed to focus on the fixation cross, and to press the start button when they were ready to begin. Participants were instructed to refrain from blinking or moving after pressing the start button until the response collection screen. Pressing the start button resulted in a 500 ms fixed delay, followed by 1500 to 2000 millisecond uniform random delay before the onset of stimulus.

Stimulus onset position was pseudo-randomly selected in the left or right field, at ±51.42○ from the midline. Stimulus motion began immediately upon onset and moved at a constant rate of 103.48 degrees per second along one of three possible trajectories, "Single Sweep", "Motion Reset" or "Motion Reversal". The Single Sweep trajectory consisted of the auditory object moving from its onset location to the opposite side of the speaker array over the course of 1 second. The Motion Reset trajectory repeated the same Single Sweep trajectory twice. During a Motion Reset trajectory the auditory object would jump from one extreme position on the auditory array to the other with no change in sound pressure level. The Motion Reversal trajectory was similar to the Motion Reset, but instead of repeating the same sweep, direction of motion was during the second sweep, returning the auditory object to its initial position. At the end of each trajectory the auditory stimulus was turned off. Each of the three trajectories was presented an equal number of times at either of the two starting locations. The six conditions (3 trajectories x 2 onset locations) were presented sixty times each over the course of the experiment. Participants were allowed a break between blocks of trials.

Sound panned along Motion Reversal trajectories impressed the perception of a single auditory object moving from one side of the auditory array to the other, then returning back to their original location. Perceptual segmentation of the auditory scene was markedly different for sounds panned along Motion Reset trajectories. In the case of Motion Reset trajectories, two independent auditory objects were perceived, coinciding with the motion reset event. This perceptual segmentation was strictly a consequence of perceived motion, rather than sound pressure level or spectral discontinuities in the source.



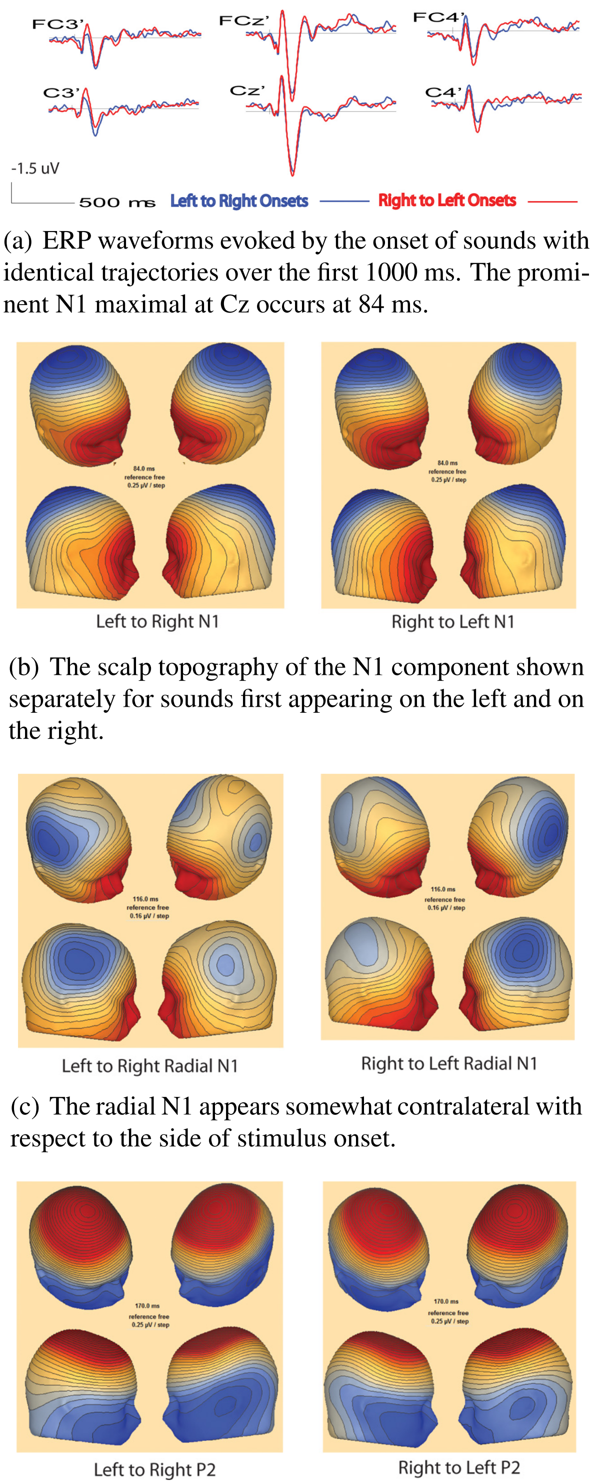
(a) Auditory motion was presented through five equidistant studio monitors spaced 25.71 degrees apart. Motion began at +/- 51.42 degrees from the midline. The stimulus was panned at 102.48 degrees per second on one of the three following possible trajectories right-side onsets are shown, left-side onsets were mirror symmetric. A spectrogram recorded with a small measurement microphone placed immediately in front of the left and right external meatus of a representative listener allows visualization of the binaural spectrotemporal modulation of the acoustic envelope over time.

(b) Upon reaching the opposite side of the speaker array, the stimulus terminated.

(c) Upon reaching the opposite side, the location reset to the original position and the motion repeated.

(d) Upon reaching the opposite side the motion reversed direction and returned to the onset location.

