CHARACTERIZATION OF THE HEAD-RELATED TRANSFER FUNCTION USING CHIRP AND MAXIMUM LENGTH SEQUENCE EXCITATION SIGNALS

BY

JEREMY A. TIDEMANN

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Adviser:

Associate Professor Mark A. Hasegawa-Johnson

ABSTRACT

Both chirp (or sweep) and maximum length sequence (MLS) excitation signals are used to obtain measurements of the head-related transfer function (HRTF) for the Knowles electronic manikin for acoustic research (KEMAR), resulting in remarkably low relative error between the two measurement methods. Due to the low relative error, it is asserted that applying both methods in experiments involving system characterization can be used to confirm the accuracy of individual measurements and reduce the likelihood that erroneous measurements go unnoticed.

An MLS burst-augmented excitation signal is presented to compensate for unpredictable delay added in sound-acquisition software that lacks the capability of synchronized record/playback.

The HRTF measurements obtained with an external probe microphone are compared with those obtained with the standard in-ear microphones that the KEMAR is equipped with, to reveal that care should be taken to select a probe microphone that has a low susceptibility to multipath interference.

An experiment is performed to investigate the feasibility of using two prefiltered chirp responses to obtain the relative transfer function between the two locations of measurement.

TABLE OF CONTENTS

CHAPTER 1

BACKGROUND

This chapter will introduce the reader to what is known as the head-related transfer function (HRTF), followed by a discussion of several pertinent fundamentals of acoustics and digital signal processing that will be referenced in subsequent chapters.

1.1 Introduction to the HRTF

As sound propagates from an acoustic source to the ear of a listener, it undergoes a complicated acoustic transformation. The location of the source, physical aspects of the environment, and physiological properties of the listener's body all contribute to the uniqueness of this transformation. Furthermore, since sound must take a different path to reach either the left or the right ear, this transformation is unique for each ear. The general features that are most commonly considered to influence the transformation of sound are an individual's head, torso, and external ear.

The size and shape of the external ear, or pinna, greatly influences this sound transformation. This effect of the pinna is of particular importance because as its shape varies from individual to individual so does the overall HRTF [1]. The position of the source with respect to the listener is characterized by *azimuth* (horizontal angle), *elevation* (vertical angle), and *range* (distance from ear). For a given angle of incidence, the specific way in which frequencies are scaled and delayed before arriving at the eardrum can actually provide the brain with important information to discern where the sound originated, with relatively high accuracy.

If a sound is passed through an artificial HRTF and played for a listener through headphones, the listener can experience an *externalized* perception of that sound, tricking the brain into believing the sound came from the di-

rection corresponding to the HRTF. In the field of psychoacoustics, research is being conducted to determine what *cues* the brain uses to arrive at a conclusion regarding the position of a sound source. While it is not exactly known what factors the brain takes into account, research has led scientists to group cues into two groups: binaural difference cues, primarily for localization in the horizontal plane, and spectral cues, primarily for localization in the vertical plane.

An example of a binaural difference cue is interaural time difference (ITD), which is the difference between the arrival times of a sound at the left and right ears. The magnitude of the HRTF with respect to frequency is considered a spectral cue.

This document will discuss how this acoustic transformation can be measured using standard audio equipment, and the considerations that should be taken into account for accurate measurement of the HRTF.

1.2 Overview of Human Auditory System

The human auditory system is the final link in the communication chain for applications seeking to immerse a listener in a virtual auditory soundscape. As more becomes known about how the auditory system receives, manipulates, and processes acoustic signals, advancements can be made in the engineering of products designed to deliver sound for virtual applications. Each of the subsystems within the auditory system has its own unique physical characteristics, for which complex mathematical models have been created that seek to explain the physics of sound propagation. Many of the physical and mathematical interdependencies between physical subsystems of the auditory system are too complex to explore in this paper. However, the following section provides a superficial explanation of the external ear, middle ear, internal ear, and the brain in an attempt to develop a broad and general understanding of the system auditory system as a whole.

1.2.1 The External Ear

As sound propagates from a free-field source to the eardrum of a listener, it must pass through the external ear, consisting of the pinna and the ear canal,

or concha. The pinna is irregularly shaped and varies from listener to listener. However, it serves an important role in filtering the sound incident on the ear. In fact, the pinna can provide a different acoustic frequency response, depending on the specific angle of an incident plane wave. It is commonly accepted that this filter from the pinna and concha provides the brain with additional aural cues that help a listener to determine the elevation of a sound source. Research has also shown that because the human is equipped with two ears, the brain is capable of using interaural difference cues such as interaural time difference to determine the horizontal or *azimuthal* location of a sound source.

1.2.2 The Middle Ear

Once an acoustic wave makes its way through the pinna and concha, it becomes incident on the eardrum and enters the part of the auditory system called the middle ear. It is here that the acoustic wave changes mediums as the eardrum transfers the acoustic energy from air into a system of three small bones or ossicles: the malleus, incus, and stapes (also the smallest bone in the human body). As the acoustic signal travels through these bones, it is amplified until it reaches the junction between the stapes and the oval window.

1.2.3 The Internal Ear

At the oval window, the acoustic signal changes mediums again from bone to fluid as it enters the internal ear. The oval window is at the base of an extremely complex and highly tuned physical organ called the cochlea. The cochlea is shaped like a spiral that makes approximately three rotations before it reaches the helicotrema at the cochlear apex. A cross-section of the cochlea is shown in Figure 1.1.

If the cochlea were rolled out, it would look something like a long acoustic tube with a horizontal divider called the basilar membrane along the center, except at the cochlear apex. As the basilar membrane stretches from the cochlear base to the cochlear apex, its thickness gradually decreases. This varying thickness of the basilar membrane in conjunction with its distance

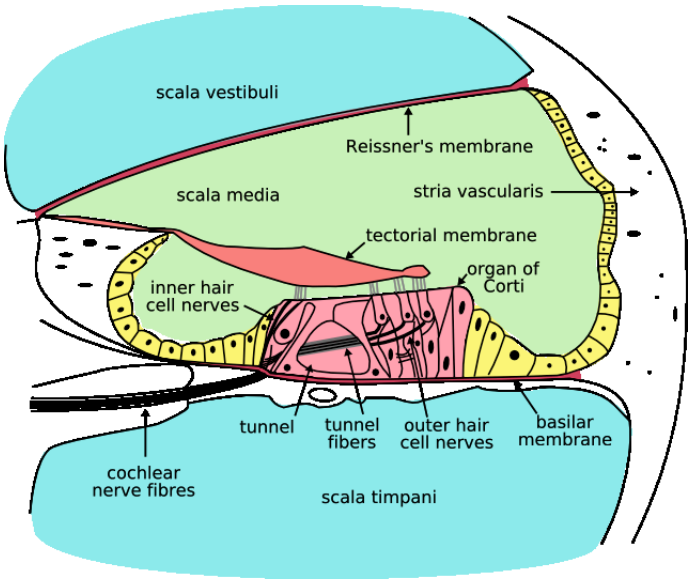

Figure 1.1: Cross-section of cochlea

from the base enables it to act like a transmission line, with finely tuned resonances at specific frequencies. High frequencies resonate at the base of the cochlea, whereas the lowest frequencies resonate near the apex.

When a frequency resonates at a specific location within the cochlea, it stimulates the inner hair cell nerves. These inner hair cell nerves are efferent, which means that they direct an electrical signal to the central nervous system. The outer hair cells have a primarily afferent function, which means that they receive signals from the central nervous system. The function of the outer hair cells is to provide feedback that "lowers the volume" that is allowed to stimulate the inner hair cells when the sound volume increases over a long period of time.

The term *firing rate* is used to describe the rate at which the inner hair cells along the cochlea discharge an efferent electrical signal. Every location along the cochlea has a unique *tuning curve* which is the sound pressure level (SPL) required to excite the inner hair cells at a particular frequency. The point at which this curve is minimum is the frequency which requires the least power to excite the inner hair cell.

The firing rate at this location does not, however, increase linearly with pressure or SPL. In fact, there are several different synaptic connections (around 15) for each inner hair cell, and each each synapse begins to fire at

relatively different sound pressure levels. It is thought that this nonlinear relationship between SPL and the firing rate has to do with the perceived loudness of a sound source. A discussion of loudness is left for the next section.

1.2.4 Perception of Loudness

The concept of loudness has been investigated since around the 1840s. Loudness was a difficult thing to measure, because it is based on *perception* rather than a physical unit. The more powerful the sound source is, and the closer it is to the ear, the greater the intensity of sound will be at the eardrum. It is obvious to assume that the higher this intensity is, the *louder* the sound will be perceived. However, exactly how the perception of loudness is related to sound intensity was the question that prompted Weber to seek what is now called the *loudness function*. An accurate representation of this loudness function became the focus of research for many years to come.

Around 1840 Weber came up with an idea for how to quantify loudness. He had been working on this problem for some time; and one night, the answer happened to come to him in his sleep.

Weber realized that the solution to such a function would rely on measurements based on perceptual experiments in which a listener would indicate whether sound "one" was louder than sound "two." If the differences in intensities of the first and second sounds were perceivable, then Weber sought to find the smallest intensity difference that a listener would be able to discern. This difference in intensity was called the intensity just-noticeable difference (JND). Later Fechner made the assumption about the loudness function, that between each intensity JND was a *constant* difference in loudness, ΔL. The catch here is that the difference in intensity ΔI was found to increase as the intensity increased, resulting in a function that tends towards a maximum constant loudness as $\Delta I_{JND} \to \infty$. This finding makes sense, as the derivative of the loudness function can be approximated by

$$\frac{\Delta L}{\Delta I_{JND}} \to 0 \quad as \quad \Delta I_{JND} \to \infty \tag{1.1}$$

The actual experiments that Weber conducted calculated ΔI_{JND} by introducing additive white Gaussian noise (AWGN) with a known variance (in

intensity) to a tone with a known intensity. This introduction of noise effectively shifted the probability distribution of the sound intensity for sound "one" around mean, μ_{I1}, with variance, σ_{I1}. A second sound with a higher average intensity, μ_{I2}, but with variance, $\sigma_{I2} = \sigma_{I1}$ was also played. The listener would then choose whether the first or second sound seemed louder, making this an intensity forced-choice (IFC) experiment.

Fechner proposed a method of using such an IFC experiment to count the number of JNDs over an interval, N_{JND}, which would then be capable of producing a representation of the loudness function.

$$N_{JND} = \int_{I_1}^{I_2} \frac{\mathrm{d}I}{\Delta L(I)} \tag{1.2}$$

Under his assumption, ΔL was a constant for each JND, he was able to simplify Equation (1.2).

$$N_{JND} = \int_{I_1}^{I_2} \frac{\mathrm{d}I}{\Delta L} \approx \frac{1}{\Delta L} \int_{I_1}^{I_2} \mathrm{d}I \tag{1.3}$$

This assumption led Fechner to find a loudness function based on integrating over the JNDs, resulting in the function

$$L_1 - L_0 = \frac{1}{C} \log\left(\frac{I_2}{I_1}\right). \tag{1.4}$$

It was later determined that the assumption of a constant ΔL was false. It was Harvey Fletcher who first devised an experiment that more accurately measured the loudness function (he was also the first person to measure the charge of an electron). In his experiment, he sought to find loudness curves of $L(I)$ and $2L(I)$. He did this to determine when a sound perceived as twice as loud. By collecting experimental data, and comparing the difference of intensity between equal loudness of the two functions, he was able to determine the following relationship

$$L(I) \; \alpha \; I^{1/3} \; , \tag{1.5}$$

which is now commonly referred to as Steven's Law after his 1957 publication on the psychophysical law. [2]

The later chapters of this document rely heavily on various fundamental principles in both acoustics and digital signal processing (DSP). For this

reason, it is pertinent to develop a strong background concerning the fundamental principles of acoustics, such as the propagation of sound, and to review several digital signal processing techniques.

1.3 Pressure, Volume, and Temperature

The means by which sound propagates in an acoustic environment is an adiabatic process, in which energy is transferred from particle-to-particle within air. The knowledge of acoustics has greatly advanced over the past century, after the discovery of the atom. To name all of the key players in the development of modern-day acoustics is beyond the scope of this paper. However, some names will be linked to key concepts.

Avogadro had hypothesized that equal volumes of gases, at the same temperature and pressure, contain the same number of molecules. It wasn't until much later that the actual number was discovered by Loschmidt to be

$$N_A = 6.02214199 \times 10^{23} \ , \tag{1.6}$$

and was named Avogadro's number (also equal to one mole).

So if air consists of such a great number of small gaseous particles, where does sound come from? In order to address this question, it is important to gain a greater understanding of the underlying physical system. Imagine a small box containing a finite number of gas molecules, each molecule having some kinetic energy along its three degrees of freedom. Although Einstein proved that, under extreme circumstances, conservation of energy breaks down, it can be reasonably assumed that energy is in fact conserved within this gaseous system. Each air molecule is free to move, as well as, to bump and bounce off of other air molecules within the box, as long as energy is conserved. The concept of pressure arises if we look at what is happening at the boundaries of the box.

Assuming the box is rigid, if a molecule is to bounce off of it, the box must supply an impulse that is equal and opposite the impulse experienced by the molecule, in order to maintain the energy of the system. From classical mechanics, the impulse required to reverse the molecule's normal velocity component would be equal to the integral over time of the force exerted by

the boundary or,

$$Impulse = \int_{t_{contact}}^{t_{depart}} f(t)\ \mathrm{d}t\ , \tag{1.7}$$

where f is the force. Now if a great number of air molecules collide with the boundary of the box, the rigid side of the box must exert a total force over the time for the cumulative molecular collisions. This force normalized over the area of the box's side becomes the pressure that it would have to exert at any point in time to keep the air molecules inside. Even though this pressure might be changing erratically as random molecules collide, the average pressure over a small time frame is relatively constant. This finding makes sense when one thinks of the extremely high number of molecules in the box, and their relative probability of hitting the side of the box.

So what would happen to the pressure if we could speed up each molecule, and add to the overall kinetic energy of the system? It seems obvious from the impulse Equation (1.7) that the instantaneous force required to keep the molecules inside the box would increase, which would increase the average force applied by the boundary of the box, ultimately resulting in increased pressure.

Now what happens if the the volume of this box were reduced, keeping the same molecules inside? The first assumption is that the energy must still be conserved. So in order to collapse the box, an excess pressure must be applied to the boundary. Now over the course of each impulse, the molecule would experience a force slightly higher than what is required to simply reverse its motion. In fact, the average molecule would return from the boundary with an increased kinetic energy. So decreasing the volume of the box results in an increase in overall kinetic energy and a decrease in area, which would cause the pressure to increase. Similar reasoning can be used to explain how pressure is reduced when the volume increases.

Boyle was the first to formalize this relationship between the pressure and volume of a gas with Boyle's Law,

$$PV = constant\ , \tag{1.8}$$

where P is pressure, and V is volume.

It was Boltzmann, however, who took this a step further, and related pressure, volume, temperature, and the number of molecules in the system,

N, with his equation.

$$PV = NkT \tag{1.9}$$

Equation (1.9), includes Boltzmann's constant, which relates temperature to the average kinetic energy of a molecule by

$$k = 1.3806505(24) \times 10^{-23} \left[\frac{joules}{kelvin}\right] \tag{1.10}$$

It was also determined that Boyle's Law was not entirely accurate, because Equation (1.8) didn't take temperature into account. It was later found that for adiabatic expansion (expansion sufficiently rapid to change the temperature of the gas), it could be corrected by

$$PV^{\gamma} = constant\ , \tag{1.11}$$

where γ is the ratio of specific heats.

$$\begin{aligned} \gamma &= \frac{C_P}{C_V} \approx 1.402 \text{ in air} \\ C_P &= \text{ molar specific heat under constant pressure} \\ C_V &= \text{ molar specific heat under constant volume} \end{aligned} \tag{1.12}$$

The constant, γ, also has significance in relating air volume and pressure within an enclosed space to the acoustic compliance of that space. The acoustic compliance, denoted C here, is a significant property in acoustics that has similar properties to electrical capacitance and can be found by

$$C = \frac{V_o}{\gamma P_o}\ . \tag{1.13}$$

1.4 Acoustic Wave Propagation

The previous section was important in establishing the interdependence of pressure, volume, and temperature in an acoustic system. Although it was not directly mentioned, density was also discussed, as it represents mass per volume (related to the number of molecules per volume). In acoustics,

density is denoted by the variable, ρ, where

$$\rho = \left[\frac{kg}{m^3}\right] \ . \tag{1.14}$$

When it is possible for air to change temperature, pressure, and volume, the relationship between pressure and density is

$$\frac{P}{P_o} = \left(\frac{\rho}{\rho_o}\right)^{\gamma} \ , \tag{1.15}$$

where γ is the ratio of specific heats from Equation (1.12). In Equation (1.15), P_o and ρ_o are the equilibrium acoustic pressure and equilibrium acoustic density. In air at standard tempereature at sea level these values are

$$P_o = 101325 \text{ [Pascals]} = [\text{N/m}^2] \tag{1.16}$$

$$\rho_o = 1.18 \text{ [kg/m}^3\text{]} \ . \tag{1.17}$$

The equation governing acoustic pressure in a gas is

$$\nabla^2 p = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \ , \tag{1.18}$$

which can be simplified to the one dimensional case of

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \ . \tag{1.19}$$

The solution of the one-dimensional wave equation from (1.19) results in two plane waves traveling in both the positive and negative x directions, of the form

$$p(x,t) = \mathbf{p}^{+} e^{j\mathbf{k}(ct-x)} + \mathbf{p}^{-} e^{j\mathbf{k}(ct+x)} \ , \tag{1.20}$$

where $\mathbf{p}^{+}$ and $\mathbf{p}^{-}$ are the phasors of the forward and backward traveling waves, respectively. In Equation (1.20), the constant, $\mathbf{k}$, is the wave number and is related to the frequency of the acoustic source, and c is the speed of sound, both of which can be found by

$$c = \sqrt{\gamma r T_k} \tag{1.21}$$

$$k = \frac{\omega}{c} \ , \tag{1.22}$$

where r is the specific gas constant and T_k is the absolute temperature in kelvins [3].

The three-dimensional wave equation can also be solved in spherical coordinates, which results in a radially outward propagating wave of the form

$$p(r,t) = \frac{\mathbf{p}^+}{r} e^{j\mathbf{k}(ct-r)} \tag{1.23}$$

Although it is not traditionally taught in textbooks, the spherical wave equation may also consist of an *inward* traveling wave, causing the equation to take the form

$$p(r,t) = \frac{\mathbf{p}^+}{r} e^{j\mathbf{k}(ct-r)} + \frac{\mathbf{p}^-}{r} e^{j\mathbf{k}(ct+r)} \tag{1.24}$$

The conditions under which the inward traveling wave is produced are rare and would probably not occur in nature. However, an inward traveling wave that converges to a relative singularity has been used to initiate the explosive fission of a nuclear bomb. Interesting as it may be, inward traveling spherical waves are almost never considered in practical acoustics.

1.4.1 Approximation of Spherical Source

It is relatively desirable to model acoustic sources as spherical ones, because of the straightforward analysis of spherical wave propagation. An added convenience of modeling a source as spherical is that it is possible to obtain a relatively simple and intuitive approximation of the source's impedance.

The spherical wave equation from (1.24) can be expressed in the form

$$p_r(r,t) = \frac{1}{r} f(ct-r) + \frac{1}{r} g(ct+r) \tag{1.25}$$

By using phasor notation and disregarding the inward propagating wave, the outward wave can be represented as

$$p_r(r,t) = \frac{P^+(\omega)}{r} e^{j\frac{\omega}{c}(ct-r)} \tag{1.26}$$

Through mathematical manipulation, the phasor of the volume velocity,

$U^+(\omega)$ can be related to the pressure by

$$U^+(\omega) = \frac{P^+(\omega)}{j\omega\rho_o}\left(\frac{1}{r} + \frac{j\omega}{c}\right) \tag{1.27}$$

Then, by noting that acoustic impedance is found by $P(\omega)/U(\omega)$, the acoustic admittance of the spherical source can be found.

$$Y_r(\omega) = \frac{U^+(\omega)}{P^+(\omega)} = \frac{1}{j\omega\rho_o r} + \frac{1}{\rho_o c} \tag{1.28}$$

It can be seen that Equation (1.28) has the form of an inductive admittance in parallel with a resistive admittance,

$$Y_r(\omega) = \frac{1}{sL} + \frac{1}{R}\ , \tag{1.29}$$

where $s = j\omega$, $L = \rho_o r$, and $R = \rho_o c$.

1.4.2 Sound Pressure Level

A simple but important matter of notation is the logarithmic measure of acoustic pressure called the sound pressure level, or SPL. The sound pressure level is defined with respect to a reference pressure of $P_{ref} = 20\mu Pa = 2 \times 10^{-}5\ Pa.$ according to

$$SPL = 20\log\left(\frac{P}{P_{ref}}\right) \tag{1.30}$$

1.5 Energy and Power

This section discusses energy in electroacoustic systems and reviews how to power results from these manifestations of energy. To gain a deeper understanding of how energy is related, it is often necessary to perform dimensional analysis using fundamental units. This section compares the main units of energy and power for electrical, acoustical, and mechanical systems.

Before dimensional analysis can be performed, it is first necessary to agree upon common units of measurement. Because of its widespread use, the metric system will be used to express the units in this section. Also when

it is possible to confuse a unit's abbreviation with that of another unit, the entire word will be used to eliminate ambiguity. Otherwise, abbreviations will be used when their meaning is relatively clear within the context.

The *fundamental units* of the metric system that are of most interest are

kilogram	[kg]	unit of mass,
meter	[m]	unit of distance.
second	[s]	unit of time,
coulomb	[C]	unit of charge .

In addition to these fundamental units, there are also units that aid in the comprehension of a system. They are called *convenience units* and include

- *cycles* - the period of a single sinusoid or one complete revolution around the unit circle
- *radians* - there are 2π radians in one complete cycle around the unit circle
- *samples* - a discrete measurement of an analog signal with regard to either time, space, etc.

Convenience units are important in understanding a system; for example, the simple formula for angular frequency, $\omega_c = \Omega_c T$, is well-known in signal processing and has the units

$$\omega_c = \Omega_c T \tag{1.31}$$

$$\left[\frac{radians}{sample}\right] = \left[\frac{radians}{second}\right]\left[\frac{seconds}{sample}\right] . \tag{1.32}$$

It can be seen from Equation 1.32 that without convenience units, discrete angular frequency would be unitless and would have a much less intuitive meaning.