Figure 6.1: Auditory Object Trajectories



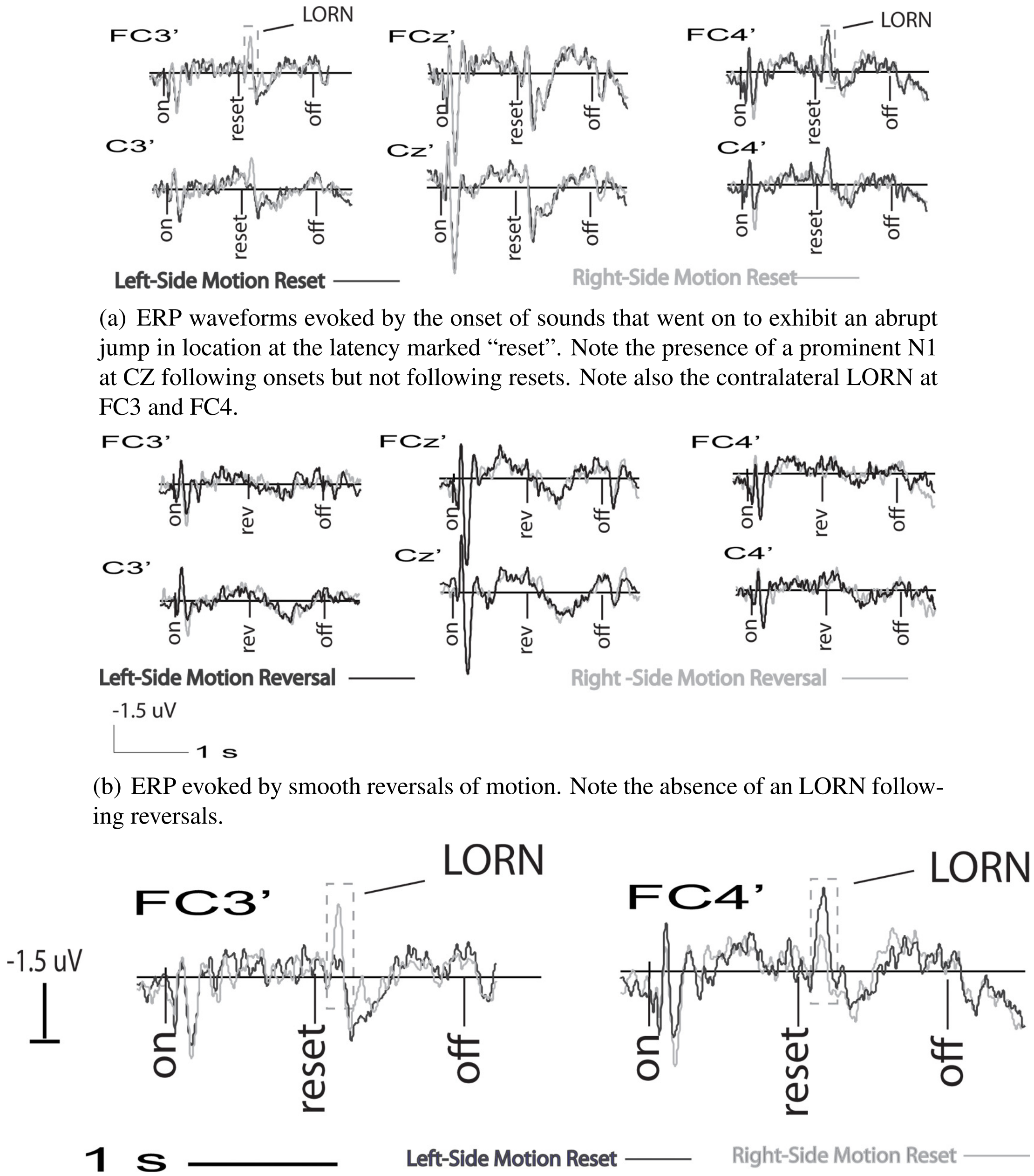
(a) ERP waveforms evoked by the onset of sounds with identical trajectories over the first 1000 ms. The prominent N1 maximal at Cz occurs at 84 ms.

(b) The scalp topography of the N1 component shown separately for sounds first appearing on the left and on the right.

(c) The radial N1 appears somewhat contralateral with respect to the side of stimulus onset.

(d) The P2 component.