To begin dimensional analysis of energy, let's look at it's metric unit, the *joule*. Energy results from applying a force over some distance, where the units of force are Newtons, $[N]$ or $[kg * m/s^2]$. The joule [J] can have dimensions as follows:

$$1.0\ J = (1.0\ N) \times (1.0\ m)$$

$$1.0\ J = (1.0\ C) \times (1.0\ V) . \tag{1.33}$$

$$\tag{1.34}$$

It can seen from Equation 1.33 that the dimensions of the Volt, [V] are joules per coulomb, or

$$[V] = \frac{[J]}{[C]} \ . \tag{1.35}$$

Equation (1.35) becomes important in the dimensional analysis of power resulting from current flow. The units of electrical current are amperes, [A], where

$$[A] = \frac{[coulombs]}{[second]} = \left[\frac{C}{s}\right] \ . \tag{1.36}$$

Power results when a *motive force* is applied to a *moveable*. The unit of power is the watt [W], and it has dimensions of energy per second. In a mechanical system, the power is found through the product of force with velocity.

$$\begin{aligned} Power &= Force \times velocity = \frac{Energy}{time} && (1.37)\\ [W] &= [N] \times \left[\frac{m}{s}\right] && (1.38)\\ [W] &= \frac{[N \times m]}{[s]} = \left[\frac{J}{s}\right] \ . && (1.39) \end{aligned}$$

The power in an electrical system is similar, however, the motive force is voltage (also called *electromotive force*) and the moveable is the electric current. The unit of power in an electrical circuit is also the watt and can be found by

$$\begin{aligned} Power &= Current \times Voltage \\ [W] &= [amperes \times volts] \\ [W] &= \left[\frac{coulombs}{second}\right]\left[\frac{joules}{coulomb}\right] \\ [W] &= \left[\frac{joules}{second}\right] = \left[\frac{J}{s}\right] \ . && (1.40) \end{aligned}$$

It is important to note that the units of electrical energy and mechanical energy are the same, and that power can be expressed in terms of the same fundamental units as well.

1.6 Types of Impedance

Impedance is central to almost every subject matter, whether it is electrical impedance, acoustic impedance, or mechanical impedance. Because of their frequent occurrence in electroacoustic systems, it is worth discussing the distinctions between these impedances, as well as how they are derived. Each of the following impedances (with the exception of specific acoustic impedance) may be a complex function of frequency and written in the form Z, $Z(j\omega)$, $Z(s)$, etc.

Mechanical impedance is perhaps the most familiar, and is defined as the ratio of force to velocity. The units of force are Newtons, $[N]$ or $[kg * m/s^2]$.

$$Z_m = \frac{F}{v} \quad \left[\frac{N}{m/s}\right] \tag{1.41}$$

In electrical engineering, voltage divided by current is recognized as electrical impedance. Because electrical voltage is the energy required to move a coulomb of charge over some load, it has units of joules per coulomb, $[J/q]$ or $[volts]$. In a wire, electrical current is the amount of charge passing through a cross-section of the wire per unit of time, $[C/s]$, or amperes. Electrical impedance can be found by the ratio of voltage to current in a circuit,

$$Z_e = \frac{V}{I} \tag{1.42}$$

$$Z_e \;=\; \frac{[volts]}{[amperes]} = \frac{[joules/coulomb]}{[coulomb/sec]} = \frac{[joules]\,[sec]}{[coulomb]^2} \;.$$

Electrical impedance is the ratio of electromotive force to the rate of electric charge flow. Stated this way, it may be easer to understand how it is similar to mechanical impedance. Both represent the force required to sustain a given velocity or flow.

Acoustic pressure is force per area and has units of Pascals, $[Pa]$, or Newtons per square-meter, $[N/m^2]$. It can can be thought of as a force normalized over an area. Similarly acoustic volume velocity, with units $[m^3/s]$, can be thought of as velocity through an area.

$$[m^3/s] = [m^2] \times \left[\frac{m}{s}\right] \tag{1.43}$$

Acoustic impedance is the ratio of acoustic pressure to the volume velocity of air.

$$Z_A = \frac{P}{U} \ . \tag{1.44}$$

The acoustic impedance of Equation (1.44) is to be distinguished from the impedance experienced by an particle in the air, or *specific acoustic impedance* of air,

$$Z_o = \rho_o c \ , \tag{1.45}$$

where ρ is the density and $c \approx 343\ [m/s]$ is the speed of sound at standard temperature and pressure.

1.7 Signal Processing Techniques

Digital signal processing is an integral part of data acquisition and analysis for HRTF measurement. Many steps in the measurement involve the processing of a digitally sampled audio signal. The following sections address several key concepts required for the manipulation of discrete signals.

1.7.1 Frequency Domain Transformations

There are several types of frequency domain representations of signals. However, it is important to know when a particular transformation is best suited for the current application. For this reason, several transformations, as well as the best time to use them, are addressed below.

- **Fourier Series** The Fourier series can be applied to continuous periodic signals, transforming them into the frequency domain at discrete multiples of the signal's fundamental frequency.

$$F(\omega_k) = \frac{1}{T} \int_{T/2}^{T/2} f(t)_T e^{-j\frac{\omega_k}{T}t} \ \mathrm{d}t \tag{1.46}$$

- **Fourier Transform** The Fourier transform is used in the transformation from continuous time to continuous frequency. The signals do not need to be periodic; however, some signals that exhibit transients such as exponential decay might not have a Fourier transform.

$$F(\omega) = \int_{t=-\infty}^{\infty} f(t)e^{-j\omega t}\ \mathrm{d}t \tag{1.47}$$

- **Laplace Transform** The double-sided Laplace transform is unique because it uses an argument, s, in the exponent that has both a real and a complex component, $s = j\omega + \sigma$, whereas the Fourier Transform uses only the complex $j\omega$. The real value in the exponent allows the Laplace transform to converge for most causal signals that exhibit exponential behavior.

$$F(s) = \int_{t=-\infty}^{\infty} f(t)e^{-st}\ \mathrm{d}t \tag{1.48}$$

  A one-sided Laplace transform, where the lower limit of the integral is 0 rather than $-\infty$, is commonly used so that the transform converges for functions that are analytic in a half-space of the complex s-plane.

- **DTFT** The discrete-time Fourier transform (DTFT) can be used to transform a discrete signal in time to a continuous frequency domain representations. Although it is mathematically valid, it cannot be applied to real-world signals, because it would require computation over infinite time.

$$X(\omega) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \tag{1.49}$$

- **DFT** The discrete Fourier transform (DFT) can be applied to discrete signals, and results in a discrete spectrum. It is related to the DTFT in that it is a sampled version of the DTFT, where the sampling occurs in the *frequency domain*. The reason that it is not continuous in the frequency domain is that it is computed over a finite window of time. The fact that an infinite window of time is not necessary makes it possible to apply the DFT to real-word signals.

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi nk}{N}} \tag{1.50}$$

- **FFT** The fast Fourier transform (FFT) is mathematically equivalent to the DFT; however, it is implemented in a computationally efficient way that groups calculations together in either the time-domain or frequency-domain rather than performing the cumbersome summation in Equation (1.50) directly. The speed of the FFT is the reason it is almost always used in practice to calculate the DFT of a signal.

- **STFT** The short-time Fourier transform (STFT) is very similar to the DFT, however the signal of interest is multiplied by a *window* function, $w[n]$ before the Fourier transform is applied. Actually, the DFT can be thought of as a special case of the STFT, when the windowing function is rectangular. The window can be moved across the signal to take "snapshots" of the signal in different regions of time. If the window is set to move at constant increments of R samples at a time, then the STFT takes the form.

$$X[k,m] = \sum_{n=0}^{N-1} w[n-mR]s[n]e^{-j\frac{2\pi nk}{N}} \tag{1.51}$$

- **Z-Transform** The double-sided Z-transform is very similar to the Laplace transform because it uses an argument, $z = re^{j\omega}$, that accounts for exponential behavior. Notice that the argument, $z = re^{j\omega}$, describes a complex plane in polar coordinates. Most causal discrete signals have a z-transform that converges in some region of the complex z-plane.

$$X[z] = \sum_{n=-\infty}^{\infty} x[n]z^{-n} \tag{1.52}$$

1.7.2 Windowing

When processing signals such as the head-related impulse response, it is often necessary to look at only a finite duration of a signal at a time. Even if the signal is simply truncated before and after the region of interest, this introduces spectral distortion to the signal. It is a mathematically proven fact that multiplying two signals in the time domain (or in discrete-time) has the effect of convolving the respective spectra of the two signals in the frequency domain. Even simple signal truncation has the mathematical sig-

nificance of multiplying the observed signal with a rectangular window in the time domain, resulting in the convolution with a sinc function in the frequency domain. For this reason, it is often important to carefully consider the spectral properties of the window itself when attempting to extract the spectral properties of a windowed signal.

Examples of windows that are commonly applied are rectangular, Hamming, and Kaiser windows. The mathematical formulas along with some properties of these windows are discussed below.

- *Rectangular Window* The rectangular window is the simplest window conceptually, as it has a magnitude of 1 over the signal's region of interest and a magnitude of 0 everywhere else. Its advantage in the frequency domain is that it has a narrow mainlobe which can be helpful in resolving frequencies that are close. Its main downfall, however, is that its sidelobes are quite high, approximately 13 dB below the main lobe.

- *Hamming Window* The Hamming window is one of the most widely used windows, because it has the desirable property of low sidelobes in the frequency domain, which are at least 40 dB below the main lobe, as shown in Figure 1.2.

$$w[k] = 0.54 - 0.46\cos\left(2\pi\frac{k}{N-1}\right) \quad , \quad k = 1, ..., N \tag{1.53}$$

  Another attractive feature of the Hamming window is that it has a low *time-bandwidth product* relative to many other windows. The time-bandwidth product, γ, is simply the product of the length of the window (in time) and the width of the mainlobe of the window's Fourier transform.

- *Kaiser Window* The Kaiser window can take on many forms, because it is defined in terms of several adjustable parameters. One thing that makes the Kaiser window desirable is that it can be designed to be application specific, with either lower or narrower sidelobes when needed.

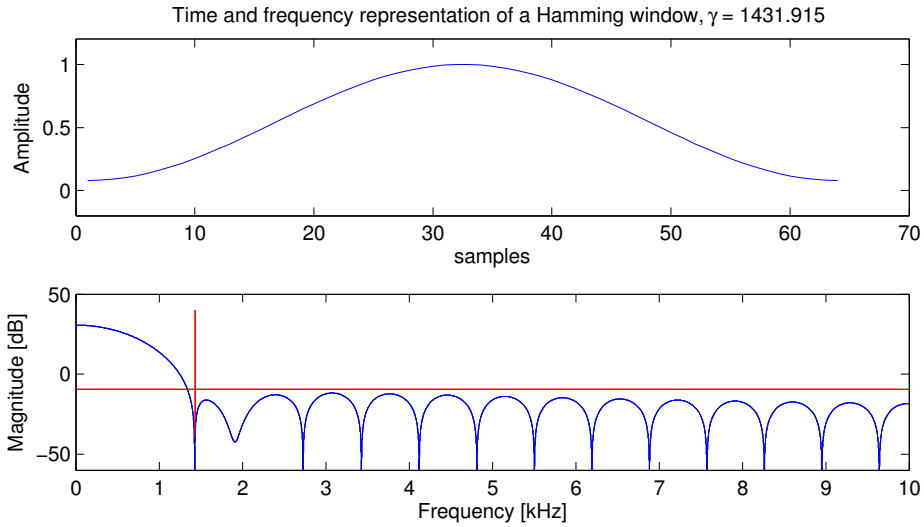

Figure 1.2: Top: Plot of a Hamming window in time. Bottom: Spectrum of Hamming window with wide mainlobe and low sidelobes, assuming a 44.1 kHz sampling frequency.

1.8 2-Port Networks

1.8.1 ABCD Matrix

An electroacoustic transducer is a specific type of 2-port network that is capable of making the transformation between electrical voltage and current to and from acoustic pressure and volume velocity, respectively. It is often

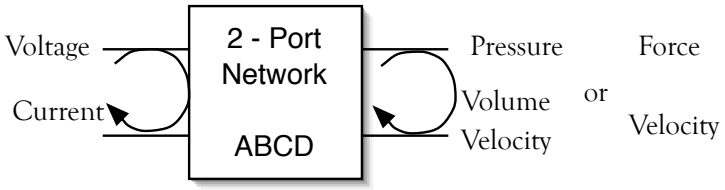

Figure 1.3: Example of a 2-port network, represented by an ABCD matrix.

possible to model this transformation in the form of an ABCD matrix (sometimes called a transmission matrix). Figure 1.3 shows an example of a 2-port network, represented by an ABCD matrix that is capable of converting from electrical voltage and current to either mechanical force and velocity or to acoustic pressure and volume velocity.

If it is desired to transform voltage and current to pressure and volume

velocity, the ABCD matrix would appear as

$$\begin{bmatrix} P \\ U \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V \\ I \end{bmatrix} \tag{1.54}$$

1.8.2 Impedance Matrix - Z

It is also possible to model a 2-port network using an impedance matrix as shown in Figure 1.4. The mathematical representation of Figure 1.4 can be

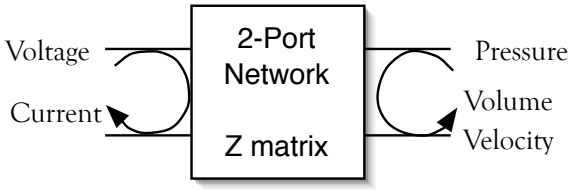

Figure 1.4: Example of a 2-port network, represented by a impedance matrix.

shown using the matrix notation

$$\begin{bmatrix} V \\ P \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} I \\ U \end{bmatrix} \quad . \tag{1.55}$$

Notice that in impedance matrix form, the current is always defined as going *into* the 2-port Network. This convention is different from the ABCD matrix form, where the current is always oriented clockwise, enabling concatenation of transmission networks.

1.8.3 Conversion Between Z and ABCD Matrices

Once a 2-port network has been found in either the impedance matrix form or transmission matrix form, it can be converted to its alternate form. If the elements of the impedance matrix are known, it is possible to find the elements of the T matrix (representing the ABCD matrix) by

$$T = \frac{1}{z_{21}} \begin{bmatrix} z_{11} & \Delta_z \\ 1 & z_{12} \end{bmatrix} \quad , \tag{1.56}$$

where Δ_z is the determinant of the impedance matrix.

Similarly, if the elements of the ABCD matrix are known, the Z-matrix can be found by

$$Z = \frac{1}{C}\begin{bmatrix} A & \Delta_T \\ 1 & D \end{bmatrix} \ . \tag{1.57}$$

1.8.4 Properties of 2-Port Networks

A two-port network may have several properties that characterize its functionality. It can be shown that a network satisfies each property if it satisfies the postulate corresponding to the property. While an in-depth discussion of each of these properties would be worthy of its own paper, seven specific network postulates are outlined below.

1. *Linearity* A network is considered linear if it obeys the principles of superposition. For example if it performs the transformation Γ, then it is linear if and only if

$$\Gamma(c_1 x(t) + c2y(t)) = c_1\Gamma(x(t)) + c_2\Gamma(y(t)) \ , \tag{1.58}$$

   where c_1 and c_2 are constants.

2. *Time Invariance* If every input corresponds to a specific output, a network is considered time-invariant if, when an input is delayed, the only change in the output is that it is delayed by the exact same amount.

3. *Passivity* A network is considered passive if it is not a power source and does not add any energy to the system. Passivity holds true if the transmission matrix is positive definite.

4. *Causality* A network is causal if the current output is determined only by the current and past inputs.

5. *Real* A network consists of real-valued parameters and functions of time such as voltage, current, pressure, etc.

6. *Reciprocity* If the input and output quantities can be scaled in such a way that the impedance matrix becomes Toeplitz, then the network is

reciprocal. A reciprocal network can communicate energy simultaneously in both directions, e.g., it can play back recorded signals while simultaneously recording the acoustic response.

7. *Reversibility* A network is reversible if the output terminals of the system can be used as the inputs, creating the corresponding "output" at the side of the input terminals. Or stated another way, if specific inputs create specific outputs, then the specific outputs can be used to create the specific inputs. It is important to note that reversibility is NOT the same as reciprocity.

1.8.5 Examples of 2-Port Networks

Two-port networks can be used to model a variety of real-world systems. This section outlines how 2-port networks are also used to model piezoelectric transducers, speakers, and transmission lines [4]. The ABCD matrices for these systems are provided below.

- **Speaker Coil** - The tricky part of the impedance matrix of a speaker coil is that it requires the cross-product of $\mathbf{E} = l(\mathbf{v} \times \mathbf{B})$. For this reason, Hunt used what is called the *quaternion* operator, $\mathbf{k}$, to rewrite the equation in the form, $E = Bl\mathbf{k}v$. The quaternion is a spatial operator capable of encompassing the spatial dimensionality of the cross-product [4]. Its use in the impedance matrix of the speaker coil is significant, because it causes the matrix to be symmetric when it would otherwise be antisymmetric. Using the quaternion operator, the impedance matrix for the speaker coil becomes,

$$Z_{coil} = \begin{bmatrix} Z_e & Bl\mathbf{k}v \\ Bl\mathbf{k}v & Z_m \end{bmatrix} . \tag{1.59}$$

- **Piezo Electric Transducer** - The impedance matrix for a piezoelectric transducer consists of four important impedance elements.

$$\begin{bmatrix} E \\ F \end{bmatrix} = \begin{bmatrix} Z_e & T_{em} \\ T_{me} & Z_m \end{bmatrix} \begin{bmatrix} I \\ v \end{bmatrix} \tag{1.60}$$

$$Z_{piezo} = \begin{bmatrix} Z_e & T_{em} \\ T_{me} & Z_m \end{bmatrix}, \tag{1.61}$$

where

$$\begin{aligned} Z_e(s) &= R_e + sL_e + \frac{1}{sC_e} = \text{blocked electrical impedance} \\ Z_m(s) &= r_m + sl_m + \frac{1}{sc_m} = \text{open-circuit mechanical impedance} \\ T_{em} &= r_T + sL_T + \frac{1}{sC_T} = \text{electrical-to-mechanical transimpedance} \\ T_{me} &= R_T + sl_T + \frac{1}{sc_T} = \text{mechanical-to-electrical transimpedance} \end{aligned}$$

- **Transmission Line** - For a transmission line of length, L, and impedance, Z_o the ABCD matrix is

$$T_{line} = \begin{bmatrix} \cosh(\gamma L) & Z_o \sinh(\gamma L) \\ \frac{\sinh(\gamma L)}{Z_o} & \cosh(\gamma L) \end{bmatrix}, \tag{1.62}$$

where $\gamma = \sqrt{ZY} = \sqrt{(j\omega)^2 LC}$.

CHAPTER 2

PREVIOUS WORKS

The chapter will present a brief history of HRTF measurement methods and proceed to summarize the results of several modern studies, each of which seeks to improve the accuracy of the head-related transfer function by the use of physical modeling based on fundamental acoustic principles. In addition, a psychoacoustic study pertaining to the importance of HRTF accuracy will be presented.

2.1 Early Measurements of the HRTF

Although the head-related transfer function has not always gone by the same name, researchers have been seeking to characterize it for over 70 years, using various testing methods. In 1974, Shaw made an effort to combine the results of 12 studies dating back to 1933 which sought to uniquely characterize the HRTF. In these studies, both direct and indirect methods of measurement had been used in the development of the head-related transfer function. In the direct method, the transformation of the sound pressure level (SPL) from the free field to the eardrum was measured by placing a small probe tube microphone in a subject's ear.