Figure 6.2: Scalp Distribution Of Voltage

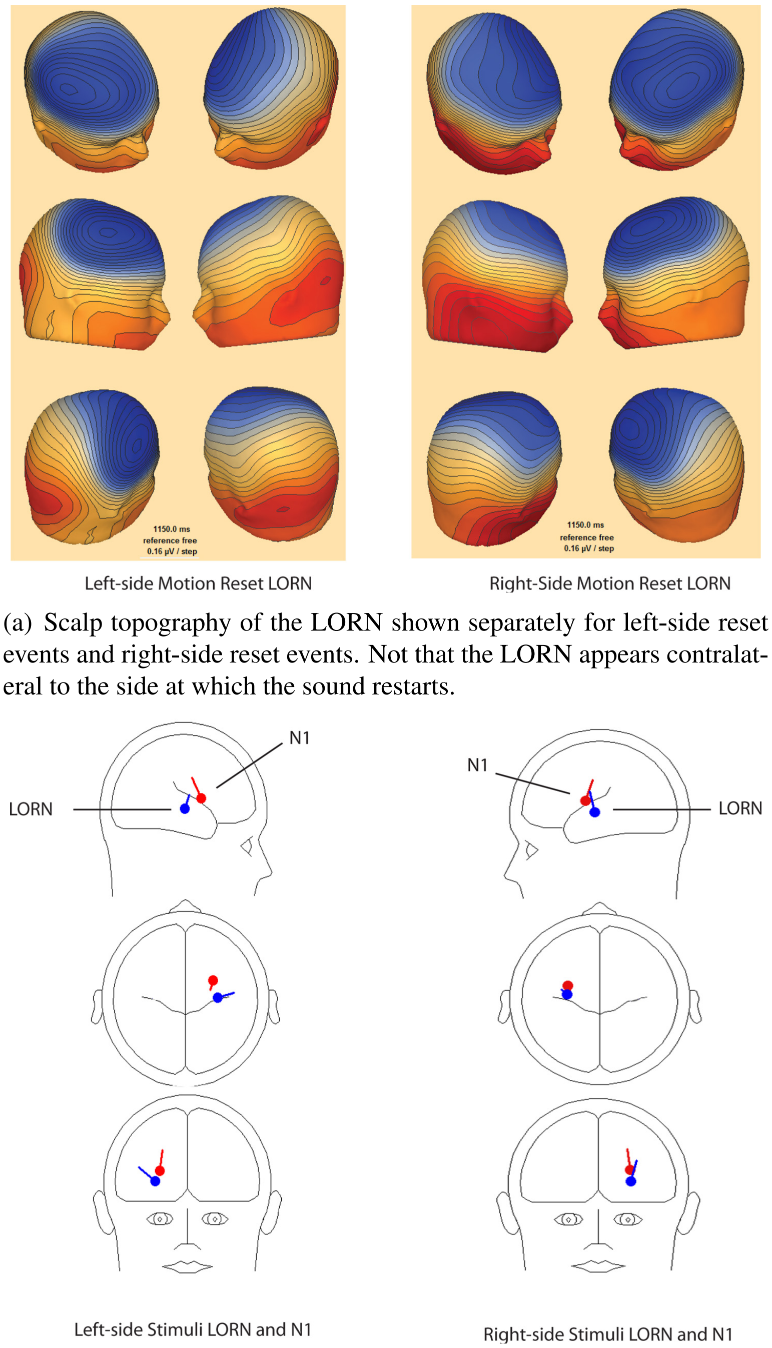


(a) ERP waveforms evoked by the onset of sounds that went on to exhibit an abrupt jump in location at the latency marked "reset". Note the presence of a prominent N1 at CZ following onsets but not following resets. Note also the contralateral LORN at FC3 and FC4.

(b) ERP evoked by smooth reversals of motion. Note the absence of an LORN following reversals.

(c) Detailed view of the LORN at fronto-lateral electrodes.

Figure 6.3: LORN Generated In Motion Reset Events


1150.0 ms
reference free
0.16 µV / step
1150.0 ms
reference free
0.16 µV / step
Left-side Motion Reset LORN
Right-Side Motion Reset LORN
N1
LORN
N1
LORN
Left-side Stimuli LORN and N1
Right-side Stimuli LORN and N1

(a) Scalp topography of the LORN shown separately for left-side reset events and right-side reset events. Not that the LORN appears contralateral to the side at which the sound restarts.

(b) The relative configurations of the N1 (red) and LORN (blue) dipoles contralateral to the side of the stimulus event. These dipoles indicate the averaged location and orientation coordinates across individual participants.

Figure 6.4: Isopotential Maps And Modelled Dipoles

6.2.2 EEG recording and analysis

Participants were fitted with an Electrical Geodesics Inc. 128 channel silver/silver-chloride electrode cap. Voltage readings were amplified using a NetAmps 200 amplifier, sampled at 500 Hz using *NetStation* (Electrical Geodesics Inc., Eugene, OR, USA), and stored for offline analysis using *BESA* (Megis Software, Grafelfing, Germany).

Impedence was measured immediately before the experiment began, and was subsequently maintained below 100 kΩ through the course of the experiment. EEG was visually inspected for per-channel artifacts. Artifacts present over a large number of channels resulted in the rejection of the data set. Artifacts in the remaining sample population were limited to a small number of channels. In these cases recorded data was replaced by an interpolated signal from surrounding electrodes. EEG was filtered with a high pass (0.5 Hz 12 dB/octave) and a low pass (30 Hz, 24 dB/octave) zero-phase Butterworth filters.

6.2.3 Construction and analysis of ERP waveforms

Since all stimuli were identical for the first 1000 ms except for the side on which they started, we first investigated the period immediately after the onset of a moving sound by computing ERP waveforms grand-averaged across all three trajectory types while preserving the distinction between left-to-right and right-to-left motion. These had a 200 ms prestimulus baseline (i.e. the epoch was -200 – +1000). To investigate and contrast the ERP evoked by reset and reversal events, we created long ERP epochs using windows beginning with a 200 ms baseline prior to the sound onset and ending at sound offset (thus 2200 ms for the Motion Reset and Motion Reversal trajectories). These epochs thus captured ERP events for sound onsets as well as resets and reversals. Epochs with ocular artifact were rejected automatically and data from two participants was discarded due to excessive eye movements, leaving 14 participants in the data set. Data from each participant was interpolated to a standard 81-channel montage prior to grand averaging to facilitate co-registration of sensors at potentially different scalp locations. Scalp topography of the ERP waveform was visualized using a spline interpolation of the grand average waveforms. We were particularly interested in the time window of 100–200 ms after a motion reset/reversal event because of the extensive prior work on the ORN and POR. We identified an N1-like negative peak during this interval (at 150 ms) in the Motion Reset condition but not during the Motion Reversal condition. Lateralization with respect to the hemifield in which the sound appeared was analyzed with a 2 hemifield (left/right) 3 electrode (FC3, FCZ, FC4) repeated-measures ANOVA using the mean amplitudes between 140 and 160 ms. Motion Reset ERPs were compared to Motion Reversal ERPs by a 2 hemifield (left/right) 3 electrode (FC3, FCZ, FC4) 2 trajectory (Reset, Reversal) ANOVA followed by paired post-hoc t-tests.

6.2.4 Electrical source analysis

Determining source localization from EEG data is an under-determined problem, providing an unbounded set of possible solutions. A number of algorithms have been proposed in an attempt to “solve” the problem, often introducing additional constraints over energy distribution or the number of sources. These constraints also imply a set of limitations specific to the constraint set used. Source localization is widely employed in EEG research, but the locations determined are not themselves informative, even when a technique has previously demonstrated a strong positive correlation to spatially sensitive neuroimaging techniques. Comparing localization parameters for different conditions can be used to indirectly evaluate similarity of scalp distributions between conditions.

We performed dipole fitting for onset and motion reset events from individual participants, and examined the resulting location and orientation parameters with a multivariate analysis of variance. Head models used in dipole modelling were provided by BESA, based on the average of 50 individual anatomical MRI scans, and transformed into standard Talairach coordinates. A bilaterally symmetric pair of dipoles was seeded at A1, and was allowed to rotate and move simultaneously, to achieve best possible fit, but only the contralateral dipoles were considered in the MANOVA. Tukeys LSD paired comparisons, uncorrected for multiple comparisons, were used to evaluate the effect of different peaks (N1 vs. LORN) on each of the six configuration parameters. In all dipole modeling, the head model was based on the average anatomical MRI scans of 50 individuals warped to the Talairach coordinate system. A conductivity ratio of 80 (most appropriate for adults) was used.

6.3 Results

Participants discriminated the kind of trajectory (Single Sweep, Motion Reset or Motion Reversal) with near perfect accuracy for all conditions.

Since the first 1000 ms of each trajectory was identical, to examine the ERP waveforms evoked by onsets we collapsed across conditions while retaining the distinction between left and right onsets. Onsets of the stimuli (which included a large instantaneous amplitude increase from background) evoked a robust auditory evoked ERP with the well-known P1-N1-P2 complex (Figure 6.2 A).

A prominent N1 wave peaked at 84 ms after stimulus onset, was maximal at fronto-central electrodes and was only slightly lateralized with respect to the hemispace in which the onset occurred (Figure 6.2 B). A radial N1 component appeared at 116 ms after sound onset with a somewhat greater magnitude over contralateral relative to ipsilateral scalp (Figure 6.2 C). The P2 appeared at 170 ms after sound onset and was not lateralized with respect to the stimulus (Figure 6.2 D). Finally, a small and temporally diffuse fronto-central negative deflection appeared in every condition at about 600 ms (or about 100 ms after the sound crossed the midline) possibly time-locked to the moment the stimulus crossed the midline of the speaker array. Of particular interest was the ERP response to the motion reset event that occurred at 1000 ms after sound onset on Motion Reset trials. This event triggered the percept of a new sound onset rather than the percept of a sudden discontinuous jump of the same sound to a different location (i.e. it was perceived as an auditory edge between two temporally distinct objects). Figure 6.3 A depicts waveforms for left-side resets and right-side resets. These events evoked a robust set of components that were notably different from those evoked by the onsets in figure 6.1 A. Given the prior interest in the ORN and POR between 100 and 200 ms, we focused on a prominent negative deflection occurring at 150 ms after the spatial transient. We refer to this below as the Lateralized Object-Related Negativity (LORN). Unlike the onset-evoked N1 at 84 ms, this peak was strongly lateralized with respect to the side on which the sound reappeared (see waveforms in figure 6.3 A and C and isopotential maps in figure 6.4 A). This lateralization is reflected in a significant stimulus side (Left/Right) by Sensor (FC3, FCZ, FC4) cross-over interaction $[F_{2,26} = 17.8; p < .001]$ for the mean amplitude of this peak (spanning $+/-$ 10 ms on either side of the peak) across participants. Neither of the main effects in this ANOVA reached significance. Motion reversal events did not trigger the percept of a new auditory object, but rather were perceived as a change in direction of a continuous object. Likewise, these events did not evoke a robust ERP (Figure 6.3 B). Comparing the prominent LORN peak (between 140 and 160 ms) evoked by motion reset stimuli to the mean amplitude over the corresponding latency window for motion reversal stimuli yielded a significant 3-way (side-sensor stimulus type) interaction $[F_{2,26} = 11.762; P < .001]$. Post-hoc paired-sample t-tests comparing each peak evoked by motion resets with its counterpart latency in the motion reversal condition revealed significant differences particularly for the contralateral sensor ($t_{13} = 0.02$ at FC3 for left events; $t_{13} = 0.0008$ at FC3 for right events; $t_{13} = 0.003$ at FCz for left events; $t_{13} = 0.167$ at FCz for right events; $t_{13} = 0.0007$ at FC4 for left events; $t_{13} = 0.966$ at FC4 for right events). Fixed bilateral A1 (Heschls Gyrus) dipoles explained 91 percent of the scalp variance for both left and right-side reset events. In both cases the contralateral A1 dipole explained most of the variance, however omission of the ipsilateral dipole resulted in a poor solution indicating that the ipsilateral cortex probably contributes signal to the scalp-measured ERP. Dipoles fixed in PT, however, could not successfully account for the LORN and left generally high residual variances. Considering the distribution of location and orientation parameters across individual participants when symmetric dipole pairs were fitted to individual data allowed us to further explore differences between the onset-evoked N1 and reset-evoked LORN peaks. The result of fitting dipoles to individual participants data and then averaging the resulting location and orientation parameters is displayed in figure 6.4. Left and right-side stimuli yielded bilaterally symmetric averaged dipoles (note this is not due to the constraint of bilateral symmetry on the fitting algorithm for each side). The LORN peak dipoles were lower, more posterior and more lateral than the N1 dipoles in both cases. For both left and right-side events the MANOVA revealed marginally significant main effects of the evoking stimulus condition (onset N1 vs. reset LORN) [Pillais Trace = 0.242, p = .062 for left-side events; Pillais Trace = 0.339, p = .023 for right-side events]. Post-hoc paired comparisons suggested that these differences were driven mainly by differences in the Z (up/down) dimension [p = .003 for left stimuli; p = .09 for right stimuli]. The N1 and LORN dipoles also differed in their orientations however this was significant only for left-side stimuli [Pillais Trace = 0.486, p = .004 for left stimuli; Pillais Trace = 0.03, p = .528 for right stimuli]. Post-hoc comparisons of orientations for left-side stimuli suggested that the y orientation parameter carried the main effect of peak [p = .005].