The subject would remain seated in an anechoic chamber as an acoustic point source of a set frequency was placed at a specific distance and angle from the stationary head. The SPL at the listener's ear was then measured over a range of frequencies for several discrete angles along the horizontal (azimuth) plane. Some of the test data in the previous studies had been collected from indirect measurements. The measurement was considered indirect if it did not measure directly at the eardrum. Placing the probe-microphone in the orifice located at the ear canal entrance, the concha, was a common way of performing an indirect measurement method. What set the direct mea-

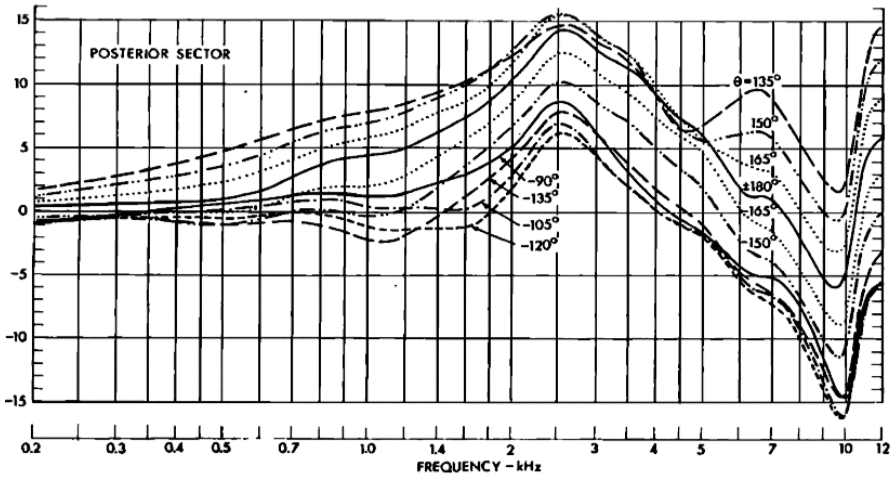

Figure 2.1: Family of 10 self-consistent curves $T_d(v)$ showing transformation of sound pressure level from free field to eardrum as a function of frequency v at 15^o intervals of azimuth θ, as shown. (included from [5])

surements apart from the indirect is that by measuring at the eardrum, the transfer function of the ear canal and the pressure distribution within it was included in the HRTF. For the indirect method, the transfer function of the ear canal needed to be modeled and added to the measured HRTF. After employing various methods to match the data sets of each previous study to a common format, Shaw produced curves representing the transformation of sound pressure level from the free-field to the eardrum as a function of frequency for 15 degree intervals of azimuth. His results for the frontal sector are shown in Figure 2.1

2.2 Impact of the Head and Torso on the HRTF

Given that the head itself has a profound influence on the head-related transfer function, it is logical to assume that the rest of the body might have considerable influence as well. In order to gain a better understanding of this, Duda and Algazi performed a study on approximating the head-related transfer function using simple geometric models of the head and torso [6]. In their study they considered two models of the head and torso. The first model consisted of a large spherical surface that represented the torso of a listener, and a smaller spherical surface was placed on top of the larger to

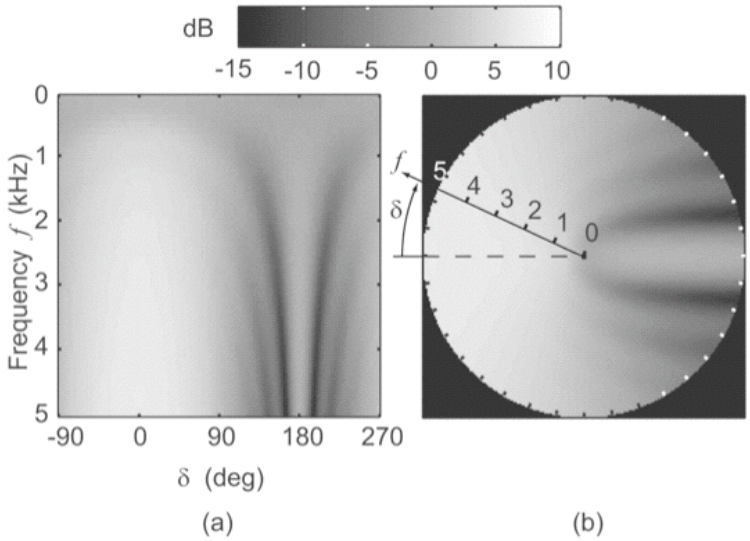

Figure 2.2: Image representations of the magnitude of the HRTF for an ideal rigid sphere. (a) shows the HRTF as a function of frequency verses elevation. in rectangular coordinates. (b) shows same data in polar coordinates. (scanned from [6])

represent the head of a listener. This first model was named the "snowman" model because of its appearance. The second model was similar to the first; however, the torso was modeled as ellipsoidal to approximately match the dimensions of the torso for the KEMAR. For both models, the pinna were removed to isolate the effect of the head and torso only. For the first experiment, a real two-sphere "snowman" model was constructed on a thin rod in the laboratory. The multipole expansion method for two spherical surfaces was used to model this system in order to obtain an approximation to the expected HRTF. Then actual measurements were taken in the laboratory to evaluate the accuracy of the multipole reexpansion method. The results of the simulation for a spherical head alone are plotted as a function of elevation versus frequency in Figure 2.2 for the right ear. Plot (a) in Figure 2.2 represents the intensity at the pinnaless ear for a given frequency and for sources existing only on a vertical plane around the listener. Plot (b) shows the same information, but in polar coordinates from the perspective of looking towards the listener's face, with his ear at the origin. In plot (b) it is much easier to see the bright spot occurring at 180 degrees, which corresponds to the source being on the exact opposite side of the head. When the torso of the snowman

model is added to the modeling, the HRTF is dramatically altered, as seen in Figure 2.3. The dark region occurring between angles between 210 and 250 degrees is the effect of the torso. The physical meaning of this shadow is that a sound above 1 kHz frequency will be greatly attenuated at the right ear if it comes from a low elevation on the opposite side of the head. When the

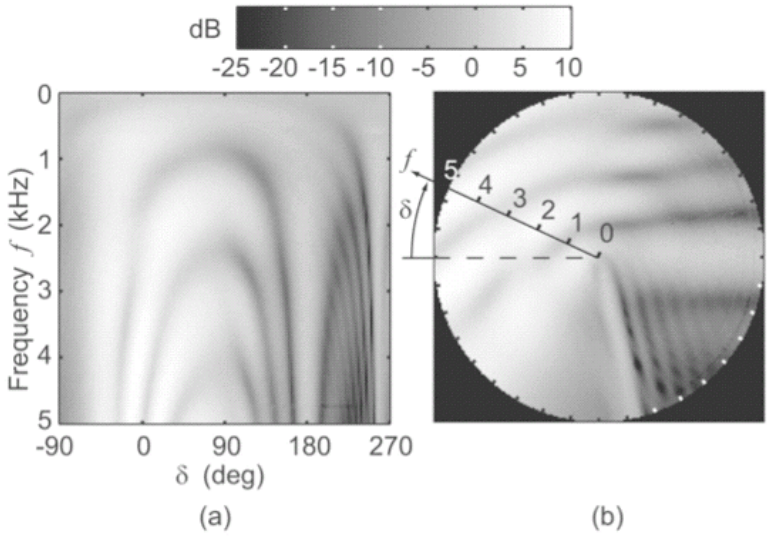

Figure 2.3: Boundaries of the torso-shadow cone for the physical showman model. (a) defines direction of the torso and the bright spot. (b) exhibits strong correspondence between geometrical boundaries and the region of reduced response (scanned from [6]).

HRTF of the physical snowman was measured in the laboratory, the HRTF shown in Figure 2.4 was obtained. It can easily be seen that this empirically measured HRTF exhibits strong similarity to the expected HRTF. This result supports the performance of multipole reexpansion in predicting sound pressure levels for a simple two sphere case.

For the second experiment, the multipole reexpansion method could not be used to model the HRTF, because it is only valid for sound scattering off spherical surfaces and this model incorporated an ellipsoidal torso roughly matching the dimensions of the KEMAR. In place of multipole reexpansion, another method called the boundary element method (BEM) was used. In the boundary element method a mesh of dense points is used to model some irregularly shaped surface.

While the boundary element method may be used to model irregularly

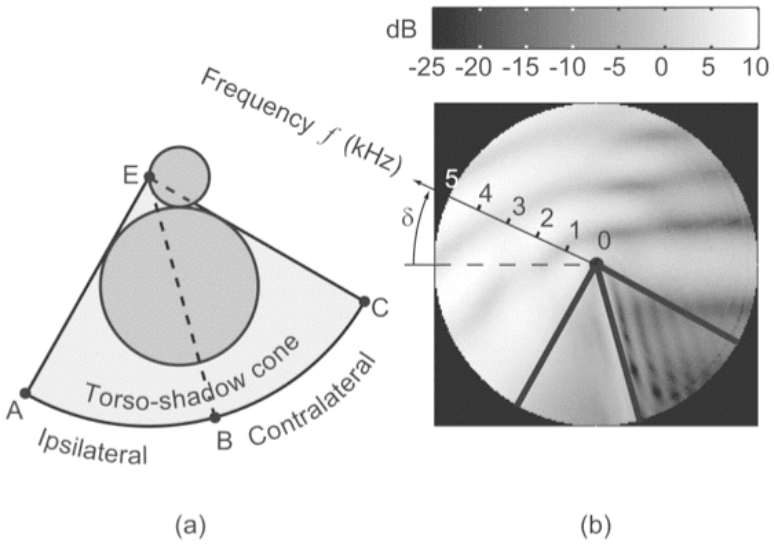

Figure 2.4: Image representations of the magnitude of the HRTF for the snowman model scaled for a head radius of 8.75 cm. (a) shows the HRTF as a function of frequency verses elevation, in rectangular coordinates. (b) shows same data in polar coordinates. (scanned from [6])

shaped surfaces, its calculation can be much more computationally intensive than the multipole reexpansion involving two spherical surfaces. As a preliminary test of the BEM, it was applied to the snowman model of the first experiment to yield the same results as multipole reexpansion. Similar to the first experiment, the HRTF was measured in the laboratory, as well as calculated using the BEM. The experimental results were then compared to the KEMAR HRTF in the pinnaless case, shown in Figure 2.5.

This study showed evidence that the introduction of the torso to a simple spherical model of the head can have a profound effect on the HRTF. This combined head and torso model was also shown to more accurately model the behavior of the pinnaless KEMAR. This finding suggests that models that don't take the torso into account could be have significant errors; especially for lower-frequency sources in the shadow region of the torso. Inaccurate rendering of such sources in virtual auditory space could yield flawed perception of sounds such as footsteps.

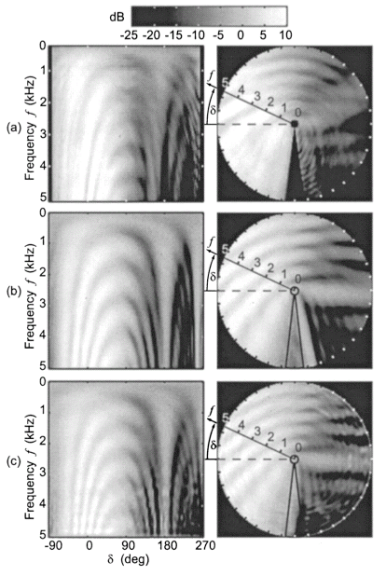

Figure 2.5: Measured and computed frontal-plane HRTFs. (a) KEMAR with the pinna removed. (b) The spherical-torso snowman approximation. (c) An ellipsoidal-torso snowman approximation. (scanned from [6])

2.3 Interpolation and Range Extrapolation of HRTFs

A characteristic of empirically measuring the HRTF is that it must be discretely sampled over a range of incident angles. Two important limitations are that (1) measured HRTFs do not directly allow one to find the transfer function for an angle that lies between two sampled points; and (2) when many source angles are considered, they are usually measured with respect to a single far-field distance. This measured HRTF would not be an accurate representation of the true transfer function of a source that is very close to the head or had a low frequency.

Time-domain methods of interpolation are commonly used to derive the HRTF between two azimuthal angles and usually employ a simple weighting of the HRIRs depending on their proximity to the desired azimuth [7]. In an attempt to address the second limitation, a study was conducted which used the multipole reexpansion method to interpolate the HRTF for angles that lie between those which were previously sampled, and to extrapolate the HRTF to a desired source range [8].

The fundamental principle behind multipole reexpansion is the *acoustic principal of reciprocity*, which states that if the locations of a small source and

small receiver are interchanged in an unchanging environment, the received signal will be the same [8].

The multipole's physical interpretation is a difficult concept; however, several equations behind the multipole reexpansion technique will be provided for completeness, as copied from [8]. The main thing to keep in mind is the multipole reexpansion is a means to compute the potential field that results when an acoustic plane wave is scattered over a spherical surface. A description for a multipole, Φ_{lm}, of order m and degree l is provided in Equation (2.1).

$$\Phi_{lm}(r,\theta,\varphi,k) = h_l(kr)Y_{lm}(\theta,\varphi), \quad k = \omega c^{-1} \tag{2.1}$$

In (2.1), k is the wavenumber, $h_l(kr)$ represents the spherical Hankel functions of the first kind, and $Y_{lm}(\theta,\varphi)$ are the spherical harmonics,

$$Y_{lm}(\theta,\varphi) = (-1)^m \sqrt{\frac{2n+1}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}} P_l^{|m|}(\cos\theta)e^{jm\varphi}, \tag{2.2}$$

where $P_l|m|$ are the associated Legendre polynomials.

Multipole reexpansion finds a solution for complex pressure amplitude, ψ, that satisfies both the the Helmholtz equation in 3–dimensional space, and the Sommerfeld radiation condition. The complex Helmholtz equation in three dimensions is the Fourier transform of Equation (1.18), thus

$$\nabla^2\psi(\mathbf{x},k) + k^2\psi(\mathbf{x},k) = 0 \ , \tag{2.3}$$

and the Sommerfeld radiation condition requires that the local wavefront must approach that of a plane wave as $r \to \infty$, thus

$$\lim_{r\to\infty} r\left(\frac{\partial\psi}{\partial r} - jk\psi\right) = 0, \quad r = |x|. \tag{2.4}$$

Instead of evaluating the potential for all possible degrees, l was chosen to be no larger than the truncation number, p. The new m values of α_{lm} are coefficients which will be used to solve for the least mean squared mapping of known potential functions to the known multipoles of the HRTF data.

The general idea behind multipole reexpansion is that if the HRTF data represents what was received when a specific combination of multipole sources

were present, then under the acoustic principle of reciprocity, it should be possible to reverse the system to find out what acoustic source was responsible for the HRTF. Imagine replacing the ear with an acoustic source with a given frequency, to calculate the potential function for that HRTF.

$$\psi = \sum_{l=0}^{\infty} \left(\sum_{m=-l}^{l} \alpha_{lm} h_l(kr) Y_{lm}(\theta, \varphi) \right) \tag{2.5}$$

In the implementation of the multipole reexpansion technique, the study noted that the choice of the truncation number is relatively important. Previous experiments with this method indicated that good interpolation quality was achieved when the truncation number was chosen to be roughly equal to the wavenumber. The resulting equations are

$$\psi_h(\mathbf{x}_1, k) = \sum_{l=0}^{p-1} \sum_{m=-l}^{p-1} \alpha_{lm} \Phi_{lm}(\mathbf{x}_1, k), \tag{2.6}$$

$$\psi_h(\mathbf{x}_N, k) = \sum_{l=0}^{p-1} \sum_{m=-l}^{p-1} \alpha_{lm} \Phi_{lm}(\mathbf{x}_N, k), \tag{2.7}$$

which can be simplified to

$$\Phi A = \Psi. \tag{2.8}$$

When the α_{lm} values of vector A are solved for, the acoustic field can be calculated for any given range, or angle.

The correlation between the truncation number and the wavenumber is supported by the underlying physics of the model. For high-frequency sources, the potential field will change rapidly as one travels through it in three dimensions. In order to fit this rapidly changing potential field with multipoles, a higher number of multipoles need to be placed around the head in the expansion, which results in a higher truncation number. Choosing the truncation number to be too large, however, can also have negative effects on the evaluation of the potential field at a smaller radius, because of the behavior of the spherical Bessel functions of the first kind, $j_l(kr)$. In order to find a truncation number, p, roughly equal to kr, the following equation was used.

$$p = \text{integer}(kr) + 1. \tag{2.9}$$

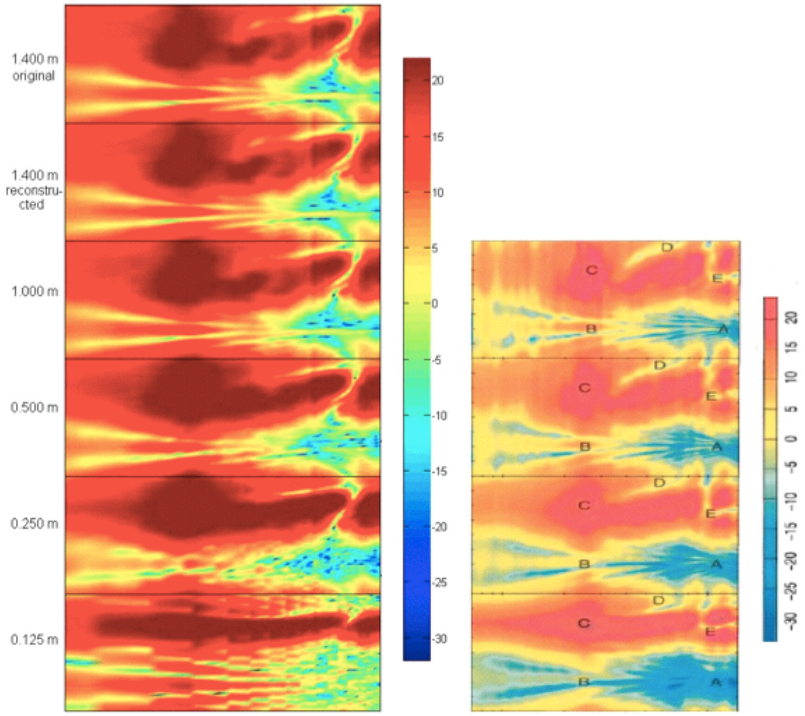

Figure 2.6: Left: KEMAR HRTF measured at 1.4 meters and results of HRTF reconstruction at 1.4 meters and HRTF extrapolation to 1.0, 0.5, 0.25, and 1.25 meters. Right: DEMAR HRTF measured at 1.0, 0.5, 0.25, and 0.125 meters (scanned in from [9] and [8]). The horizontal axis is frequency, the vertical axis is azimuth.

For verifying the multipole reexpansion technique for range extrapolation, two main tests were conducted.

The purpose of the first experiment was to confirm the performance of multipole reexpansion, for an analytical sphere HRTF model that had been calculated over a grid containing 1,636 points. Multipole reexpansion was used to compute the potential function at $r = 1$ meter, which was then used for the multipole decomposition of the potential field. This potential field was used to extrapolate the HRTF for a source at a distance of 0.5, 0.25, and 0.125 meters.

The HRTFs obtained from range extrapolation after multipole expansion were then directly compared to the analytical HRTFs, which had been calculated for the same ranges. It was found that the reconstructed HRTFs matched almost perfectly with the analytical results, which verified the functionality of the method for a simplified model.

The second experiment sought to perform range extrapolation on the HRTF of the KEMAR, empirically measured at a fixed distance of 1.4 meters, in an attempt to match the transfer function with HRTF data, empirically measured (in a separate study) over ranges of $r = \{1, 0.5, 0.25, 0.125\}$ meters. The result, seen in Figure 2.6, shows general agreement in some characteristics of the head-related transfer function

2.4 Psychoacoustic Importance of HRTF Accuracy

The previous research studies were all geared towards increasing the accuracy of the head-related transfer function in order to more accurately recreate a virtual auditory space. This motivation is logical; however, it is based on an assumption that if the modeled HRTF is more true to the actual HRTF for an individual, the individual's perception of the virtual auditory space would be more profound. A recent psychoacoustic study [1] questioned this assumption by investigating the role of spectral detail in sound-source localization.

An experiment was conducted to test the listener's ability to distinguish "real" sounds and "virtual" sounds. The real sounds were played over a loudspeaker, and the virtual sounds were introduced to the ear through open canal tube-phones. To test the importance of spectral detail, these virtual sounds were passed through an HRTF that was a "smoothed" version of the original data. This smoothed HRTF still retained the dominant spectral notches found in the original.

In a complete binaural head-related transfer function, the interaural time difference (ITD) is already incorporated to the filter. In the procedure for this experiment, however, modified HRTFs were used. The minimum phase approximation of the HRTF was served as the filter, which only accounted for frequency amplitude at the ear. Then the ITD was added after calculating the expected ITD for actual free-field stimulus.

In order to smooth the HRTF, the Fourier transform of the original HRTF was used to calculate its spectrum. Then the HRTF was reconstructed after truncating the upper spectral components of the transformed HRTF. Let's elaborate quickly on this point, as it may be a confusing concept. The head-related transfer function is already a function of the frequency spectrum. The Fourier transform of this frequency spectrum shows at what rate the HRTF

changes with a change in frequency. Truncating this second spectrum essentially has the effect of low-pass filtering the *frequency domain* representation of the HRTF. The smoothed versions of the HRTF can be seen in Figure 2.7.

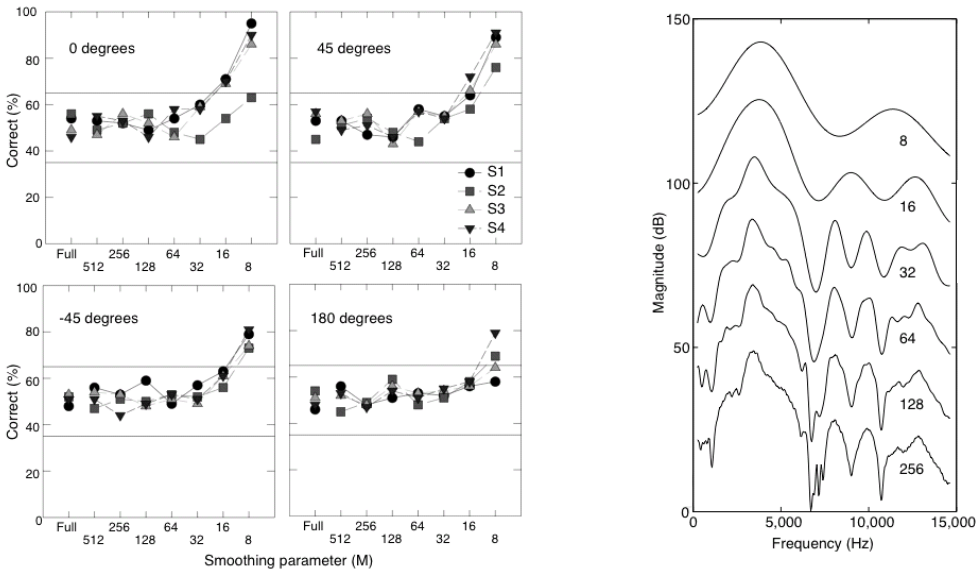

Figure 2.7: Left: Discrimination performance of four subjects (S1–S4) as a function of the smoothing parameter, M, at each azimuthal position tested. The 95% confidence bounds for chance performance are shown by the two horizontal lines in each graph. Right: Spectral smoothing of the left-ear HRTF magnitude spectra measure from a representative subject (S1) for the 0-degree location. (scanned from [1])

The subjects (listeners) were given real and virtual sounds from azimuthal angles of 0, 45, 135, and 180 degrees. They were simply asked to identify which sound was which. If both the real and the virtual sound were to produce exactly the same pressure function at the eardrum, it is expected that the listener would correctly discern the real from the virtual with 50% accuracy, after many trials.

The results of the experiment indicate that the four test subjects could not discriminate between the real and virtual sounds until the HRTF was smoothed enough to eliminate significant spectral notches. According to the results, at least in an anechoic space when no head movements occur and when the visual cues are consistent and accurate, details of the HRTFs (magnitude and phase) are not responsible for a well-localized, externalized

perception of a sound source. ([1] p.748)

They did stress however, that the individualized HRTFs are not unimportant, and that only individualized HRTFs were used in the experiment.

2.5 Summary

The search for an accurate representation of the head-related transfer function (HRTF) dates back to the 1930s. Since then, insight into what uniquely characterizes the HRTF has been gained through research modeling the acoustics of combinations of the torso, head, and pinna. The results from multipole reexpansion techniques and the boundary element method suggest that it is possible to construct an accurate HRTF based solely on the physical characteristics of a given listener. Further advancement to HRTF accuracy, however, should yield meaningful results. For this reason, psychoacoustic analysis of the relevance should complement further efforts of HRTF research.

CHAPTER 3

SYSTEM CHARACTERIZATION

This chapter discusses the subject of system characterization of a linear time-invariant (LTI) system by the estimation of the system's impulse response.

Two excitation signals are presented: the chirp signal and the maximum length sequence (MLS). Though the representations of the chirp signal and the MLS are very different in the time-domain, they can be used independently to estimate the same impulse response of a system.

If it can be confirmed that MLS and chirp measurement methods can independently obtain the same transfer function with acceptable relative error, the measurement methods can serve as an extra check to verify the validity of one another.

3.1 Impulses in Continuous and Discrete Time

There are different delta functions that serve as the mathematical definition of an impulse for continuous-time and discrete-time. As in previous sections, the notation for all discrete-time functions will be represented with bracket notation, and continuous-time will be represented with parenthesis.

The Dirac delta function, $\delta(t)$, is the unit impulse in continuous-time and can be expressed as

$$\delta(t) \equiv \lim_{a \to 0} \left[\frac{1}{a\sqrt{\pi}} e^{-t^2/a^2} \right] \ . \tag{3.1}$$

The Dirac delta function is a purely theoretical mathematical function that has desirable mathematical properties, especially for convolution; but only approximations of it can exist in reality due to the fact that $\delta(0) = +\infty$.