6.4 Discussion

6.4.1 Interpretation of the LORN

The onset of auditory stimulus is reflected in data as the “P1-N1-P2 complex”. If the LORN were a correlate of object registration it would be present as a component of this initial series of deflections during auditory onset. At the same latency as the LORN, the P2 positive deflection is starting to develop.

The stimuli used in the current study were presented over studio monitors. The Motion Reset event was characterized by a sudden changes in interaural timing and amplitude, influenced by the HRTF and subtle variations in room reverberations. These binaural changes suggest the interpretation of the LORN as a low-amplitude N1 response. These changes invoked a muted P1-N1-P2 response relative to onset, but the LORN is a larger amplitude than the N1 at fronto-lateral electrodes FC3/FC4. (Figure 6.3).

The LORN may index a change in attentional load, similar to the “Mismatch-Negativity” (MMN) – an ERP component that can be found when a stimulus varies in some dimension(s) from a previously established pattern. The classical auditory MMN is evoked only after a period of training, followed by a deviant stimuli.

Spatially deviant acoustic events are known to elicit an MMN at latencies similar to the LORN [34] [50]. Tata and Ward showed that free-field spatial deviants evoked a contralateral MMN component at 160 ms latency, while Sonnadara demonstrated HRTF localized spatial deviants produced a contralateral MMN component between 120-130 ms latency. Sonnadara showed that the amplitude of the MMN was related to the frequency of the spatial deviant, with complete attenuation of the component when the “deviant” occurred at the same frequency as the standard. The motion sweep, motion reversal, and motion reset events occurred in random order and with equal probability, which would seem to preclude an interpretation involving the classical MMN.

The "model adjustment hypothesis" states that the MMN is generated when a learned model of stimulus behaviour is no longer sufficiently predictive of sensory information, and new information must be integrated into the model [17]. A possible interpretation of the LORN is that the model of behaviour for the auditory object was violated by Motion Reset events, requiring a update to the model. The LORN may be a correlate of this adjustment, elicited an unexpected spectrotemporal "edge" in an already present stimulus. It would be instructive to attempt motion reversals at varying velocities to determine if a LORN could be generated, but the velocities required to invoke a LORN in this way may exceed other perceptual limits.

Tata and Ward proposed that the lateralized MMN may be a result of a sudden shift of attention through auditory space towards a stimulus at a previously unattended location [51]. This situation occurs in the present study at auditory onset and during Motion Reset events. Onset evoked potentials do not contain a LORN component, suggesting that neither reflexive attention reorienting, nor object registration are sufficient to elicit the response. It is possible that the changes in energy coinciding with onset and offset are sufficient to invoke the percept of a new object, and that objects defined by motion require additional processing which induce the LORN.

6.4.2 Relationship to other ERP and MEG responses

We have chosen to use the term lateralized object-related negativity because the 150 ms negative-going peak in question is evoked when spectrotemporal discontinuity in the auditory scene triggers the percept of an auditory edge between two distinct auditory objects. It consequently exhibits important similarities to the ORN described by others [3] [14] [31]: First, both the ORN and the LORN arise when the eliciting stimulus triggers the perceptual segregation of two distinct auditory objects. Second, both can be dissociated from the N1 because they do not require an abrupt energy transient (in the case of the ORN this independence is achieved by subtracting tuned- from mistuned harmonic stimuli and in the case of the LORN this independence is achieved by temporally embedding the evoking auditory edge into a continuous acoustic envelope). Third, like the LORN reported here, the ORN has a fronto-central focus and is maximal at electrodes contralateral to the mistuned harmonic [31]. Fourth, the latency of the ORN and LORN reported here are nearly identical (160 ms and 150 ms, respectively). Finally, both the ORN and the LORN can be dissociated from the classically defined MMN because they do not require that the evoking stimulus is a rare deviant.

The LORN, ORN and POR bear some similarity to a negative deflection of the auditory ERP that is evoked when a click train shifts laterality due to periodic changes in interaural time delay (ITD) [32]. That study reported an N1-like negative deflection at the vertex that occurred about 40 ms later than the N1 evoked by the onset of the click train itself. Since the percept of lateralization due exclusively to ITD requires integration and comparison of both ear inputs, whereas the onset of a sound can be registered monaurally, the authors interpreted the negative deflection as an N1 peak delayed due to the extra time required to process the binaural input relative to monaural input. The study did not record signals from fronto-lateral sites (FC3/FC4) where we found the LORN to be maximal and did not find any lateralization of the ITD-related N1 peak (they recorded only from CZ, T3 and T4). However, it is possible that the LORN we describe here is related to the negative deflection described by McEvoy et al. [32] and that the LORN could be evoked by the onset of a new perceptual object defined only by changes in ITD. The present study clearly differentiates the LORN from the N1, mainly on the basis of dense-array scalp topography and dipole modeling – approaches that were not employed by the earlier study.

There are important differences between the ORN and LORN as well: First, the ORN is defined as a difference between two ERP waveforms evoked by tuned and mistuned complex tones, whereas the LORN is readily apparent in the single ERP waveform. Second, the ORN is related to segmentation of simultaneous sounds whereas the LORN is related to segmentation of a temporal sequence of sounds. Finally, the object that gives rise to the ORN is usually defined spectrally whereas the LORN is associated with a spatially defined object. The ORN elicited by dichotic pitch stimuli [22] [24] [26] [25] [27] is an interesting exception. Although the cues that enable segregation of the dichotic pitch stimulus from the background arise from inter-aural timing differences, the percept is of a distinctly lateralized pitch. Thus the dichotic pitch paradigm may capture more than one mechanism of auditory scene segmentation.

There are also notable similarities between the LORN and the neuromagnetic POR evoked by the transition from noise to RIS noise [29] [47] [48]: The LORN and the POR are both responses to the onset of a perceptual object despite the absence of a sharp first-order energy transient; both occur at about 150 ms latency and both are differentiable from earlier responses triggered by energy onsets. Furthermore, the POR was found to be slightly anterior, medial and inferior to the generator of the N100m. A prominent theory regarding the functional anatomy of auditory cortex suggests that a functionally distinct what pathway for non-spatial information (akin to the ventral pathway in visual cortex) extends anterior to primary cortex [2] [5] [44]. The ventral pathway in visual cortex is known to subserve object identification and it may be that the registration of new auditory objects, even when defined by spatial information, triggers activity in this anterior auditory pathway. Further research will be necessary to determine if the spectral ORN and POR, and the spatial LORN do in fact reflect functionally related brain mechanisms.

Chapter 7

Summary

We have reviewed salient perceptual cues for directional hearing, and details of a number of panning algorithms designed to take advantage of perceptual mechanisms. These algorithms drove the development of a hardware and software system suitable for presenting auditory stimulus in both behavioural and EEG experiments according to the criteria set forth in chapter 4. The EEG interfacing library *libnetstation*, which was developed and released as an open-source project, has been shared with other research groups which employ the EEG technique, and has served as the basis of a python version of the library known as *pynetstation*.

Perceptual direction was compared to rendered direction for the three ambisonic panning variants (Basic, In-Phase and Max R$\vec{e}$), and two vector-base methods (vector base intensity panning, and vector base amplitude panning). The distribution of error for vector-base techniques had higher kurtosis than ambisonic panning techniques, but this trend was not statistically significant with the number of subjects in our study. In-phase panning produced a wider sound that other techniques, a consequence predicted by its mathematical formation.

The vector-base amplitude panning technique was chosen for an ERP study of perception of sound in motion. Sounds were put in motion along one of three paths – single sweep, motion reversal, and motion reset. Motion reset events evoked a previously unreported ERP component we have named the lateralized object-related negativity (LORN). The LORN may indicate processing of second order "velocity" boundaries, prior to the formation of a gestalt in the auditory system.

Chapter 8

Appendix

8.1 Papers, Posters And Presentations

The following documents made use of the *libnetstation* library.

1. Tata, Matthew S. (2010) In Search of the Neural Correlates of Gambling: Evidence From Human Neuroimaging. Behind The Mask: A Symposium On Women Problem Gamblers. Banff, AB

2. Tata, Matthew S. (2010) Atypical Frontal Cortex Activity in Early-Stage Problem Gamblers (2011) Hotchkiss Brain Institute Visiting Lecture Series, Calgary, AB

3. Butcher, A., Govenlock, S. and Tata, M.S. (2010) A lateralized auditory evoked potential elicited when auditory objects are defined by spatial motion. Hearing Research.

4. Tata, M.S., Alam, N., Mason, A.L.O, Christie, G.J. and Butcher, A. (2010) Selective Attention and Optic Flow Interact in Human MT/MST. Vision Research. 50, 750 760.

5. Christie, G. J., and Tata, M.S. (2009) Risk-Taking in a Gambling Task Increases Theta-Band Oscillatory Activity in Right Frontal Cortex. Neuroimage. 48, 415 - 422.

6. Oberg, S. and Tata, M.S. (2011) Atypical frontal activity is exhibited in early-stage problem gamblers. Talk presented at the Canadian Spring Conference on Brain and Behaviour, Fernie, BC.

7. Ponjavic, K. and Tata, M.S. (2011) Distraction Decoherence: a correlate of attentional distraction in the dynamics of the human auditory system. Talk presented at the Canadian Spring Conference on Brain and Behaviour, Fernie, BC.

8. Dowdall, J. and Tata, M.S. (2010) Neural mechanisms of failed perception in object substitution masking. Talk presented at the Canadian Spring Conference on Brain and Behaviour, Fernie, BC.

9. Dowdall, J. R. and Tata, M. S. (2010) An electrophysiological investigation of effective and ineffective masks in object substitution masking. Poster presented at the annual meeting of the Cognitive Neuroscience Society, Montreal, Canada.

10. Christie, G. J. and Tata, M. S. (2010) Anterior cingulate undergoes theta-band phase locking with right frontal cortex during feedback processing in a gambling task. Poster presented at the annual meeting of the Cognitive Neuroscience Society, Montreal, Canada.

11. Christie, G. J. and Tata, M. S. (2010). Theta phase locking within human frontal cortex during Gambling. Talk presented at the Canadian Spring Conference on Brain and Behaviour, Fernie, BC.

12. Ponjavic, K. Kalynchuk, M. and Tata, M. S. (2010). Auditory distraction in ADHD doesnt depend on distractors? Talk presented at the Canadian Spring Conference on Brain and Behaviour, Fernie, BC.

13. Dowdall, J., and Tata, M. S. (2010). Neural Mechanisms of Failed Perception in Object Substitution Masking. Talk presented at the Canadian Spring Conference on Brain and Behaviour, Fernie, BC.

14. Christie, G. J., Butcher, A. and Tata, M. S. (2009) Risk taking in a gambling task increases oscillatory theta-band activity in right medial frontal cortex. Poster presented at the annual meeting of the Cognitive Neuroscience Society, San Francisco, CA.