The Kronecker delta function, $\delta[n]$, comprises the unit impulse in discrete-

time, and can be expressed as

$$\delta[n] = \begin{cases} 1 & \text{, if } n = 0 \\ 0 & \text{, otherwise .} \end{cases} \tag{3.2}$$

Unlike the Dirac delta function, the value of the Kronecker delta function is finite at $n = 0$ and does not take on a value greater than one.

The key point to take away from this section is that the Kronecker delta is *not* a sampled version of the Dirac delta function and there is no direct transform between the two. However, both the Dirac and Kronecker delta functions have the property shown in Equation (3.3) and Equation (3.4), respectively.

$$\delta(t) * x(t) = x(t) \tag{3.3}$$

$$\delta[n] * x[n] = x[n] \tag{3.4}$$

3.2 The Need for Excitation Signals

The impulse response of a continuous-time LTI system, $h(t)$, relates the excitation signal, $x(t)$, to the excitation response, $y(t)$, by

$$x(t) * h(t) = y(t) \ . \tag{3.5}$$

Note that if the Dirac delta function could be substituted for the excitation signal in Equation (3.5), then

$$\delta(t) * h(t) = h(t) \ . \tag{3.6}$$

Equation (3.6) indicates that $h(t)$ is the theoretical response of the system when the excitation signal is an impulse. Figure 3.1 illustrates an analog LTI system with an impulse response, $h(t)$, if an analog impulse in the form of a Dirac delta were supplied as the excitation signal.

Although it is technically impossible to create a perfect Dirac delta to be used as an impulse, pulse generators have been made to approximate an impulse in physical systems and used to obtain impulse response measurements.

This work seeks to obtain the discrete-time characterization of an acoustic

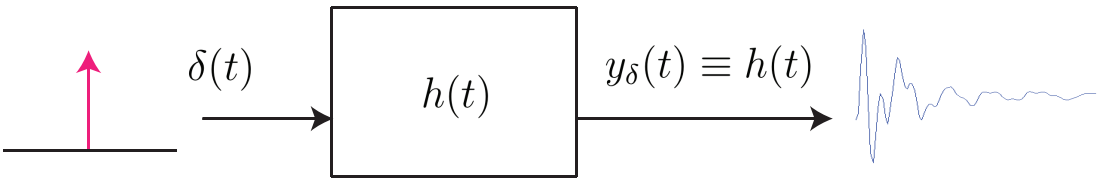

Figure 3.1: Impulse Response

system as it relates to the head-related transfer function. The discrete-time characterization requires the sampling of Equation (3.5) at intervals of $t = nT$ to obtain

$$x[n] * h[n] = y[n] \ , \tag{3.7}$$

where T is the sampling period in seconds.

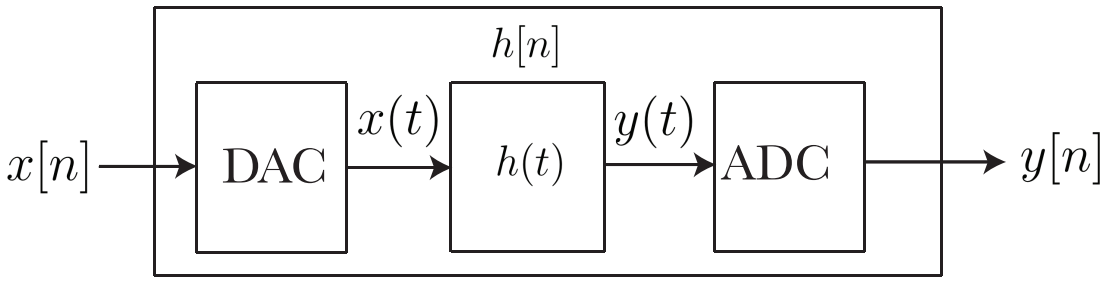

Figure 3.2: Overall System

Figure 3.2 illustrates where the digital-to-analog converter (DAC) and the analog-to-digital converter (ADC) are located with respect to the physical acoustic system that is represented in continuous-time.

It is a natural question to ask why the Kronecker Delta function cannot be used as the excitation signal to obtain the discrete-time impulse response directly. The answer to this lies in the fact that the Kronecker Delta function is not a sampled version of the Dirac delta function.

When presented with a Kronecker delta, any normal DAC would not recognize that it is a signal that should be converted in a special way. A typical DAC would attempt to create an analog signal by using either a zero-order hold or a first-order hold method for signal reconstruction. This conversion method would cause the DAC to effectively output the nominal output voltage corresponding to $\delta[0] = 1$ for a duration of a single audio sample. Figure 3.3 shows that after the result is passed through a low-pass anti-imaging filter, the output would be a small ripple ($\text{sinc}\left(\frac{\pi t}{T_s}\right)$ if the DAC is ideal), which has much less energy than a Dirac delta impulse and would

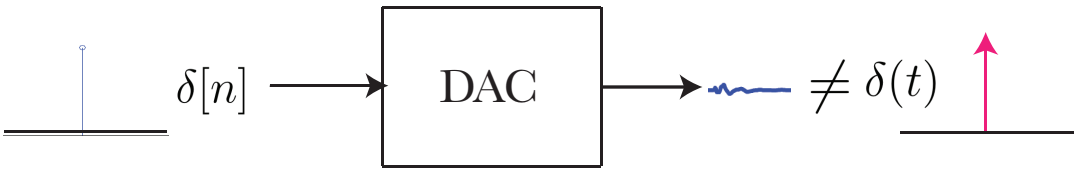

Figure 3.3: Kronecker Delta after Digital-to-Analog Conversion

most likely be inaudible.

In order to characterize the acoustic system seen in Figure 3.2, it is necessary to choose a digital excitation signal, $x[n]$, that is rich in frequency content and energy that can undergo digital-to-analog conversion with a standard DAC to produce an analog excitation signal, $x(t)$, that is a faithful representation of its digital counterpart.

3.3 The Chirp Excitation Signal

The chirp is effectively a sinusoid that increases in frequency over time, so it is a signal that is rich in frequency content and has substantially more energy than the Kronecker delta function. Since modern digital-to-analog converters excel at accurately converting sinusoid-like signals over the audible range, the chirp is a good candidate for an excitation signal that will not be distorted due to hardware limitations. Figure 3.4 shows a chirp excitation signal, $x_c[n]$,

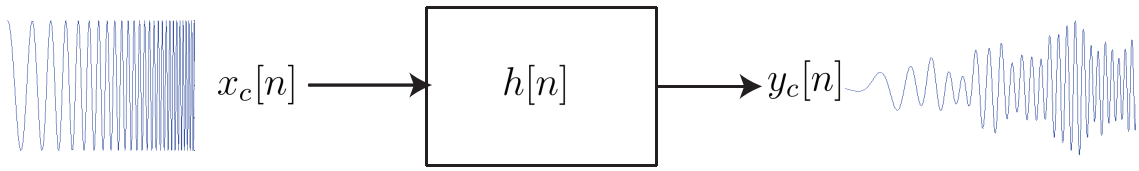

Figure 3.4: Chirp Response

applied to a system with an discrete-time impulse response, $h[n]$, to obtain a chirp response, $y_c[n]$.

The chirp signal is also commonly referred to as a *sweep*, and the terms may be used interchangeably. However, in this document, the signal will be most frequently referred to as a chirp because of the author's familiarity with that term.

The rate at which the instantaneous frequency changes with time can either be linear or logarithmic. The chirp signal used for experiments was a

logarithmic chirp, as the instantaneous frequency increased with the square of time.

3.3.1 Generation of a Chirp Signal

For all experiments, a frequency domain-based implementation by Jont Allen was used to generate a chirp that would have a power spectral density of unity across all frequencies, when evaluated via the FFT [10]. This section presents a derivation of an alternative chirp in the time-domain to illustrate its time-varying instantaneous frequency.

Since the chirp is a time-varying sinusoid, it can be expressed in its most generic form by

$$x(t) = \sin(\theta(t)) \ . \tag{3.8}$$

The chirp can be either linear or logarithmic, depending on the definition of $\theta(t)$ in Equation (3.8). The instantaneous frequency of the sinusoid from Equation (3.8) is the derivative of $\theta(t)$ and can be expressed with respect to time as

$$\Omega(t) = \frac{d}{dt}\theta(t) \ , \tag{3.9}$$

where $\Omega(t)$ the frequency in units radians per second. The conversion to frequency in Hertz leads to a more intuitive representation and can be be done using the relationship

$$\Omega(t) = 2\pi f(t) \ , \tag{3.10}$$

where $f(t)$ is in units of Hertz. By substituting Equation (3.10) into Equation (3.9), we obtain the relationship

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}\theta(t) \ . \tag{3.11}$$

Farina presents a derivation for a linear chirp [11], where the function for the instantaneous frequency takes the form

$$f(t) = f_o + kt \ . \tag{3.12}$$

Combining Equation (3.11) with Equation (3.12) we can begin to solve for a

specific $\theta(t)$ by

$$\frac{1}{2\pi}\frac{d}{dt}\theta(t) = f_o + kt \tag{3.13}$$

$$\theta(t) = 2\pi \lim_{t\to\infty} \int_0^t (f_o + k\tau)d\tau \tag{3.14}$$

$$\theta(t) = 2\pi(f_o + \frac{k}{2}t)t \ . \tag{3.15}$$

Finally, by substituting the result from Equation (3.15) into Equation (3.8), the linear chirp signal can be expressed as

$$x_c(t) = \sin\left(2\pi(f_o + \frac{k}{2}t)t\right) \ . \tag{3.16}$$

The derivation by Farina is useful in illustrating how a chirp can be derived in continuous-time for a given $f(t)$. However, the derivation from Equation (3.12) has limited usefulness as it does not correspond to a well-defined frequency range and is not a discrete-time signal. In order to obtain a more generic solution for a chirp excitation signal, the author took steps to solve for a chirp that would range from $\omega = 0$ to $\omega = \pi$ over N samples. The remainder of this section presents the derivation for an r^{th} order chirp that can be used to generate both linear and logarithmic chirps of length N that cover the entire frequency range of a DAC.

The instantaneous frequency can be expressed as

$$f(t) = f_1 + \left(\frac{t}{T_c}\right)^r (f_2 - f_1) \ , \tag{3.17}$$

where f_1 is the start frequency, f_2 is the end frequency, T_c is the intended duration of the transition from f_1 to f_2, and r is the order of the exponentiation.

For the purposes of system characterization over the entire frequency range of a system, Equation (3.17) can be simplified by making the substitutions of $f_1 = 0$ and $f_2 = f_N$, where f_N is the Nyquist frequency and is related to the sampling frequency, f_s, by

$$f_N = \frac{f_s}{2} \ . \tag{3.18}$$

This reduces Equation (3.17) to

$$\begin{aligned} f(t) &= 0 + \left(\frac{t}{T_c}\right)^r (f_N - 0) && (3.19)\\ &= \frac{f_N}{T_c^r} t^r && (3.20)\\ &= K_r t^r \ , && (3.21) \end{aligned}$$

where K_r has been defined as a constant for an order r chirp.

Replacing Equation (3.12) with Equation (3.21) and solving for $\theta(t)$, we have

$$\begin{aligned} \frac{1}{2\pi}\frac{d}{dt}\theta(t) &= K_r t^r \ , && (3.22)\\ \theta(t) &= 2\pi \lim_{t\to\infty} \int_0^t (K_r \tau^r) d\tau && (3.23)\\ \theta(t) &= \frac{2\pi K_r}{(r+1)} t^{(r+1)} \ . && (3.24) \end{aligned}$$

It is desirable to obtain a discrete-time representation of $\theta(t)$ by sampling at $t = nT_s$, where T_s is the sampling period. Before sampling, note that if the chirp is to be N samples long, then the duration of the chirp, T_c, can be converted to

$$\begin{aligned} T_c &= NT_s \\ &= \frac{N}{f_s} \ . && (3.25) \end{aligned}$$

After sampling $\theta(t)$, Equation (3.24) can be reduced to

$$\begin{aligned}
\theta[n] &= \frac{2\pi K_r}{(r+1)}(nT_s)^{(r+1)} \qquad (3.26)\\
&= \frac{2\pi}{(r+1)}\frac{f_N}{T_c^r}(nT_s)^{(r+1)}\\
&= \frac{2\pi}{(r+1)}\frac{f_s}{2}\frac{1}{T_c^r}(nT_s)^{(r+1)}\\
&= \frac{2\pi}{(r+1)}\frac{f_s}{2}\left(\frac{f_s}{N}\right)^r(nT_s)^{(r+1)}\\
&= \frac{2\pi}{(r+1)}\frac{f_s}{2}\frac{f_s^r}{N^r}\left(\frac{n}{f_s}\right)^{(r+1)}\\
&= \frac{2\pi}{(r+1)}\frac{f_s}{2}\frac{f_s^r}{N^r}\left(\frac{n}{f_s}\right)^{(r+1)}\\
&= \frac{2\pi}{(r+1)}\frac{1}{2}\frac{f_s^{(r+1)}}{N^r}\frac{n^{(r+1)}}{f_s^{(r+1)}}\\
&= \frac{\pi}{N^r(r+1)}n^{(r+1)} \qquad (3.27)
\end{aligned}$$

Finally, by sampling Equation (3.8) and substituting Equation (3.27) for $\theta[n]$, a discrete-time representation of a chirp can be expressed as

$$x_c[n] = \sin\left(\frac{\pi}{N^r(r+1)}n^{(r+1)}\right) \ . \qquad (3.28)$$

If $r = 1$ is selected, the resulting chirp will be linear, and if $r > 1$ the chirp will be logarithmic.

Note that Equation (3.28) has no dependence on the sampling frequency. It is simply designed to sweep from a start frequency of $\omega_1 = 0$ to an end frequency of $\omega_2 = \pi$ over a duration of N samples. Therefore a chirp, $x_c[n]$, that is constructed from Equation (3.28) can be reused at any sampling frequency to sweep over the entire range of the DAC. The only difference in switching from one sampling rate to another will be the change in the physical duration of the chirp, T_c, which is determined by the number of samples and the sampling frequency.

It should be restated that the chirp from Equation (3.28) was not used in experiments; however, the first order chirp with $r = 1$ was verified to measure simulated impulse responses in MATLAB. It should also be noted that Equation (3.28) does not encompass all possible solutions for linear and

logarithmic chirps but was derived by the author for the specific purpose of providing a closed solution capable of generating either a linear or logarithmic chirp that spans the maximum frequency range.

3.3.2 The Chirp Train

In Figure 3.5 a system is shown that has an impulse response $h[n]$. If this is an LTI system, then it will theoretically also have a unique chirp response, denoted $y_c[n]$. However, because of noise in the system, every measurement will be corrupted by some degree. For this reason, a distinction is also made between the theoretical chirp response, $y_c[n]$, and the i^{th} chirp response window, which will be denoted $\tilde{y}_c^i[n]$.

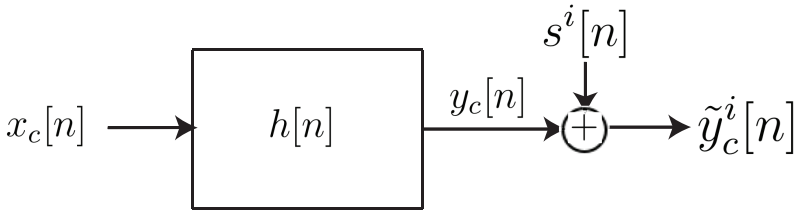

Figure 3.5: Chirp Response with Noise

The estimated chirp response, $\hat{y}_c[n]$, should contain exactly the same number of samples as the chirp excitation signal $x_c[n]$. Also, any delay in the chirp response can be recovered only if it causes $\hat{y}_c[n]$ to be a circularly shifted version of $x_c[n]$. This constraint can make things tricky, since the delay in the system is not known until after the impulse response calculation.

To address the problem with delay and the need for a circularly shifted $\hat{y}_c[n]$, a chirp must be repeated at least once in a *chirp train*. After a chirp train is passed through a system, the resulting train of chirp responses are segmented into a series of contiguous chirp response windows, $\tilde{y}_c^i[n]$, where the i^{th} window should contain the response for the i^{th} chirp, $x_c^i[n]$, in the chirp train.

The first measured chirp response window, $\tilde{y}_c^1[n]$, will almost certainly contain zero values from the first sample up to the onset of the chirp response because of system delay. The delay will also shift important samples of the chirp response into the next chirp response window. Since it is very likely

that ideal chirp response $y_c[n]$ will not be contained $\tilde{y}_c^1[n]$ due to shifting, this window should be discarded from the measurement calculations.

Unlike $\tilde{y}_c^1[n]$, the second measured chirp response window, $\tilde{y}_c^2[n]$, will begin with all measured samples that should have been in $\tilde{y}_c^1[n]$ but were shifted out of the window due to system delay. Directly following the samples shifted in from the previous response, this window will also contain the initial response from the second chirp excitation signal, $x_c^2[n]$. The windowing of a chirp train effectively satisfies the requirement that, in addition to filtering, $\tilde{y}_c^2[n]$ is a circularly shifted version of $x_c[n]$.

The subsequent chirp response windows should all effectively obtain circularly shifted chirp response measurements; however, all of these measurements will independently be corrupted to some degree by noise.

3.3.3 Improving SNR using a Chirp Train

In order to improve the resilience of the chirp response in the presence of noise, multiple chirps presented: in a chirp train can be averaged.

After obtaining M independent windows of the chirp response, an improved estimate of the system chirp response can be made by averaging over the M independent windows by

$$\hat{y}_c[n] = \frac{1}{M} \sum_{i=2}^{M} \tilde{y}_c^i[n] \tag{3.29}$$

where $\hat{y}_c[n]$ is the chirp response estimate.

Using the model from Figure 3.5, each measured chirp response, $\tilde{y}_c^i[n]$, can be represented as the ideal chirp response, $y_c[n]$, corrupted by a vector of noise on the i^{th} measurement, $s^i[n]$, which leads to

$$\hat{y}_c[n] = \frac{1}{M} \sum_{i=2}^{M} \left(y_c[n] + s^i[n] \right) \ . \tag{3.30}$$

Since the ideal chirp response, $y_c[n]$, does not change for each measurement, it can be pulled out of the summation, which leaves

$$\hat{y}_c[n] = y_c[n] + \frac{1}{M} \sum_{i=2}^{M} s^i[n] \ . \tag{3.31}$$

If it is assumed that $s^i[n]$ is stationary Gaussian noise, Equation (3.31) indicates that doubling the number of chirps in the chirp train, M, will have the effect of reducing the noise power by a factor of two.

3.3.4 Impulse Response from a Chirp Response

The impulse response can be found through the deconvolution of the chirp response with the chirp excitation signal. [12].

Deconvolution is a process in which a response signal, $y[n]$, is convolved with the inverse of the excitation signal, $x'[n]$, resulting in

$$h[n] = y[n] * x'[n] \ . \tag{3.32}$$

The inverse of the signal here means that it is the inverse with respect to convolution and has the property that

$$x[n] * x'[n] = \delta[n] \ . \tag{3.33}$$

The inverse of the excitation signal can be calculated using the FFT and IFFT by

$$\begin{aligned} X[k] &= FFT(x[n]) & (3.34)\\ X'[k] &= \frac{X[-k]}{|X[k]^2|} & (3.35)\\ x'[n] &= IFFT(X'[k]) \ . & (3.36) \end{aligned}$$

Deconvolution in the time domain can be difficult to apply and is computationally inefficient. However, deconvolution is much more straightforward when applied in the frequency domain. Just as convolution can be performed in the frequency domain by multiplication, deconvolution can be performed by the division of $Y[k]$ by $X[k]$. In this method, the deconvolution is performed much faster because of the use of the FFT and IFFT algorithms.

$$h_c[n] = IFFT\left(\frac{FFT(\hat{y}_c[n])}{FFT(x_c[n])}\right) \ . \tag{3.37}$$

Equation (3.37) illustrates how to apply deconvolution in the frequency domain to efficiently compute the impulse response from the chirp excitation

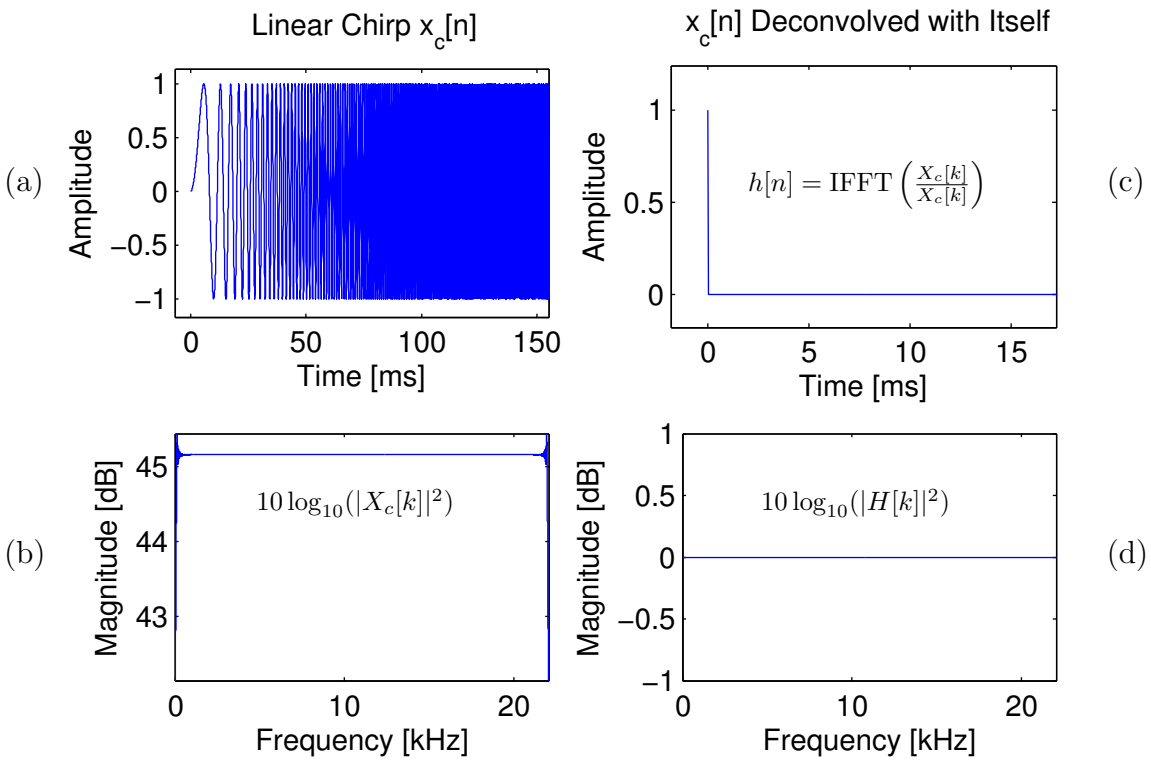

Figure 3.6: Properties of Linear Chirp

signal and an averaged chirp response. For clarity, an impulse response estimate obtained from a chirp response is denoted $h_c[n]$.