15. Christie, G.J. and Tata, M.S. (2009) Tracking theta-band neural activity in frontal cortex during feedback processing in a gambling game. Talk presented at the Canadian Spring Conference on Brain and Behaviour, Fernie, BC.

16. Dowdall, J., Kalynchuk, M. and Tata, M.S. (2009) The effect of distraction on low-level auditory processing: Evidence from Auditory ERP. Talk presented at the Canadian Spring Conference on Brain and Behaviour, Fernie, BC.

17. Scott A.K. Oberg, Gregory J. Christie, Andrew Butcher, Matthew S. Tata (2011). Problem Gamblers Exhibit Atypical Reward Processing in Frontal Cortex Following Feedback During Gambling. Poster presentation at the Cognitive Neuroscience Society 18th Annual Meeting.

18. Dowdall, J.R., Luczak, A., and Tata, M.S. (2011). Visual search for a popout target increases induced theta power over contralateral visual cortex. Poster presented at the Cognitive Neuroscience Society Annual Meeting, San Fransisco, CA.

19. Ponjavic, K. Dowdall, J.R. and Tata, M.S. (2011). Electrophysiological Correlates of Auditory Distraction as Manifested in Post-secondary Adults with Attention Deficit Hyperactivity Disorder. Oral presentation at the Hopewell Professorship Alberta Imaging Symposium, University of Calgary, Health Sciences Centre

20. Ponjavic, K.D. and Tata, M.S. (2011). Electrophysiological correlates of auditory distraction in normal listeners and listeners with Attention Deficit Hyperactivity Disorder. Poster presentation at the Cognitive Neuroscience Society, 18th Annual Meeting, San Francisco, California

Bibliography

[1] Intel 64 and ia-32 architectures software developers manual: Volume 3 (3a 3b): System programming guide. 3:1–26, 2011.

[2] C. Alain, S. R. Arnott, S. Hevenor, S. Graham, and C. L. Grady. ”what” and ”where” in the human auditory system. *Proc Natl Acad Sci U S A*, 98(21):12301–6, 2001. 0027-8424 (Print) Journal Article.

[3] Claude Alain, Benjamin M Schuler, and Kelly L McDonald. Neural activity associated with distinguishing concurrent auditory objects. *J. Acoust. Soc. Am*, 111(2):990–995, feb 2002.

[4] JS Bamford and J Vanderkooy. Ambisonic sound for us. *Journal of the Audio Engineering Society -Preprint*, 1995.

[5] P. Belin and R. J. Zatorre. ’what’, ’where’ and ’how’ in auditory cortex. *Nat Neurosci*, 3(10):965–6, 2000. Comment Letter Review United states.

[6] Jacob Benesty. *Springer Handbook of Speech Processing*. Springer, 1 edition, dec 2007.

[7] JC Bennett and K Barker. A New Approach to the Assessment of Stereophonic Sound System Performance. *Journal Of The Audio Engineering Society*, 1985.

[8] S Bertet, J Daniel, L Gros, and E Parizet. Investigation Of The Perceived Spatial Resolution Of Higher Order Ambisonics Sound Fields: A Subjective Evaluation Involving Virtual And Real 3D Microphones. *30th AES Int. Conference*, jan 2007.

[9] Jens Blauert. *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization*. The MIT Press, rev sub edition, oct 1996.

[10] David Boersma, Paul Weenink. Praat: doing phonetics by computer [computer program]. version 5.2.44, retrieved 23 september 2011 from http://www.praat.org/.

[11] Albert S Bregman. *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press, sep 1994.

[12] Andrew Butcher, Stanley W Govenlock, and Matthew S Tata. A lateralized auditory evoked potential elicited when auditory objects are defined by spatial motion. *Hearing research*, 272(1-2):58–68, feb 2011.

[13] J Daniel, R Nicol, and S Moreau. Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging. *Journal of the Audio Engineering Society -Preprint*, 2003.

[14] B. J. Dyson and C. Alain. Representation of concurrent acoustic objects in primary auditory cortex. *J Acoust Soc Am*, 115(1):280–8, 2004. 0001-4966 (Print) Journal Article Research Support, Non-U.S. Gov't.

[15] A Fishbach, I Nelken, and Y Yeshurun. Auditory edge detection: a neural model for physiological and psychoacoustical responses to amplitude transients. *Journal of neurophysiology*, 85(6):2303–2323, jun 2001.

[16] Alon Fishbach, Yehezkel Yeshurun, and Israel Nelken. Neural model for physiological responses to frequency and amplitude transitions uncovers topographical order in the auditory cortex. *Journal of neurophysiology*, 90(6):3663–3678, dec 2003.

[17] Marta I Garrido, James M Kilner, Klaas E Stephan, and Karl J Friston. The mismatch negativity: a review of underlying mechanisms. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 120(3):453–463, mar 2009.

[18] Michael A. Gerzon. General metatheory of auditory localisation. In *Audio Engineering Society Convention 92*, 3 1992.

[19] S Getzmann, J Lewald, T Geyer, HJ Muller, S Ghorashi, LN Jefferies, SM Givens, and DB Boles. Localization of moving sound. *Perception and Psychophysics*, 69(6):1022, 2007.

[20] M Gröhn. Localization Of A Moving Virtual Sound Source In A Virtual Room, The Effect Of A Distracting Auditory Stimulus. *International Conference on Auditory Display*, jan 2002.

[21] H Haas. The influence of a single echo on the audibility of speech. *J. Audio Eng. Soc*, 1972.

[22] M. J. Hautus and B. W. Johnson. Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch. *J Acoust Soc Am*, 117(1):275–80, 2005. 0001-4966 (Print) Clinical Trial Journal Article Randomized Controlled Trial.