3.3.5 Linear vs. Logarithmic Chirps

Figure 3.6 shows the properties of a linear chirp constructed from Equation (3.28) when a value of $r = 1$ is selected with $N = 2^{16}$, and assumes $f_s = 44.1$ kHz. In Figure 3.6(a) the first 150 ms illustrates the sinusoidal characteristics of linear chirp as it increases in frequency. Figure 3.6(b) shows that the power spectral density (PSD) of the linear chirp is evenly distributed across all frequencies. The power spectral density of Figure 3.6 was estimated using the FFT by

$$\text{PSD}(x[n]) = |\text{FFT}(x[n])|^2 \ . \tag{3.38}$$

The magnitude of the PSD is dependent on the length of the chirp, N, so a longer chirp excitation signal would raise the magnitude of the energy seen for each frequency bin in Figure 3.6(b).

A ripple of around 3 dB for frequencies under 200 Hz can be seen in

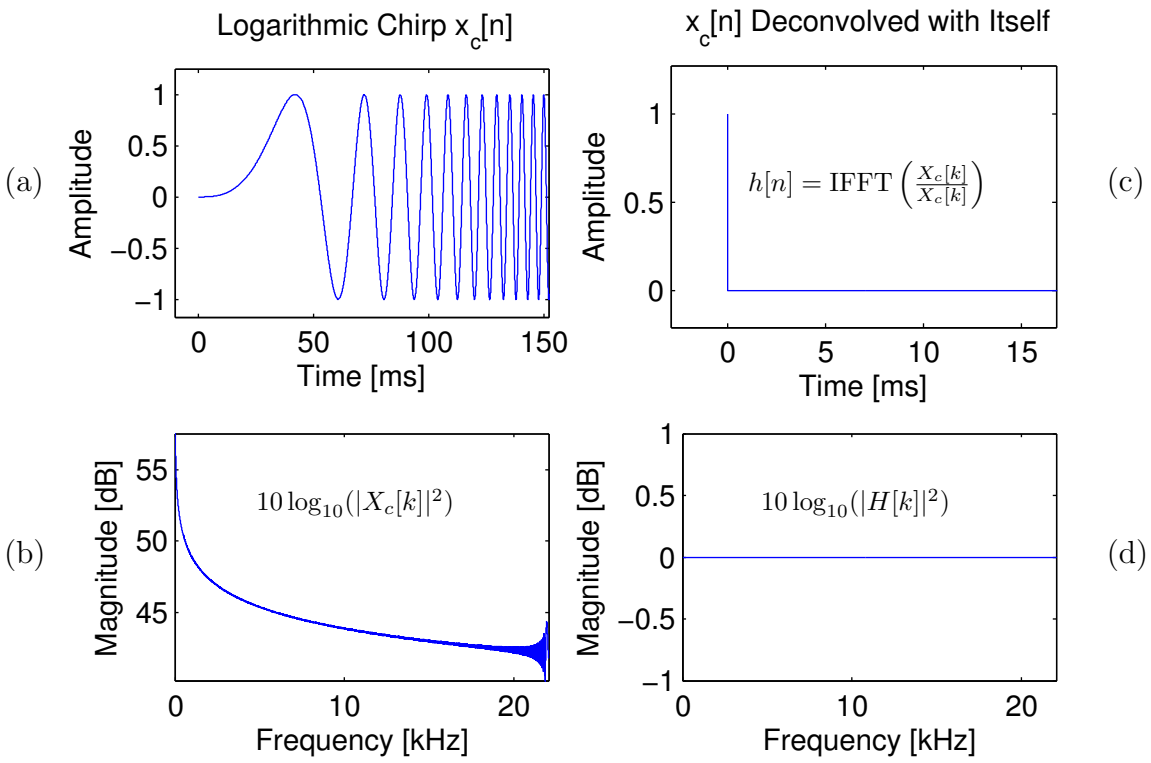

Figure 3.7: Properties of Logarithmic Chirp

Figure 3.6(b); however, Figure 3.6(d) shows that the ripple in PSD does not have any bearing in the effectiveness of the chirp to produce flat frequency response for these frequencies when passed through a system with a flat frequency response.

Figure 3.7 shows the properties of a logarithmic chirp constructed from Equation (3.28) when a value of $r = 2$ is selected. Again, the length of the chirp is $N = 2^{16}$, and it is assumed that $f_s = 44.1$ kHz. In Figure 3.7(a) it can be seen that the logarithmic chirp spends more time at lower frequencies within the first 150 ms when compared to the linear chirp. The emphasis placed on lower frequencies can also be seen in the PSD of the logarithmic chirp in Figure 3.7(b).

It can be seen in Figure 3.7(c) that, even though it does not have flat power spectral density, the logarithmic chirp is perfectly capable of characterizing a system that has an impulse response of a Kronecker delta function in the same way that the linear chirp did in Figure 3.6(c). In fact, logarithmic chirps are sometimes preferred in system characterization because the uneven spectral energy will result in a higher SNR for lower frequencies.

In Chapter 6, the ability of some chirps with uneven PSD to character-

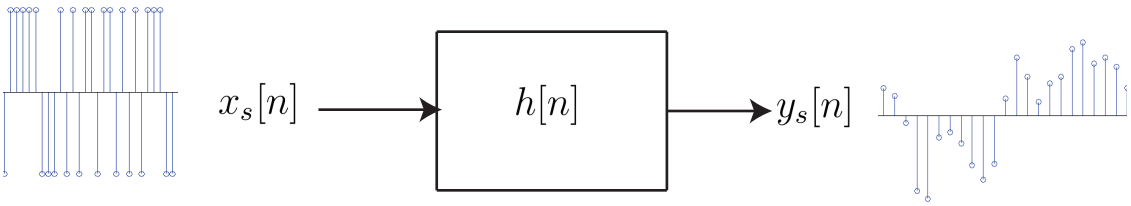

Figure 3.8: MLS Response

ize a system is used as the basis for an experimental approach to relative equalization.

3.4 The MLS Excitation Signal

Figure 3.8 shows a maximum length sequence (MLS) excitation signal, $x_s[n]$, applied to a system with a discrete-time impulse response, $h[n]$, to obtain a MLS response, $y_s[n]$.

A maximum length sequence is a discrete signal that resembles a pseudorandom sequence but has special properties that make it desirable for impulse response measurements. Since the MLS is rich in frequency content and energy, it is a viable choice for use as an excitation signal.

However, because the MLS requires abrupt transitions from -1 to $+1$, questions have been raised concerning whether a standard DAC will introduce nonlinear artifacts during the digital-to-analog conversion that may need to be compensated for in impulse response calculations [13]. According to previous research, it is possible that the MLS can expose nonlinear effects in several aspects of the sound production, such as the loudspeaker itself [14].

It can be difficult for a researcher to be confident in the experimental results when presented with such potential drawbacks in the MLS measurement methods. In particular, if a feature of the transfer function abruptly changes, it is desirable to be able to identify it as a characteristic of the physical system with confidence and reduce doubt that it is distortion introduced by the measurement method.

3.4.1 Generation of a Maximum Length Sequence

The maximum length sequences were generated using an implementation by Ludwig [15] based on a shift-register generator designed by Peterson and Weldon [16].

To begin generation of a MLS, the length must be chosen based on rounding the desired number of samples to the nearest power of 2. A k-bit shift-register generator is capable of generating a sequence of length $2^k - 1$.

Next, a specific coefficient vector, $\bar{\alpha}$, must be found that will generate a sequence with a periodicity of $N = 2^k - 1$, which is considered the *maximum length* of a sequence that can be generated by a k-bit shift-register. Each value in the coefficient vector, α_i, can take on a value of either 0 or 1, and there is a specific solution for $\bar{\alpha}$ that is unique for a k-bit implementation of the shift-register generator. The purpose of $\bar{\alpha}$ is to effectively choose which values of the shift-register to include in the summation before shifting takes place. The derivation of the coefficient vector is beyond the scope of this document, but they are solved for many values of k in [16].

Using the correct $\bar{\alpha}$ derived for the chosen value of k, the maximum length sequence is recursively defined relative to its previous k values by the difference equation,

$$x[n] = \left(\sum_{i=1}^{k} \alpha_i x[n-i] \right) \mod 2, \tag{3.39}$$

where the modulo operation is performed on the result of the summation. Note that if Equation (3.39) is implemented using a physical shift-register, the modulo operation may be implicitly performed as the underlying hardware stores the result of the summation in a single-bit position of the register, which can only take on a value of 0 or 1.

In the shift-register implementation, the values of $x[n-1]$ to $x[n-k]$ are each stored as a single bit in a shift-register and are initialized to 1. If all bits in the shift-register were ever zero, then the recursive defininition in Equation (3.39) would result in $x[n] = 0 \ \forall \ n$. These initial shift-register values for $n < 1$ are not considered to be a part of the MLS.

The value for $x[n]$ is calculated by summing a subset of $x[n-1]$ to $x[n-k]$ according to $\bar{\alpha}$, and then performing the modulo 2 on the result to limit $x[n]$ to 0 or 1. When the value of $x[0]$ is calculated, all values in the register are shifted so that $x[0-k]$ is discarded, and $x[0]$ takes the place of the open bit.

After shifting, Equation (3.39) is performed again to calculate the value of $x[1]$. These steps are repeated to calculate $x[2]$, and so forth.

A k-bit shift-register is capable of having 2^k states, each corresponding to the unique binary number represented by the bit in each location of the shift-register. In the generation of a maximum length sequence, the shift-register will take on all possible states, except for the zero vector. The omission of the zero vector is the reason a MLS generator will not produce a MLS that is periodic with an even power of two but rather have a periodicity of $2^k - 1$ [16].

3.4.2 Mathematical Properties of MLS

Figure 3.9(a) shows the first 5 ms of a MLS excitation signal taking on values of ± 1. Figure 3.9(b) shows the power spectral density of the entire MLS excitation signal, which is nearly flat over all frequencies.

The following section discusses the properties of maximum length sequences that make them suitable for obtaining an estimation of the impulse response of a system. The derivation presented here is slightly simplified from [17] to deal with a digital system where the MLS excitation signal can only take on values of ± 1.

In any discrete system, the output signal, $y[n]$, is related to the input signal, $x[n]$, through linear convolution with the impulse responses, $h[n]$, by

$$y[n] = x[n] * h[n] \tag{3.40}$$

$$= \sum_{k=-\infty}^{\infty} x[k]h[n-k] \ . \tag{3.41}$$

It should be emphasized that *linear* convolution, denoted by the * operator in equation 3.40, is needed to transform the input signal to the output signal via the impulse response.

The property that makes a MLS well suited to estimate the impulse response of a system stems from the fact that its circular autocorrelation,

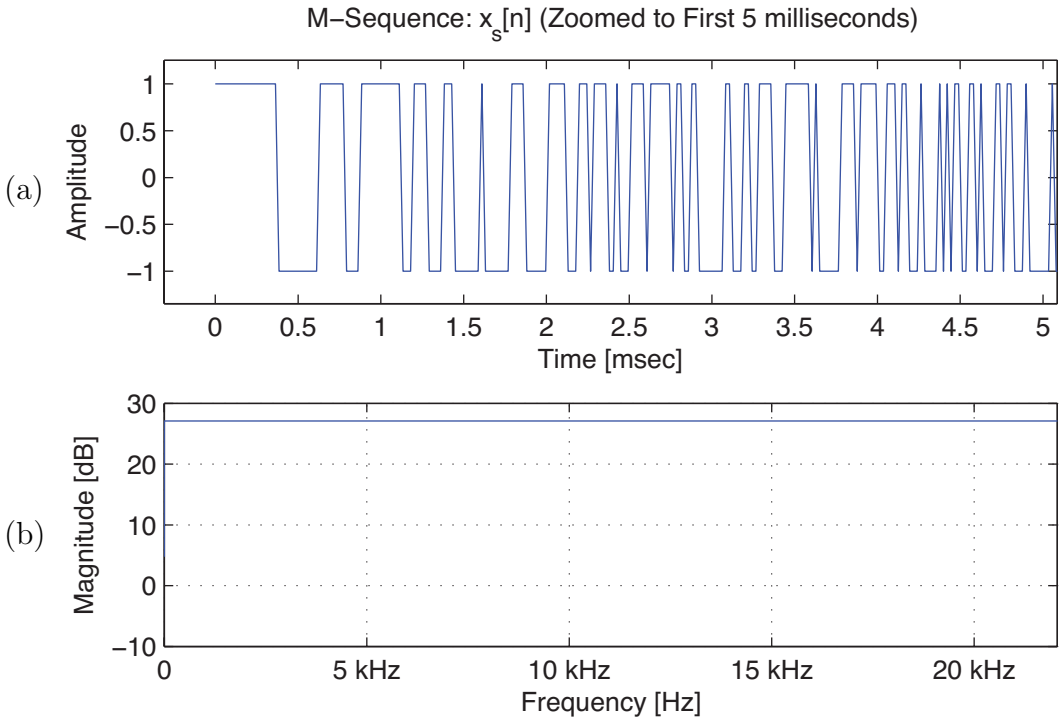

Figure 3.9: Maximum Length Sequence: $x_s[n]$

$R_{x_s x_s}[n]$, is nearly a delta function, as seen in Equation (3.43). [17]

$$R_{x_s x_s}[n] = \frac{1}{L+1}\sum_{k=0}^{L-1} x_s[k]x_s[n+k] \tag{3.42}$$

$$= \delta[n-l] - \frac{1}{L+1} \tag{3.43}$$

The function, $R_{x_s y_s}[n]$, will be used to denote the circular cross correlation of the MLS excitation signal, $x_s[n]$, and the MLS response, $y_s[n]$. Combining the MLS property of Equation (3.43) with the definition of the impulse response from Equation (3.40), $R_{x_s y_s}[n]$ can be expressed in terms of the

impulse response by

$$
\begin{aligned}
R_{x_s y_s}[n] &= \frac{1}{L+1}\sum_{k=0}^{L-1} x_s[k]y[n+k] && (3.44)\\
&= \frac{1}{L+1}\sum_{k=0}^{L-1}\sum_{l=-\infty}^{\infty} x_s[k]x_s[n+k-l]h[l] && (3.45)\\
&= \sum_{l=-\infty}^{\infty} h[l]\frac{1}{L+1}\sum_{k=0}^{L-1} x_s[k]x_s[n+k-l] && (3.46)\\
&= \sum_{m=-\infty}^{\infty}\sum_{l=0}^{L-1} h[l+mL]\left[\delta[n-l]-\frac{1}{L+1}\right] && (3.47)\\
&= \sum_{m=-\infty}^{\infty} h[n+mL] - \frac{1}{L+1}\sum_{m=-\infty}^{\infty} h[m]\ . && (3.48)
\end{aligned}
$$

The cross-correlation of the MLS excitation signal with the MLS response is essentially the impulse response that will wrap around to the beginning after L samples and is very slightly shifted by the second term of 3.48.

Therefore, an estimate of the impulse response from a MLS excitation signal will be defined as

$$\hat{h}_s[n] \equiv R_{x_s y_s}[n] \tag{3.49}$$

When L is sufficiently large, the second term of Equation (3.48) tends to zero and can safely be discarded, leaving an impulse response that is repeated every L samples, as in Equation (3.50).

$$
\begin{aligned}
\hat{h}_s[n] &= R_{x_s y_s}[n]\\
&= \sum_{m=-\infty}^{\infty} h[n+mL] - \frac{1}{L+1}\sum_{m=-\infty}^{\infty} h[m]\\
&= \sum_{m=-\infty}^{\infty} h[n+mL] - \epsilon\\
&\doteq \sum_{m=-\infty}^{\infty} h[n+mL] && (3.50)\\
&\approx h[n] && (3.51)
\end{aligned}
$$

If the length, L, of the MLS excitation signal is chosen carefully to ensure that it is sufficiently longer than the impulse response, the fact that $h[n]$

is wrapped every L samples will be of little or no consequence because the value of the wrapped samples of $h[n]$ will essentially be zero value. When this criterion is met, the approximation of Equation (3.51) is valid.

3.4.3 Impulse Response from a MLS Response

If circular cross-correlation is to be performed, it requires that the excitation signal and the estimated excitation response are the same number of samples. This constraint was also imposed for the chirp response, and the difficulty of creating a response with a circular delay was solved by creating a chirp train and discarding the first windowed response. A similar solution can be applied to the MLS response by immediately repeating the MLS and discarding the first window.

Cross-correlation in the time-domain is computationally inefficient and is mathematically equivalent to the convolution operation with a time-reversed signal as

$$R_{x_s y_s}[n] = y_s[n] * x_s[-n] \ . \tag{3.52}$$

The convolution in Equation (3.52) can be efficiently applied in the frequency domain to obtain the impulse response by

$$\hat{h}_s[n] = IFFT\left(FFT(y_s[n]) \times FFT(x_s[-n])\right) \ , \tag{3.53}$$

recalling that the circular cross-correlation, $R_{x_s y_s}[n]$, was defined to be the estimator of the impulse response in Equation (3.49).

CHAPTER 4

EXPERIMENT 1: EXCITATION RESPONSE SYNCHRONIZATION

4.1 Variability in Sound-Acquisition Software

To ensure valid measurements, extra care was taken to verify that a measurement would be reproducible for each angle of the HRTF. In order to test for reproducibility after the KEMAR had been positioned, each excitation signal was repeated several times and the response signal was collected for each repetition. Since repeated measurements were separated by only a few seconds, it was assumed that all environmental factors would remain constant, including the position of the probe microphone, cables, and the KEMAR.

If all systems are operating correctly, then the only theoretical difference between independent measurements would be due to system noise. However, surprisingly, after comparison of independent measurements, it was found that the resulting impulse responses were shifted in time relative to one another. In fact, these repeated impulse response measurements could shift up to 120 milliseconds.

Visually, each impulse response appeared to be similar but was shifted in time. To perform an initial verification of reproducibility, all samples before the impulse response were manually removed in MATLAB, effectively shifting each impulse response to the beginning of the measured signal. After each manually shifted measurement was superimposed in MATLAB, it was apparent that the impulse responses were nearly identical but had been shifted due to an unknown reason.

After inspecting the system and troubleshooting, the issue was finally root caused. The program to issue the excitation signal and record its response was written in the MATLAB scripting language using the built-in *wavplay* and *wavrecord* functions. As of MATLAB version R2007b, there was no functionality available to synchronize the functionality of recording and playback

so that they would always be reading in lockstep with respect to the audio buffer in the soundcard. In the execution of the program, the excitation was initiated by the *wavplay* function, which would presumably spawn a thread to configure the soundcard for output and begin writing to the output buffer. Then the program would continue to the next line in the MATLAB script where the *wavrecord* function was invoked. This function would also presumably spawn a separate thread to configure the soundcard for input and would independently begin to read from the recorded audio buffer.

The delay was caused due to the independence of threads spawned by the *wavplay* and *wavrecord* matlab functions. The starting time of these threads was therefore largely dependent on the operating system and any other activity that was happening on the PC at the time the MATLAB script was executed. Sometimes this could result in sub-millisecond discrepancies and other times it could result in up to 120-millisecond discrepancies.

This variability presented a major challenge in validating the reproducibility of each measurement, since manually shifting and inspecting several impulse responses for each angle of the HRTF would be too costly and time-consuming to be feasible. It was clear that the synchronization issue across independent measurements would need to be rectified in another way.

Some time was spent investigating if a custom interface to the soundcard could be written in C, but this would also require a large amount of cost and would be customized for the particular soundcard in use. Another drawback of this approach is that the record/playback program would have limited portability across operating systems. Ultimately, it was concluded that the most desirable solution would be to devise a MATLAB function capable of synchronizing independent measurements relative to some reference time. If such a function were capable of ensuring sample-to-sample accuracy across measurements, it would be sufficient for automating the task of synchronization of repeated measurements, as well as for independent measurements of the HRTF for different azimuthal angles.

4.2 Response Synchronization via the MLS Burst

In the previous chapter it was explained that when a maximum length sequence is sent through a system, the impulse response of the system can be

obtained by cross-correlating the MLS response with the original maximum length sequence.

It was the cross-correlation properties of the maximum length sequence that made it a desirable signal for the function of time synchronization. If a maximum length sequence is shifted by N samples to the right and cross-correlated with its unshifted version, the result will be a Kronecker delta function shifted by N samples to the right and scaled by the number of samples in the sequence.

Similarly, if a MLS excitation signal is filtered and delayed, the maximum value of the cross-correlated response will most likely be delayed by the same number of samples each time. It is foreseeable that an exception may occur when the numerical value for the maximum is shared (or almost the same) for two adjacent samples of the impulse response, but this was not observed in this experiment.

4.2.1 An MLS Burst Augmented Excitation Signal

In order to synchronize measurements, a MLS burst was prepended to each excitation signal. Figure 4.1(a) shows the response data obtained from two independent excitation responses augmented with the following:

- *Initial Silence*: This is a sequence of zero-valued samples. It is expected that some of these samples may be missing from the response data, depending on when the record function begins. The length of this sequence should be used to increase the likelihood that recording has begun before the MLS synchronization burst is sent. For this experiment, a duration of 0.5 seconds was chosen for the initial silence.

- *MLS Synchronization Burst*: This is a short maximum length sequence, whose sole purpose is to provide a reference signal for synchronization. The length of the MLS should be chosen so that a suitable maximum can be found after cross-correlation. In this experiment, the length of 2^{16} samples was used at $44.1kHz$ to result in a 1.5-second MLS synchronization burst.