[23] Maurice Herlihy and Nir Shavit. *The Art of Multiprocessor Programming*. Morgan Kaufmann, 2008.

[24] B. W. Johnson, M. Hautus, and W. C. Clapp. Neural activity associated with binaural processes for the perceptual segregation of pitch. *Clin Neurophysiol*, 114(12):2245–50, 2003. 1388-2457 (Print) Journal Article Research Support, Non-U.S. Gov't.

[25] B. W. Johnson, M. J. Hautus, D. J. Duff, and W. C. Clapp. Sequential processing of interaural timing differences for sound source segregation and spatial localization: Evidence from event-related cortical potentials. *Psychophysiology*, 44(4):541–51, 2007. 0048-5772 (Print) Journal Article.

[26] B. W. Johnson, M. J. Hautus, A. L. Hayns, and B. M. Fitzgibbon. Differential cortical processing of location and pitch changes in dichotic pitch. *Neuroreport*, 17(4):389–93, 2006. 0959-4965 (Print) Journal Article.

[27] B. W. Johnson, S. D. Muthukumaraswamy, M. J. Hautus, W. C. Gaetz, and D. O. Cheyne. Neuromagnetic responses associated with perceptual segregation of pitch. *Neurol Clin Neurophysiol*, 2004:33, 2004. 1526-8748 (Electronic) Journal Article.

[28] J Jot, V Larcher, and J Pernaux. A comparative study of 3-D audio encoding and rendering techniques. *Proceedings of the AES 16th international conference*, jan 1999.

[29] K Krumbholz, R D Patterson, A Seither-Preisler, C Lammertmann, and B Lütkenhöner. Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral cortex (New York, NY : 1991)*, 13(7):765–772, jul 2003.

[30] R Litovsky, H Colburn, and W Yost. The precedence effect. *The Journal of the Acoustical Society of America*, jan 1999.

[31] K. L. McDonald and C. Alain. Contribution of harmonicity and location to auditory object formation in free field: evidence from event-related brain potentials. *J Acoust Soc Am*, 118(3 Pt 1):1593–604, 2005. 0001-4966 (Print) Journal Article.

[32] L. K. McEvoy, T. W. Picton, S. C. Champagne, A. J. Kellett, and J. B. Kelly. Human evoked potentials to shifts in the lateralization of a noise. *Audiology*, 29(3):163–80, 1990. McEvoy, L K Picton, T W Champagne, S C Kellett, A J Kelly, J B Research Support, Non-U.S. Gov't Switzerland Audiology : official organ of the International Society of Audiology Audiology. 1990;29(3):163-80.

[33] J Middlebrooks and D Green. Sound localization by human listeners. *Annual Review of Psychology*, jan 1991.

[34] R Näätänen, P Paavilainen, T Rinne, and K Alho. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 118(12):2544–2590, dec 2007.

[35] M Neukom. Ambisonic Panning. *aes.org*.

[36] M Neukom. Decoding Second Order Ambisonics to 5.1 Surround Systems. *AES 121st Convention, San Francisco, CA, USA*, 2006.

[37] SH Nielsen. Auditory distance perception in different rooms. *Audio Engineering Society Preprint*, 1991.

[38] J Pernaux, P Boussard, and J Jot. Virtual Sound Source Positioning and Mixing in 5.1 Implementation on the Real-Time System Genesis. *Proc. Conf. Digital Audio Effects (DAFx-98).*

[39] V Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal Of The Audio Engineering Society*, 45(6):456–466, 1997.

[40] V Pulkki. Uniform Spreading Of Amplitude Panned Virtual Sources. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, page 4, oct 1999.

[41] V Pulkki. Localization of amplitude-panned virtual sources. II: Two- and three-dimensional panning. *Journal Of The Audio Engineering Society*, 49(9):753–767, 2001.

[42] V Pulkki. Spatial sound generation and perception by amplitude panning techniques. 2001.

[43] R Ramamoorthi. An efficient representation for irradiance environment maps. *SIGGRAPH '01 Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001.

[44] JP Rauschecker. Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 2000.

[45] L Rayleigh. On our perception of sound direction. *Philosophical magazine*, 13:214–232, 1907.

[46] Franz Schwabl. *Quantum mechanics*. Springer Verlag, 2007.

[47] A Seither-Preisler, K Krumbholz, R D Patterson, S Seither, and B Lütkenhöner. Interaction between the neuromagnetic responses to sound energy onset and pitch onset suggests common generators. *The European journal of neuroscience*, 19(11):3073–3080, jun 2004.

[48] A. Seither-Preisler, R. D. Patterson, K. Krumbholz, S. Seither, and B. Lutkenhoner. From noise to pitch: transient and sustained responses of the auditory evoked field. *Hear Res*, 218(1-2):50–63, 2006. 0378-5955 (Print) Journal Article Research Support, Non-U.S. Gov't.

[49] PP Sloan and J Kautz. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *SIGGRAPH '02 Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 2002.

[50] R Sonnadara, C Alain, and L Trainor. Effects of spatial separation and stimulus probability on the event-related potentials elicited by occasional changes in sound location. *Brain Research*, 1071(1):175–185, feb 2006.

[51] MS Tata and LM Ward. Spatial attention modulates activity in a posterior. *Neuropsychologia*, 43(4):509–516, 2005.

[52] DeLiang Wang and Guy J Brown, editors. *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Wiley-IEEE Press, sep 2006.

[53] E Wenzel. Effect of increasing system latency on localization of virtual sounds. *Audio Engineering 16th International Conference on Spatial Sound Reproduction*, jan 1999.

[54] F Wightman. Headphone simulation of free-field listening. I: Stimulus synthesis. *J. Acoust. Soc. Am*, jan 1989.