- *Intermediate Silence*: This is a sequence of zero-valued samples between the MLS synchronization burst and the excitation signal. The length

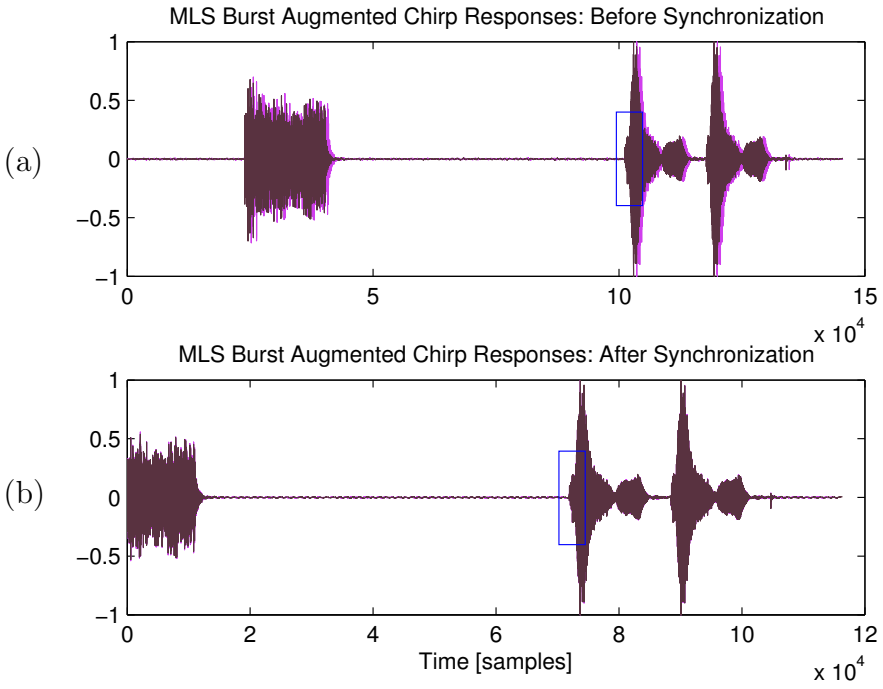

Figure 4.1: Synchronizaion of Independent Chirp Responses

of this silence should be determined by the reverberation time of the room with a reasonable safety factor, so that any reverberation from the MLS synchronization burst has dissipated before the excitation signal begins. The KEMAR laboratory had an approximate reverberation time of one second, so an intermediate silence of 3 seconds was chosen for this experiment.

- *Excitation Signal*: The excitation signal would take the form of either a train of chirps or a maximum length sequence.

To ensure that no portion of the excitation response would be inadvertently truncated, the *wavrecord* function was configured to record for one second longer than the length of the augmented excitation signal.

The boxed areas of Figure 4.1 contain the onset of the chirp responses before and after synchronization. Figure 4.2 is zoomed to the boxed area showing the onset of the chirp responses.

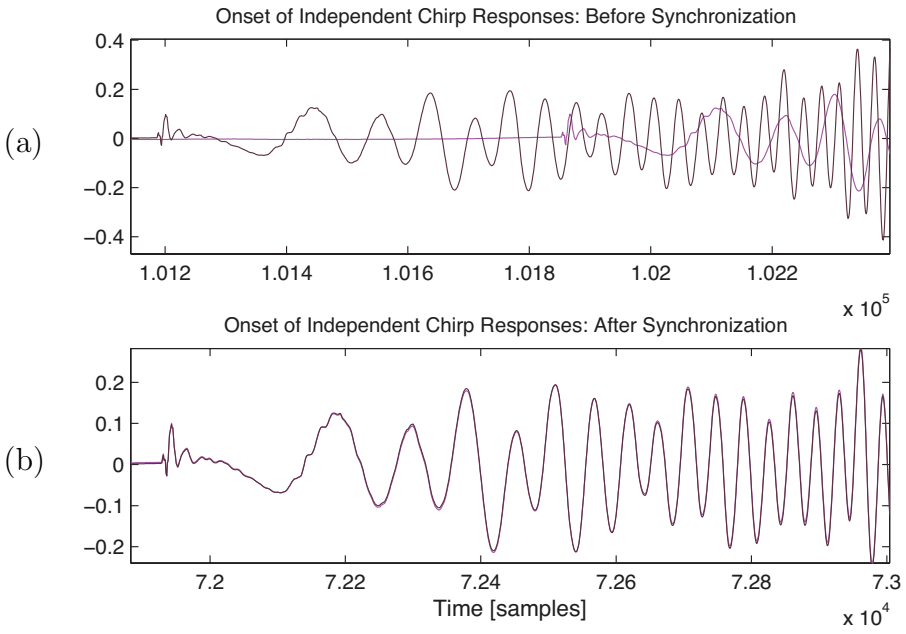

Figure 4.2: Onset of Independent Chirp Responses

4.2.2 Response Synchronization

After the response data was collected, the MLS burst was used to perform the synchronization. The first task of performing the synchronization was to isolate the MLS synchronization burst apart from the excitation signal. If no extra latency had been incurred between the playback/record of the augmented excitation signal, the MLS synchronization burst would have completed after 2 seconds (0.5 seconds of silence plus 1.5 seconds of the burst). However, it likely would have been shifted by up to 120 milliseconds or 0.12 seconds during the measurement. With this in mind, the first 3 seconds of the augmented excitation signal were selected to isolate the synchronization burst with silence on either side. For the synchronization step, all samples after 3 seconds (including the excitation signal) were discarded.

Once the response for the synchronization burst was isolated, the signal was cross-correlated with the original MLS synchronization burst. Then the index of the maximum value of the cross-correlated signal was found. This index was referred to as the *synchronization index* because it would theoretically occur at the same time relative to the beginning of the excitation signal.

To perform synchronization, the entire augmented excitation signal was first left-shifted (with truncation) by the value of the synchronization index. This shift effectively eliminated the initial silence, so that the remaining sig-

nal theoretically began at or near the first sample of the MLS synchronization burst. Next, the synchronized augmented excitation signal was left-shifted by the length of the MLS synchronization burst and the intermediate silence so that only the synchronized excitation response remained. To reduce the risk of truncating the excitation response, the last 64 samples of the intermediate silence were kept in the excitation response data.

4.3 Results

In Figure 4.2(a), the onset of the unsynchronized responses differs by approximately 700 samples, which corresponds to a discrepency of 15.9 ms for a 44.1 kHz sampling frequency. A discrepency of this magnitude would make independent measurements unreliable to measure the interaural time difference (ITD) of the HRTF, which is less than 1 ms at its maximum.

Figure 4.2(b) shows the onset of the chirp responses after the synchronization steps have been performed. These synchronized chirp responses are capable of producing impulse responses whose maximum occurs reliably at the same sample index. Thus, the synchronization method is capable of aligning independent measurements with sample-by-sample accuracy.

CHAPTER 5

EXPERIMENT 2: COMPARISON OF PROBE AND IN-EAR MICROPHONES IN HRTF MEASUREMENTS

Chapter 4 addressed the need for synchronization across independent measurements of an excitation response, before a reproducible impulse response can be calculated for a system.

This chapter builds on the results of Chapter 4 by incorporating the MLS burst method of synchronization for every measured MLS response and chirp response. When the head-related impulse response (HRIR) measurements are presented in figures, both the MLS- and chirp-derived impulse responses are superimposed for a reference of relative accuracy.

The focus of this chapter is to test for potential discrepancies that may arise when attempting to obtain a HRIR using a probe microphone. The issues that are tested include the potential for multipath interference for a probe microphone located outside the KEMAR, and an attempt to dereverberate a measurement obtained by a probe microphone.

5.1 Equipment and Room Configuration

The Event EZ Bus was used as the DAC/ADC in all experiments. A desktop PC with an Pentium 4 processor and MATLAB R2007b was interfaced to the EZ Bus via a single USB 1.0 cable, which ran from the adjacent room under a sound-treated door. All audio equipment, with the exception of the loudspeaker and stand, were grouped as far from the KEMAR as possible, on the opposite side of the room. The experiment was designed to allow for the existence of early reflections from equipment and supporting structures that were in the room. Figure 5.1 shows the position of the KEMAR relative to the sound source.

The KEMAR laboratory had been built for the intelligent hearing aid project conducted in the Beckman Institute. The room that housed the KE-

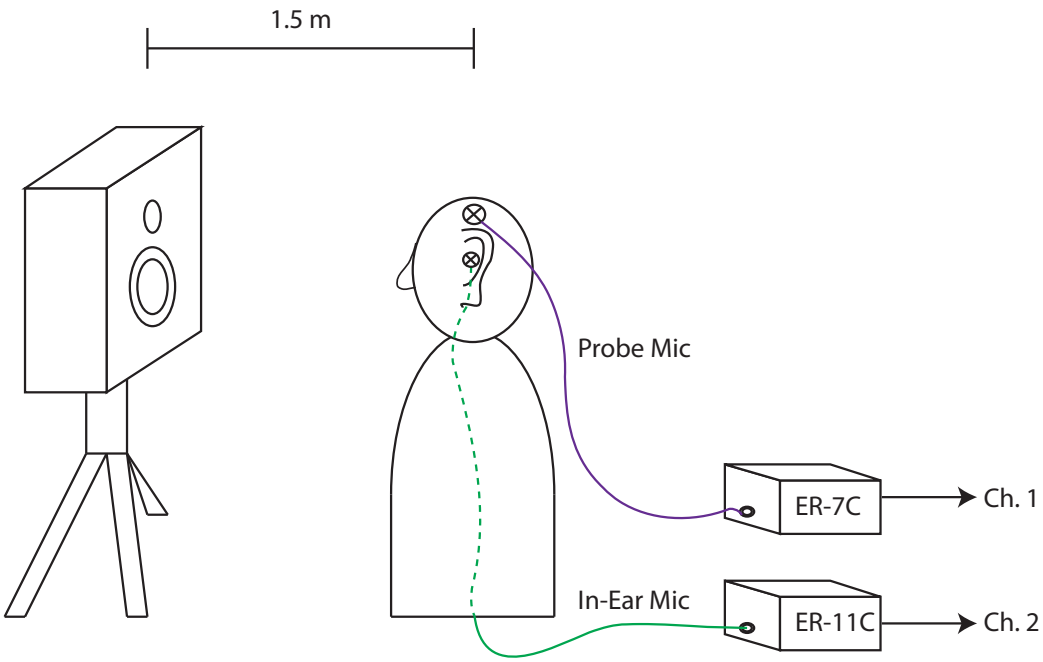

Figure 5.1: Room Configuration and Probe Microphone Placement

MAR was sound-treated with styrofoam on the walls and doors. Although the walls were treated to increase sound absorption, the room was not anechoic. Other reflective surfaces included the floor, ceiling tiles, and ventilation ducts. The floor was simply carpeted, and the ceiling tiles and ventilation ducts were left untreated. The laboratory had a notable reverberation time that was mostly due to reverberation in the ventilation system.

To reduce the reverberation time, styrofoam was also placed over the ventilation ducts when experiments were performed. The exact value for the reverberation time of the KEMAR laboratory was not needed for this experiment, but it was informally observed to be on the order of 1 second.

As previously described in section 3.4.2, in order to avoid corruption due to time-domain wrapping of the room impulse response tail, the excitation signal must be longer than the reverberation time of the room. This requirement was taken into account by using excitation signals that were at least three times as long as the estimated reverberation time.

With the reverberation time accounted for, the early reflections caused by reflective surfaces in the room needed to be addressed.

For this experiment, two types of microphones were used. The first microphone is the ER-11, developed by Etymotic Research for use with KEMAR.

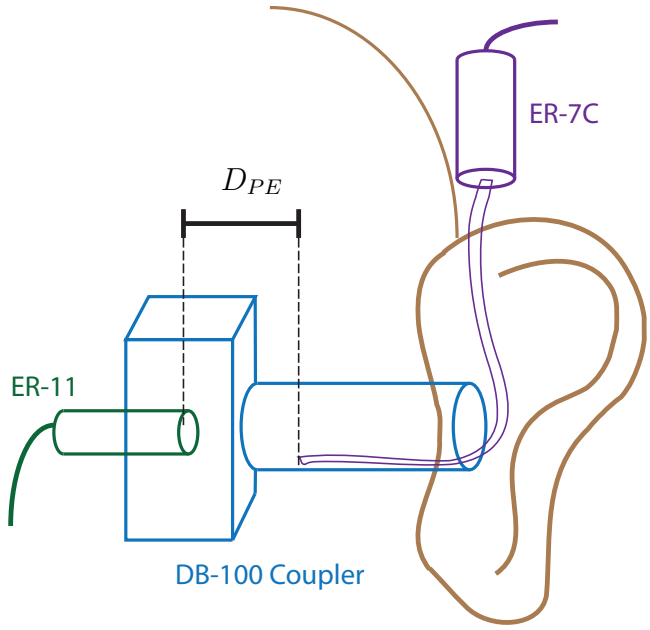

Figure 5.2: KEMAR Microphone Placement and DB-100 Zwislocki Coupler

Each ear of the KEMAR is equipped with a DB-100 to simulate the ear canal, which is terminated with an ER-11 microphone. All microphones used in these experiments were manufactured by Etymotic Research.

The Etymotic ER-11 is a microphone specifically designed to use with the DB-100 Zwislocki Coupler that is shown in Figure 5.2. The DB-100 is used to simulate the human ear canal, and the ER-11 microphone is placed so that it is analogous to the eardrum. The probe microphone used for this experiment was the ER-7C, which uses a probe tube to measure the SPL in cavities such as the ear canal. The physical microphone of the ER-7C is contained in a housing that was placed outside the KEMAR head above the ear, so that the probe tube could reach into the DB-100 without significantly altering the response seen by the ER-11 microphone.

For the rest of the experiment, the ER-11 will be referred to as the *in-ear* microphone, and the ER-7C will be referred to as the *probe* microphone. The distance between the in-ear microphone and the termination of the probe tube is denoted D_{PE} in Figure 5.2.

5.2 Addressing Early Reflections

When working in an imperfect acoustic space that is not anechoic, perhaps the most obvious form of multipath interference of concern is that of early reflections from walls and other objects in the environment. It may seem that such an environment would make it impossible to obtain an uncorrupted measurement of the head-related impulse response. However, if the distances of most reflective boundaries and objects are taken into account, the environment can be arranged in such a way that it allows for the impulse response to be isolated from nearly all other early reflections.

Previous work has indicated that the main part of the HRIR lies within the first 5 to 6 milliseconds of the impulse response [18]. The measurements contained in the CIPIC database consist of the first 256 samples of the HRIR sampled at 44100Hz, which takes a time of

$$256[samples] \times \frac{1}{44100} \frac{[second]}{[samples]} = 5.8 \; milliseconds \; . \tag{5.1}$$

Therefore, if the room can be arranged in a way that will cause early reflections to arrive more than 6 milliseconds after the direct path of the excitation signal, then these reflections will present themselves after the pertinent section of the HRIR. Therefore, for the following experiments, it is maintained that 6 milliseconds of an uncorrupted HRIR is sufficient for an accurate representation of the HRIR as though it had been measured in an anechoic chamber.

The sample length of the chirp excitation signal, L, was determined based on the sample rate of 44.1 kHz and the desired duration of the signal. To improve the computational speed of FFT calculations, it was desirable that the signal length be a power of two. In addition, in order to satisfy the criterion that the excitation signal was at least twice the length of the reverberation time of the room, a sample length of 2^{18} samples was chosen for each chirp of the chirp train. This corresponded to a 5.9-second chirp excitation signal, which was at least three times the reverberation time of the room.

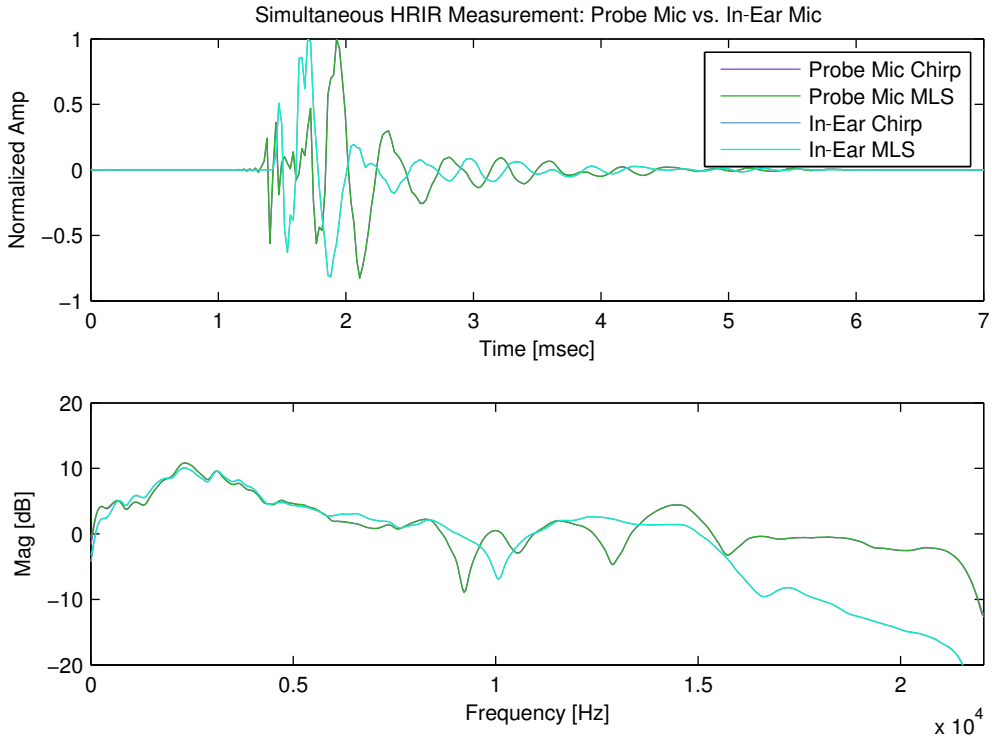

Figure 5.3: Probe vs. In-Ear: No Alignment.

5.3 Results

The DAC/ADC used in this experiment was capable of recording only two simultaneous channels, which were both needed to record the simultaneous excitation responses for both the probe and in-ear microphones. This limitation prevented using a third external reference microphone for synchronization, so the MLS burst from the in-ear microphone channel was used to synchronize all measurements.

This experiment focused on the measurement of only one azimuthal angle of the HRTF for the left ear. The azimuthal angle chosen was 90°, which theoretically corresponds to the HRIR with the most energy for the left ear.

Figure 5.3 shows the first 7 ms of an impulse response obtained simultaneously from both the in-ear and probe microphones. Both a chirp and MLS response were used to obtain the impulse response for each microphone, resulting in a total of four independent impulse response measurements plotted. The impulse responses resulting from the chirp and excitation signals are superimposed with little relative error so that effectively only two impulse

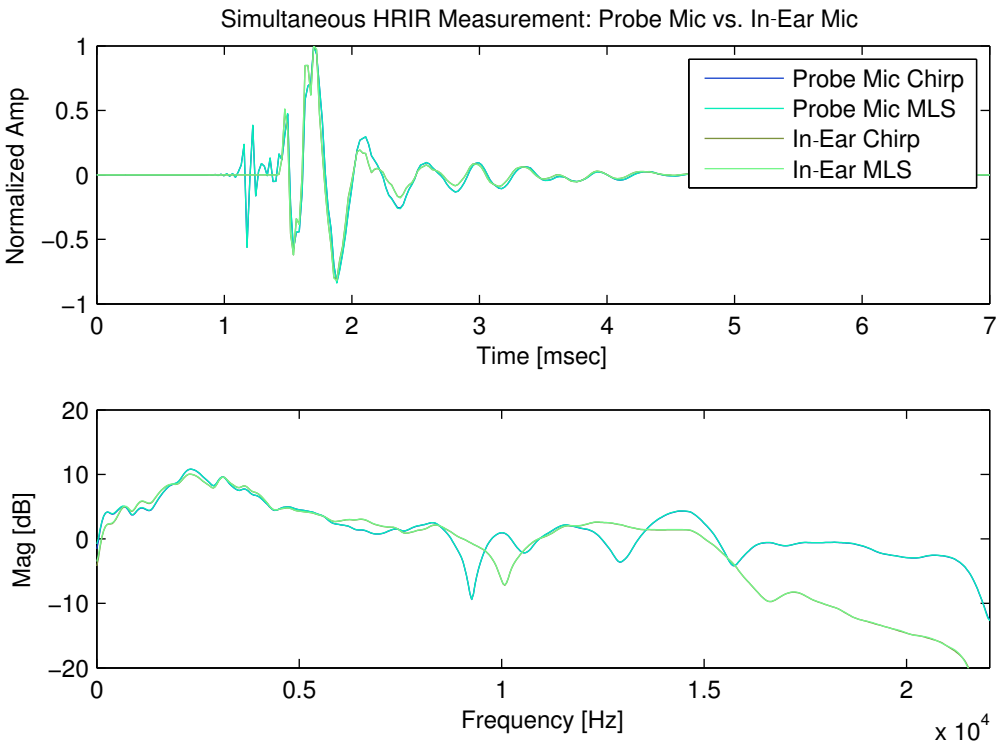

Figure 5.4: Time-Aligned Measurement with Multipath Interference

responses are visible. The similarity of MLS- and chirp-derived responses is also the reason that only two frequency responses are visible in the second plot of Figure 5.3.

It is evident from Figure 5.3, that the probe and in-ear microphones result in significantly different estimation of the HRIR for the same azimuthal angle. The most notable visible difference in the time-domain is that the impulse response estimated by the probe microphone appears to be a time-delayed version of that of the in-ear microphone. This finding generally stands to reason since an incident wave must travel a longer distance to reach the transducer of the probe microphone, as it must first enter the canal and then travel through the probe tube to the housing of the ER-7C.

Figure 5.4 shows the probe impulse response manually shifted eleven samples to the left to align the most prominent features with the in-ear impulse response. Examined closely, it can be seen that the in-ear impulse response has a very clear onset at approximately 1.5 ms but is zero up to this onset. The probe impulse response, however, shows unexpected energy preceding the more pronounced onset of the in-ear impulse response.

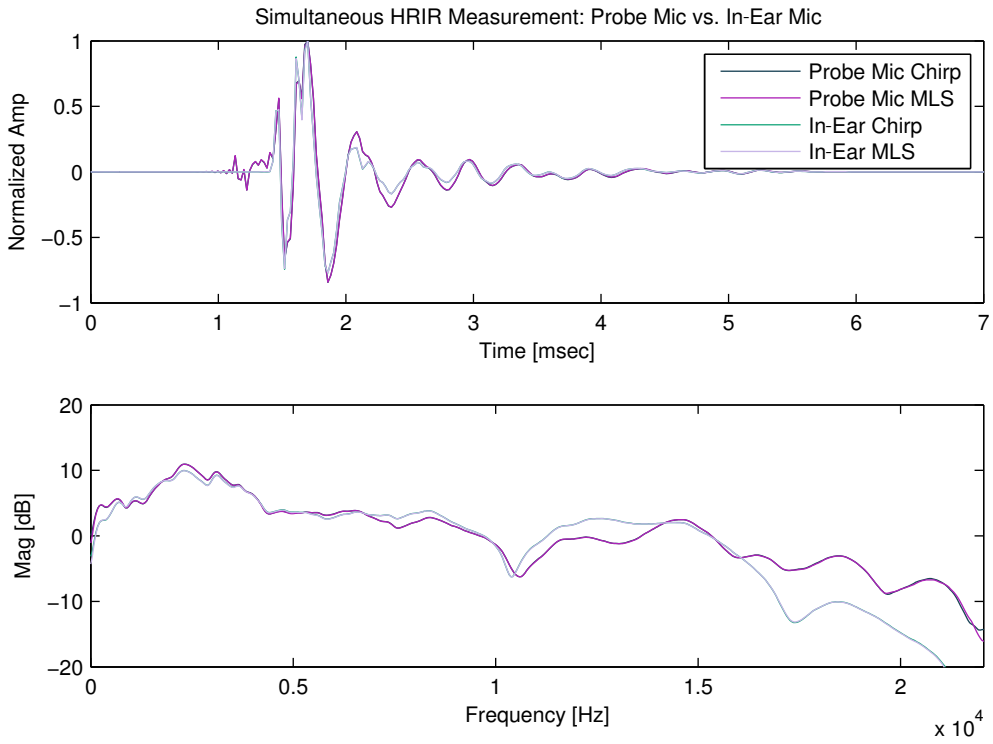

Figure 5.5: Time-Aligned Measurement with Styrofoam Dampening

There is also a major discrepancy in the frequency responses in Figure 5.4, as the frequency response obtained from the probe microphone has a notch at 9 kHz, as opposed to the frequency response from the in-ear microphone, which has the notch at 10 kHz.

Suspecting that the unexpected energy seen in the probe impulse response was due to multipath interference, an extra step was taken to shield the housing of the ER-7C microphone from direct incident waves. A hollow was cut in the center of styrofoam cube of approximately 4 cm per side, so that the entire housing of the ER-7C could fit inside with only the cord exposed to air. The styrofoam-enclosed ER-7C was replaced above the left ear of the KEMAR, and the experiment was repeated.

Figure 5.5 shows the resulting measurements after styrofoam damping with the probe impulse response shifted as it had been in Figure 5.4. The unexpected energy in the probe impulse response is clearly attenuated with respect to what is seen without the styrofoam damping in Figure 5.4.

The frequency response obtained by the probe microphone with styrofoam damping more closely matches that of the in-ear microphone up to 10 kHz

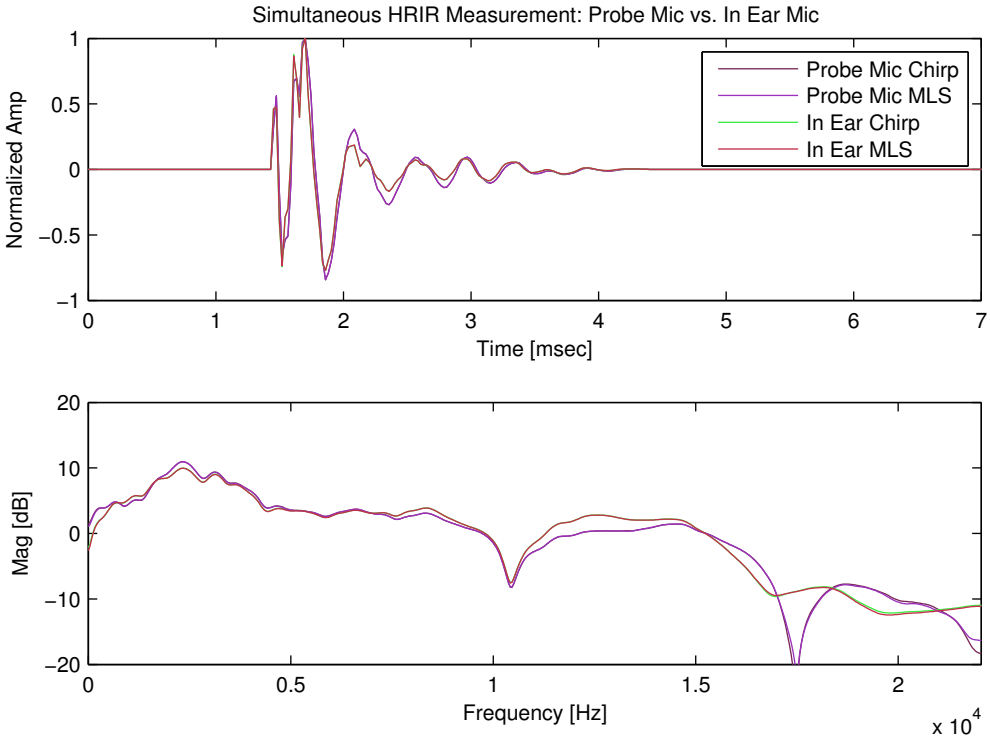

Figure 5.6: Time-Aligned Measurement with Dampening and Truncation

and does not have the notch previously located at 9 kHz. However, even with these improvements, the frequency responses of Figure 5.5 still diverge significantly above 15 kHz.

Since Figure 5.4 clearly demonstrates that the unexpected energy was due to multipath interference, a heuristic was tested that arbitrarily zeros out time-domain signal components believed to be caused by multipath interference, specifically, those preceding the first clear onset of the impulse response.

Figure 5.6 shows the resulting impulse response and frequency responses after the samples leading up to the initial onset of the in-ear impulse response were zeroed. A notch is introduced at approximately 17 kHz for the frequency response of the probe microphone after pre-truncation is applied to the impulse responses. Future work will test theoretical models in order to determine whether the 17 kHz notch is likely to match any physical property of the probe microphone system.

CHAPTER 6

EXPERIMENT 3: AN EXPERIMENTAL APPROACH TO EQUALIZATION

It was established in section 5.3 that it was likely that multipath interference was chiefly responsible for the discrepancies between impulse response estimations obtained from the probe microphone and those obtained from the in-ear microphone.

This section explains an experiment that was developed with the intention to characterize the multipath interference of the probe microphone, itself, and attempt to compensate for its corruption of the measurement of the HRIR.

6.1 Explanation of Approach

The experiment describes a method for equalization that has not been seen in previous works.

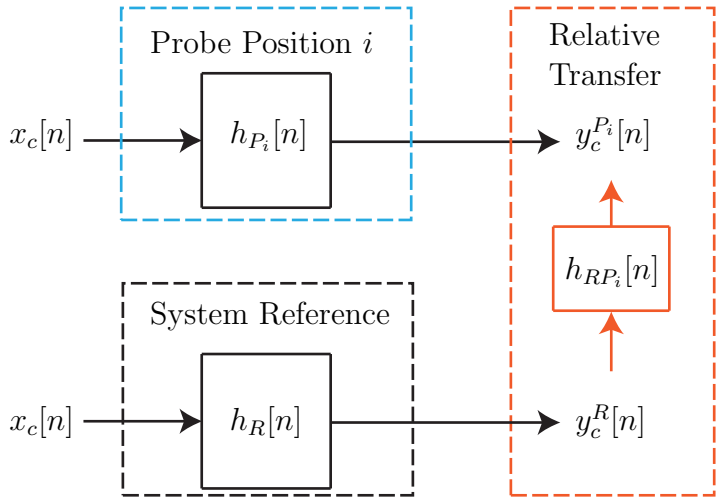

Figure 6.1: Approach to Equalization using a Reference Chirp Response.

Figure 6.1 illustrates the approach, to equalization that was attempted

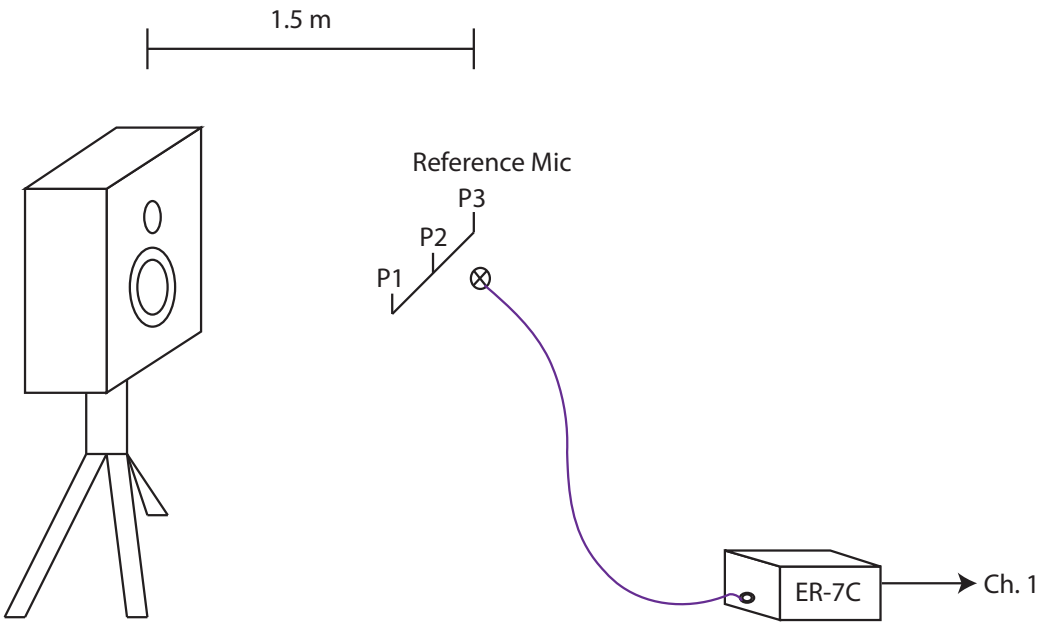

Figure 6.2: Microphone Positions for Equalization Experiment

in this experiment. In the approach the KEMAR was removed from the room and the probe microphone was suspended from the ceiling at position, P_1, shown in Figure 6.2, to obtain a reference chirp response denoted $y_c^R[n]$, which was stored for later use. This reference chirp response, $y_c^R[n]$, was meant to characterize the entire transfer function of the system at P_1 without the KEMAR, including that of the DAC/ADC, the loudspeaker, the probe microphone, and all reflective surfaces, including the cord to the microphone.

After the reference chirp response was obtained, three independent chirp responses were measured at positions P_1, P_2, and P_3. Once the chirp response, $y_c^{P_i}[n]$, was obtained for position P_i, the relative transfer function from position R to P_i was obtained from the deconvolution with the reference chirp response, $y_c^R[n]$, by

$$h_{RP_i}[n] = IFFT\left(\frac{FFT(y_c^{P_i}[n])}{FFT(y_c^R[n])}\right) . \tag{6.1}$$

Before going further, it is necessary to discuss how the relative impulse response, $h_{RP_i}[n]$, should be interpreted with respect to the system. It should not be interpreted as the transfer function that a sound will undergo as it travels from the reference point R to point P_i. Rather, when a sound source is present in the room, the relative impulse response of $h_{RP_i}[n]$ is what must

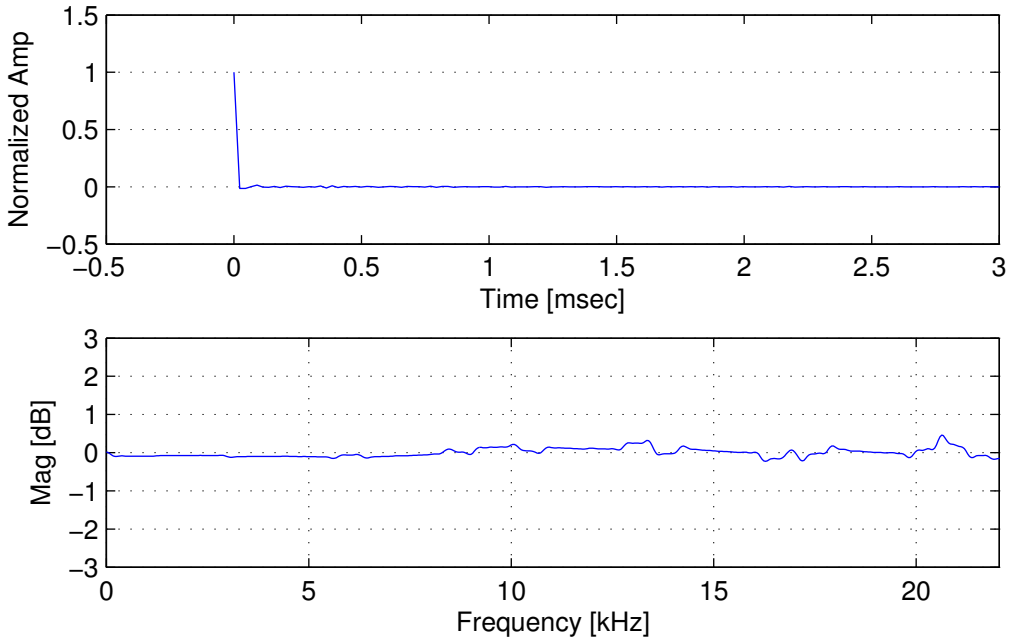

Figure 6.3: Chirps Equalization Centered

be applied to the sound *observed* at point R to transform it to what would theoretically be observed at point P_i for the same sound source.

6.2 Equalization Case 1

For the first case in this experiment, the reference point and the test point were chosen to be the same location, where $R = P_1$. When this experiment was performed, after obtaining the reference chirp response, absolutely nothing about the room configuration was changed before the measurement of the chirp response response for position $R = P_1$. This simple test is an important first step to evaluate the soundness of this approach to equalization because if the approach is not capable of equalizing hanging probe microphone in a static position, it cannot be expected to equalize the system when the position is altered.

Figure 6.3 shows the relative transfer function from point R to point P_1. The relative impulse response shows that $h_{RP_1}[n] \approx \delta[n]$, which is what would be expected for the transformation a sound must undergo to be observed from the same location in space. The transfer function is nearly flat across

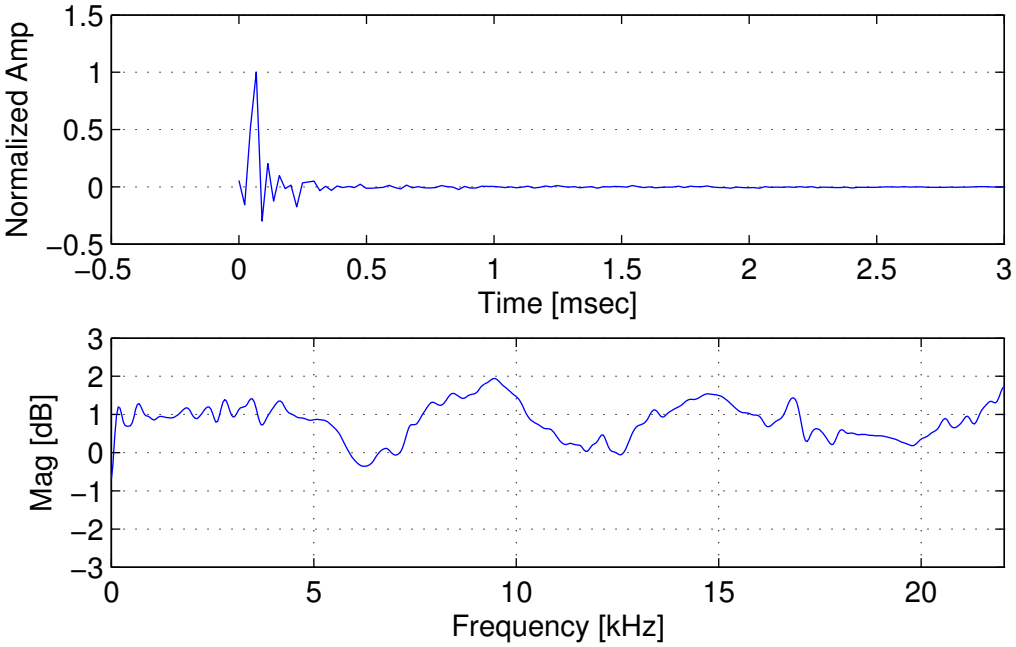

Figure 6.4: Chirps Equalization, 5-cm displacement

all frequencies but does exhibit small deviations of less than 0.5 dB from unity gain. The results from Figure 6.3 do indicate that deconvolution can be used with a chirp response which contains pre-shaped spectral content from already having been passed through the system characterized by $h_R[n]$.

6.3 Equalization Case 2

For the second case in this experiment, the hanging probe microphone was moved approximately 5 cm to the right to position P_2, shown in Figure 6.2. In addition to changing the physical location of the probe microphone, the act of moving the probe microphone slightly changed orientation of the probe tube, as well as the angle of the microphone cord.

Figure 6.4 shows the result of deconvolution of $y_c^{P_2}[n]$ with $y_c^R[n]$. Unlike the case at P_1, the relative transfer function has noticeable peaks and valleys across the entire frequency range. Still, the gain of the relative transfer function from P_1 to P_2 does not exceed ± 2 dB, which may be the case for a slight change in location within the room.

Without comparing the result relative transfer function with one obtained

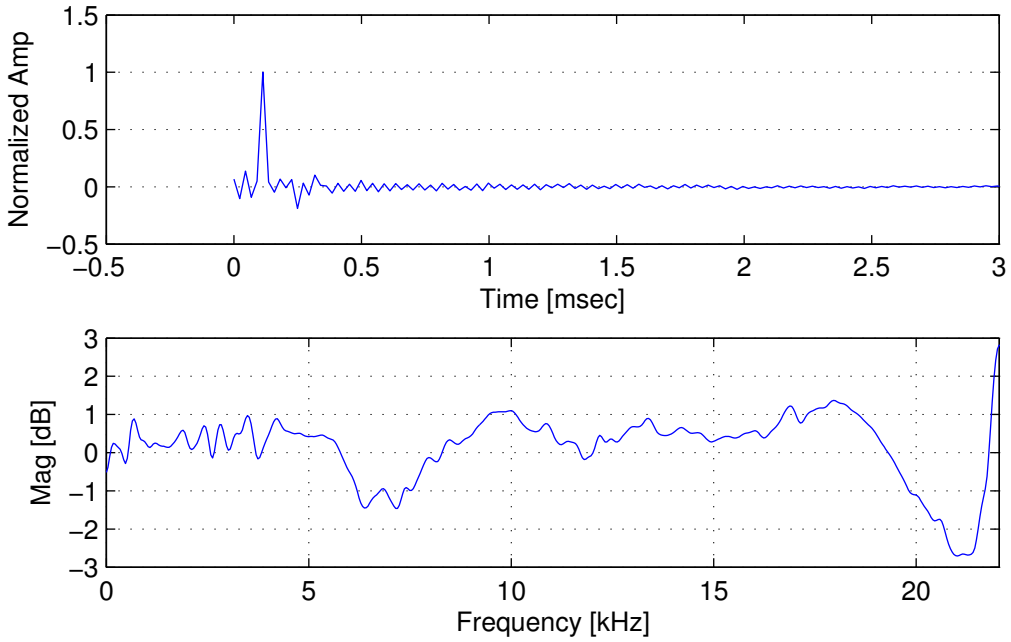

Figure 6.5: Chirps Equalization, 10-cm displacement

from another approach, it is difficult to use the results seen in Figure 6.4 to either support or refute the validity of the equalization approach. It is foreseeable that moving 5 cm would cause a listener to observe frequencies accentuated and attenuated by a few decibels as would be the case in the relative transfer function of Figure 6.4.

6.4 Equalization Case 3

For the third case in this experiment, the hanging probe microphone was moved another 5 cm to the right to position P_3, resulting in approximately 10 cm of deviation from the original reference point shown in Figure 6.2. Again, the act of moving the probe microphone slightly changed orientation of the probe tube, as well as the angle of the microphone cord.

Figure 6.5 shows the result of deconvolution of $y_c^{P_3}[n]$ with $y_c^R[n]$. The relative frequency response is still within reason, as it does not deviate from unity gain by more than ± 3 dB.

In this case, the relative impulse response, $h_{RP_3}[n]$, is the most interesting, as it shows a very noticeable resonance in the time-domain. If such artifacts

were created after applying this method in an attempt to equalize an HRIR, the process of equalization would serve to further corrupt the HRIR rather than correct it.

It should be noted that the result from Figure 6.5 does not invalidate the possibility that $h_{RP_3}[n]$ is an accurate representation of the relative impulse response between points R and P_3; however, it does suggest that the reference chirp measured at point R cannot be used to dereverberate the measurement made after the probe microphone's position and orientation had changed.

If it can be assumed that $h_{RP_3}[n]$ is an accurate representation of the relative impulse response between the chirp responses obtained R and P_3, there must be an explanation for what changed in the system that prevented $y_c^R[n]$ to dereverberate $y_c^{P_3}$.

It is possible that the manifestation of the multipath interference had changed for the probe microphone between the times that the measurements were taken at points R and P_3. One possibility, that the system had significantly changed, can be attributed to the assumption that the multipath interference is highly dependent on the orientation of the probe tube. In the case of the ER-7C probe microphone, the multipath interference was suspected to be the result of incident waves reaching the microphone housing before traveling to the microphone by way of the probe tube.

By changing the orientation of the probe tube, the relative time needed for a wave to reach the microphone via the probe tube is also altered. The time needed for the wave to reach the microphone would be at a minimum when the probe tube is oriented toward the sound source, and would be at a maximum when the probe tube is oriented away from the sound source.

6.5 Feasibility of Dereverberation for HRTF Measurements

After analyzing the results from case 3, it was reasoned that, even if a valid method of dereverberation is used, the dereverberation would only be valid to deal with multipath interference that remains constant with time.

In an experiment measuring a binaural HRTF of the KEMAR, the orientation of the probe tube and microphone cord would eventually rotate 360 degrees as the measurements are made for each chosen azimuth. This

change in orientation would cause the probe microphone to experience time-varying multipath interference. It was reasoned that, in order to dereverberate an impulse response for the i^{th} azimuthal measurement obtained from the probe microphone, the exact orientation of the probe microphone, probe tube, and microphone cord would have to be recorded as they would rest on the KEMAR. Then the exact orientation would need to be reproduced without the KEMAR present in order to obtain a valid $y_c^{R_i}$ to be used for the dereverberation.

In light of this hypothesis, it was reasoned that dereverberation for a complete set of azimuthal measurements in a binaural HRTF would be an infeasible workaround for a probe microphone that is more susceptible to multipath interference than the in-ear microphones already present in the KEMAR.

CHAPTER 7

EXPERIMENT 4: BINAURAL HRTF MEASUREMENTS USING MLS AND CHIRP EXCITATION SIGNALS

The overall goal of the HRTF measurement is to characterize the system that acts to transform sound as it travels from an acoustic source in three-dimensional space to the eardrum of a listener. More specifically, the system of interest pertains to the transformational contribution of the head, torso, and pinna. For many years, the acquisition of impulse response has been the primary means of characterizing such a system. As highlighted in previous works, the system becomes better characterized as impulse responses are collected for a variety of combinations of azimuth, elevation, and distance from the head.

The accuracy of the system characterization is, therefore, very closely tied to the accuracy of each individual impulse response measurement. Any distortion or corruption to the impulse response will, in turn, skew accuracy of the system characterization. Therefore, in order to improve the accuracy of the system characterization, all factors that may serve to corrupt or distort the impulse response measurements should be sought out and eliminated.

7.1 Experimental Procedure

The chirp and MLS excitation signals introduced in Chapter 3 are independently applied for each azimuthal measurement. Chapter 4 concluded that if synchronization is applied for both measurement methods, they can serve as an extra mutual validation of one another.

This chapter builds on the results of Chapter 4 by incorporating the MLS burst method of synchronization for every measured MLS response and chirp response.

The interaural time difference (ITD), introduced in section 1.1, is a critical aural cue that the brain uses for sound localization. Therefore, it was an im-

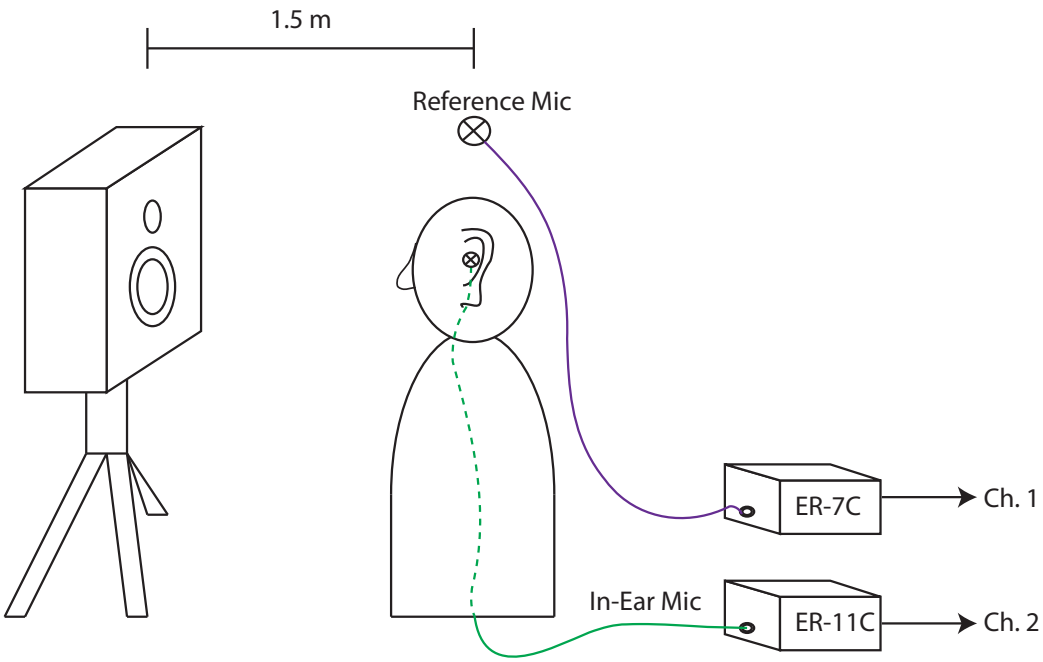

Figure 7.1: Room with Reference Microphone

portant consideration that the experiment be conducted in a way that would ensure that the ITD was accurately captured in each azimuthal measurement of the binaural HRTF.

Figure 7.1 shows the room configuration for the experiment. The ER-7C was suspended from the ceiling near the KEMAR to serve as a dedicated reference microphone to be used for synchronization of each measurement. The KEMAR was equipped with two ER-11 microphones, each enclosed in a DB-100 coupler for both the left and the right ears.

Because the DAC/ADC was capable of recording only two simultaneous channels and the statically located reference microphone channel was needed for all measurements, the impulse response for only one in-ear microphone could be measured at a time.

To begin the HRTF measurement, the in-ear ER-11 microphone for the left ear of the KEMAR was sent through channel 2 and the suspended reference microphone was sent through channel 1 to the analog-to-digital converter. The KEMAR was rotated so that the loudspeaker was positioned at an azimuthal location of -180 degrees, which is behind the head of the KEMAR. Figure 7.2 shows how the location of the sound source is mapped to azimuth in degrees relative to the KEMAR head.

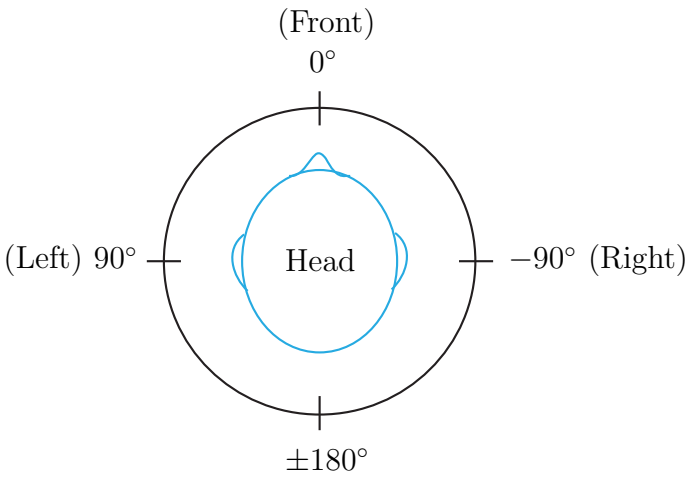

Figure 7.2: Azimuthal Location of Source

Both the chirp and MLS excitation responses were augmented with an MLS synchronization burst, explained in Chapter 4. The augmented excitation signal was used to obtain an augmented excitation response for both the reference microphone and the in-ear microphone.

For the synchronization step, the MLS burst from the suspended microphone was used to obtain the synchronization index. The reference microphone was used to obtain the synchronization index for all measurements, since its position remained constant relative to the sound source. The MLS synchronization burst from the in-ear microphone was discarded because it would have caused the impulse response to shift relative to the location of the in-ear microphone, which would effectively lose ITD information.

Both channels were synchronized using the synchronization index derived from the MLS synchronization burst from the reference microphone. After the excitation responses were synchronized, the impulse response estimates were calculated for the chirp and MLS methods via deconvolution and cross-correlation, respectively.

With the ER-11 microphone for the left ear still connected, the KEMAR was rotated by 5 degrees to locate the source at $-175°$ and the impulse responses were estimated for the new azimuthal location of the sound source. The measurement method was repeated for all angles from $-180°$ to $+175°$, resulting in 72 azimuthal estimates of the HRIR for the left ear from both the chirp and MLS excitation methods.

When a complete set of HRIRs had been collected for the left ear, the in-ear ER-11 microphone for the right ear of the KEMAR was connected to channel 2 of the analog-to-digital converter, and the experiment was repeated until the HRIRs for the same 72 azimuthal locations were obtained the right ear.

7.2 Reproducibility of Independent Measurements

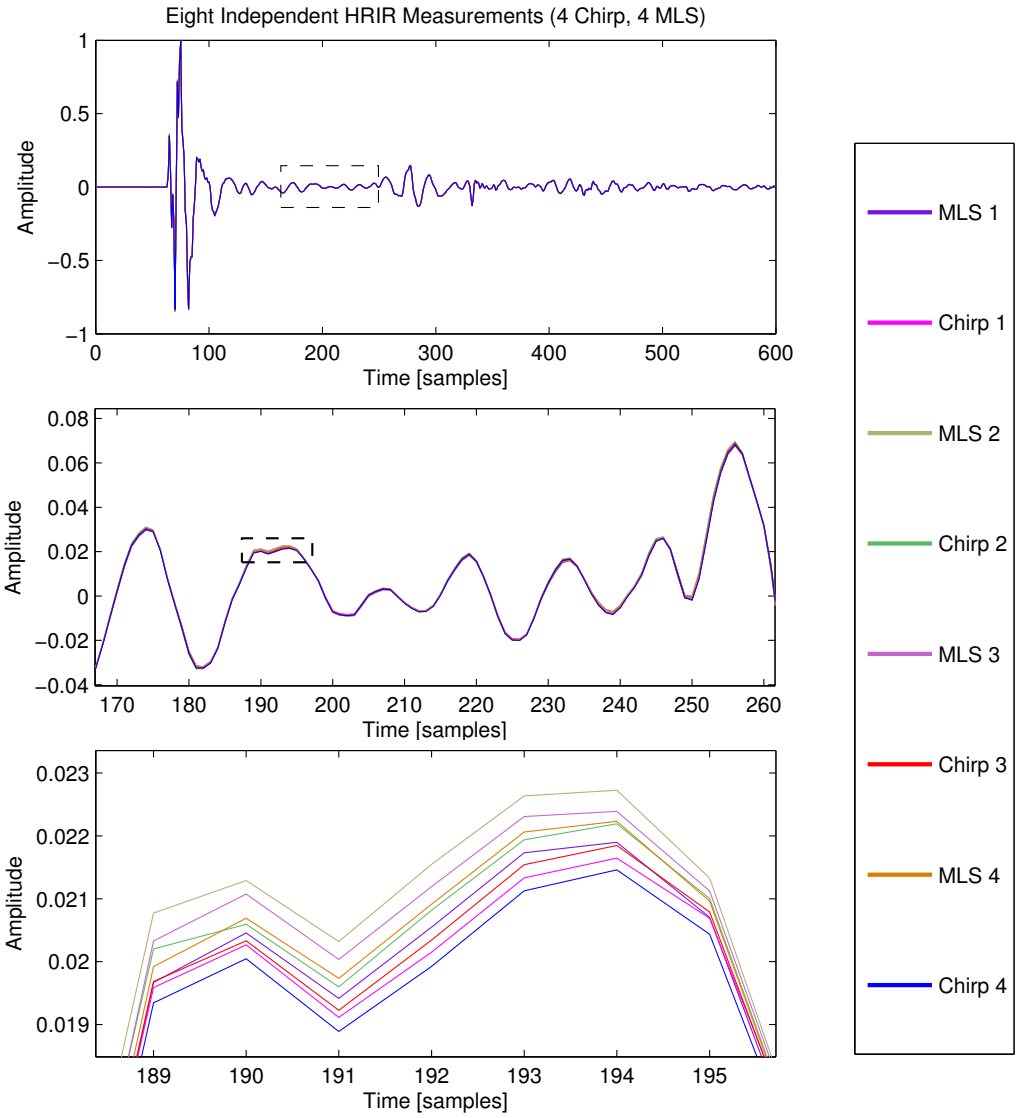

Figure 7.3: Reproducibility of Independent HRIR Measurements

Figure 7.3 illustrates the reproducibility of independent measurements of

an impulse response for the same azimuthal location.

At first glance, the impulse response in Figure 7.3(a) seems to be a single plot that shows some high-complexity features. If one is not sure of the accuracy of the measurement, some of these complexities may appear to be noiselike and possibly raise doubt concerning the accuracy of the measurement.

Figure 7.3(b) and Figure 7.3(c) reveal the redundancy that was used to verify the reproducibility of the impulse response measurement by attempting to reproduce it four times from both chirp and MLS methods. The fact that the independent measurements are almost indistinguishable from one another increases the confidence that the complex features present in the impulse response are due to physical aspects of the system and should not be dismissed as noise or corruption.

7.3 Relative Accuracy of MLS- and Chirp-Derived Transfer Functions

Table 7.1 shows the relative error between transfer functions obtained by chirp and MLS methods for each azimuthal measurement for the left ear. Given the chirp-derived transfer function, $H_c[k]$, and the MLS-derived transfer function, $H_s[k]$, the relative error between the two functions was calulated by

$$\text{Error}(H_s, H_c) = \frac{1}{N} \sum_{k=0}^{N-1} \left| 1 - \frac{|H_s[k]|}{|H_c[k]|} \right| \ . \tag{7.1}$$

The relative error metric of Equation (7.1) also provides an intuitive measure of the visual difference that would be seen when viewing a plot of the two transfer functions.

Of the 72 azimuthal data points, 81% of measurements had less than 0.01 relative error, and 95% of measurements had less than 0.02 relative error. The 25° measurement for the left ear was the only outlier in the measurements and had a value of 0.162 for relative error. Not considering the outlier, the average relative error was 0.0079 between the MLS- and chirp-derived transfer functions for the left ear.

Table 7.1: Relative Error of MLS- and Chirp-Derived Transfer Functions

Az.	Error	Az.	Error	Az.	Error	Az.	Error
-180°	0.0042	-90°	0.0089	0°	0.0072	90°	0.0047
-175°	0.0096	-85°	0.0084	5°	0.0048	95°	0.0048
-170°	0.0055	-80°	0.0064	10°	0.0071	100°	0.0153
-165°	0.0129	-75°	0.0089	15°	0.0056	105°	0.0034
-160°	0.0067	-70°	0.0254	20°	0.0083	110°	0.0052
-155°	0.0116	-65°	0.0153	25°	0.0058	115°	0.0080
-150°	0.0047	-60°	0.0074	30°	0.0059	120°	0.0057
-145°	0.0065	-55°	0.0079	35°	0.0052	125°	0.0042
-140°	0.0064	-50°	0.0161	40°	0.0061	130°	0.0045
-135°	0.0058	-45°	0.0114	45°	0.0066	135°	0.0073
-130°	0.0089	-40°	0.0091	50°	0.0061	140°	0.0059
-125°	0.0067	-35°	0.0094	55°	0.0066	145°	0.0054
-120°	0.0177	-30°	0.0059	60°	0.0087	150°	0.0051
-115°	0.0091	-25°	0.1620	65°	0.0057	155°	0.0044
-110°	0.0212	-20°	0.0057	70°	0.0068	160°	0.0038
-105°	0.0111	-15°	0.0084	75°	0.0062	165°	0.0043
-100°	0.0170	-10°	0.0075	80°	0.0054	170°	0.0051
-95 °	0.0124	-5°	0.0036	85°	0.0045	175°	0.0052

7.4 Evaluation of Measured Binaural HRIR

In Figure 7.4, the first 5 ms of all 72 head-related impulse response measurements for both the left and right ears are represented in log magnitude. Each row shows the impulse response synchronized to the reference microphone and measured independently after the KEMAR was rotated to direct the source at the given azimuthal location. The reason that log magnitude was chosen for the figure is because the amplitude of each impulse response quickly decays after its initial onset. When viewed in log magnitude, it can be seen that the binaural HRIR still has structure several milliseconds into the response that is closely related to the impulse responses of nearby azimuthal measurements.

Figure 7.4 also presents the measured ITD clearly as the onset of the impulse responses gradually change with azimuth. When the sound source is located at $-90°$, the onset of the impulse response for the right ear begins at its earliest time and the onset of the impulse response for the left ear begins at its latest time. Conversely, when the sound source is located at $+90°$, the

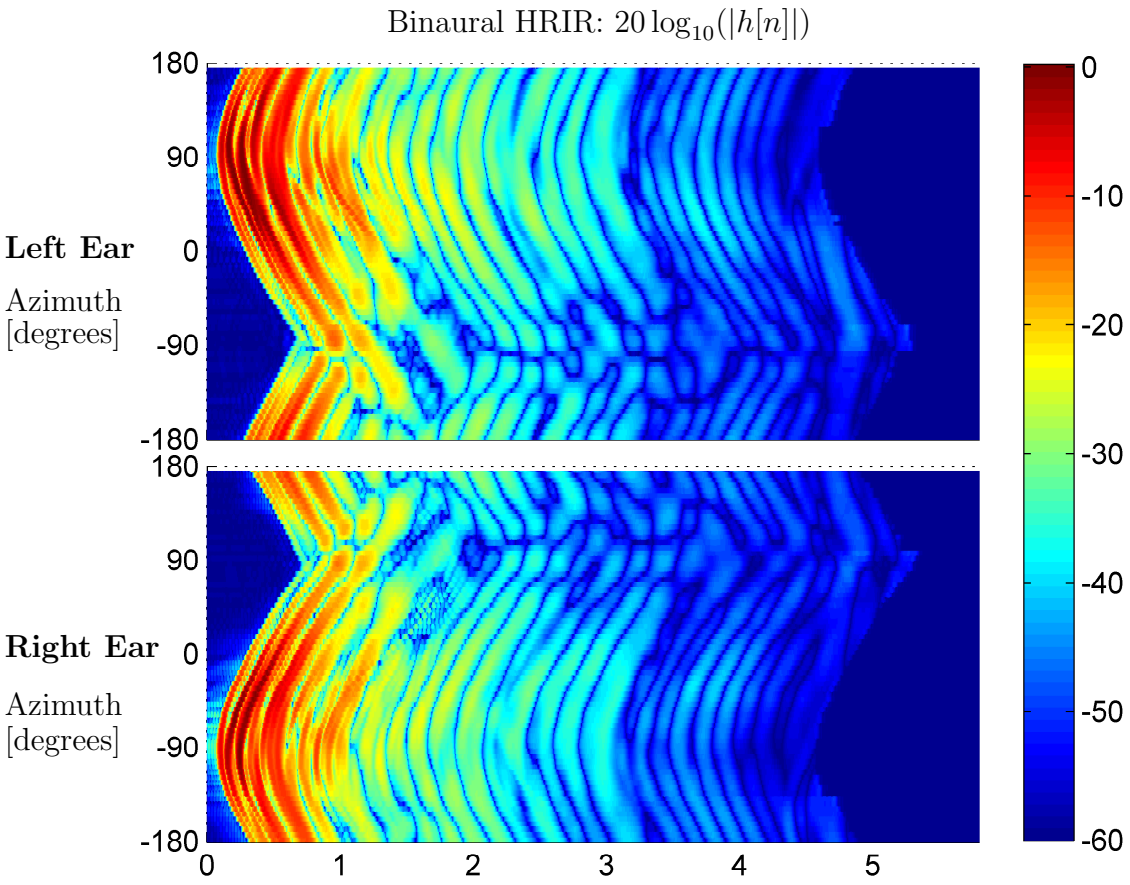

Figure 7.4: Binaural HRIR Expressed in Log Magnitude

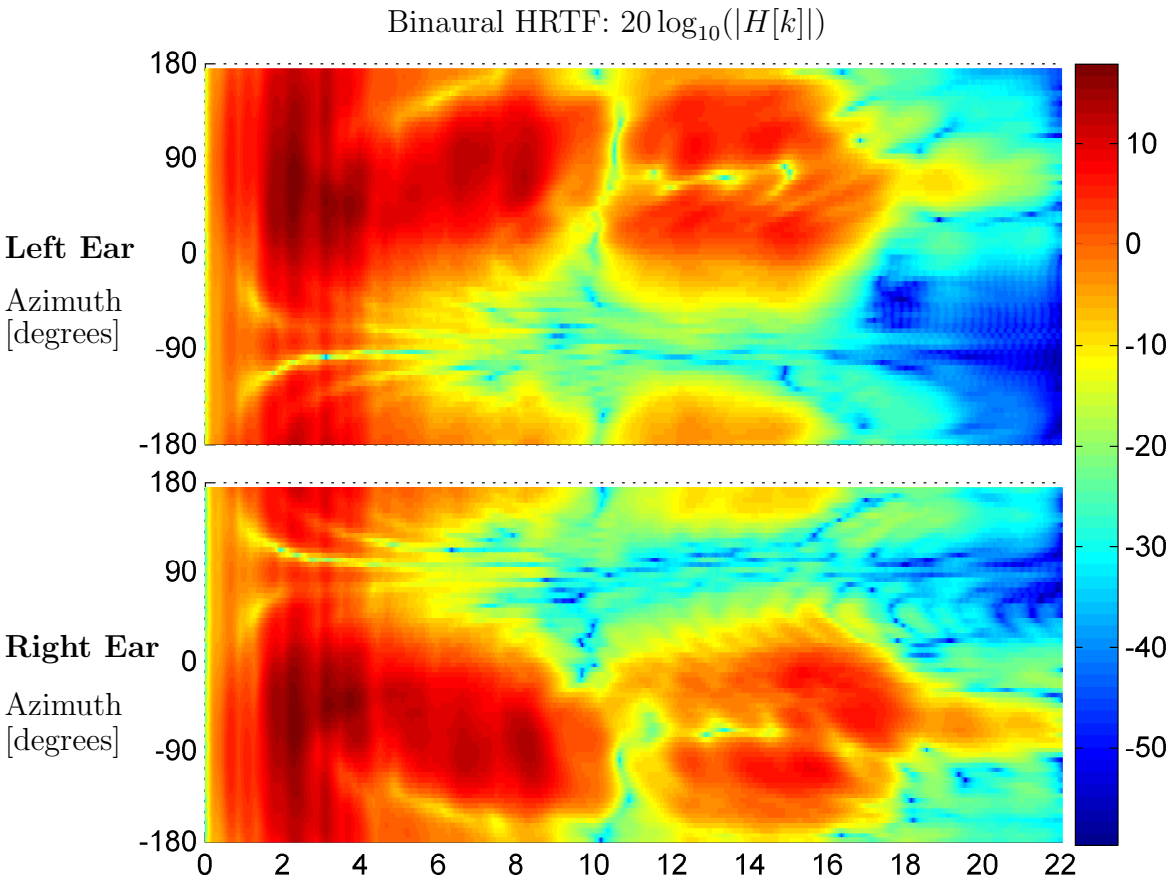

Figure 7.5: Binaural HRTF Expressed in Decibels

onset of the impulse response for the right ear begins at its latest time and the onset of the impulse response for the left ear begins at its earliest time.

7.5 Evaluation of Measured Binaural HRTF

Figure 7.5 shows the measured binaural head-related transfer function for the KEMAR expressed in decibels. In the figure, it can be seen that for a fixed frequency the gain of the HRTF is smooth with respect to changes in azimuth. This continuity indicates that the choice of 5-degree increments in azimuthal samples was sufficient to sample the HRTF without spacial aliasing of key features that may change quickly with changes in azimuth.

Figure 7.5 also shows features that are consistent with what is expected to be seen regarding the shadowing effects of the head [6]. For the left ear, the

shadowing effects of the head are clearly seen for frequencies above 6 kHz when the sound source is between $-135°$ and $-45°$, whereas for the right ear, the same shadowing effects are seen when the sound source is located between $45°$ and $135°$.

CHAPTER 8

DISCUSSION

8.1 Probe Microphone Selection

The experiment presented in Chapter 5 investigated the ability of a probe microphone to measure the same head-related impulse response as the in-ear microphone of the KEMAR. The probe microphone used in the experiment was the ER-7C and the in-ear microphone was the ER-11, both manufactured by Etymotic Research.

In the experiment, the impulse response measurements obtained by the ER-7C microphone exhibited some significant artifacts that differed from the impulse response obtained by the ER-11. Because the artifacts were suppressed after sound-isolating styrofoam was placed around housing of the ER-7C, it was concluded that the initial artifacts were the result from multipath interference caused by the microphone sensing acoustic waves directly incident on the microphone housing, in addition to what was sensed from the probe tube.

It is possible that other ER-7C microphones are not as susceptible to multipath interference as the microphone used in the experiments. However, only one ER-7C microphone was available for the laboratory experiments, so the effect of multipath interference could not be compared against another ER-7C microphone.

The results from the experiment in Chapter 5 do suggest that a probe microphone should be tested for its susceptibility to multipath interference before is chosen to measure the binaural HRTF of a human subject. According to the results shown in Figure 5.5, the discrepancies in the transfer function may be more significant for frequencies above 10 kHz.

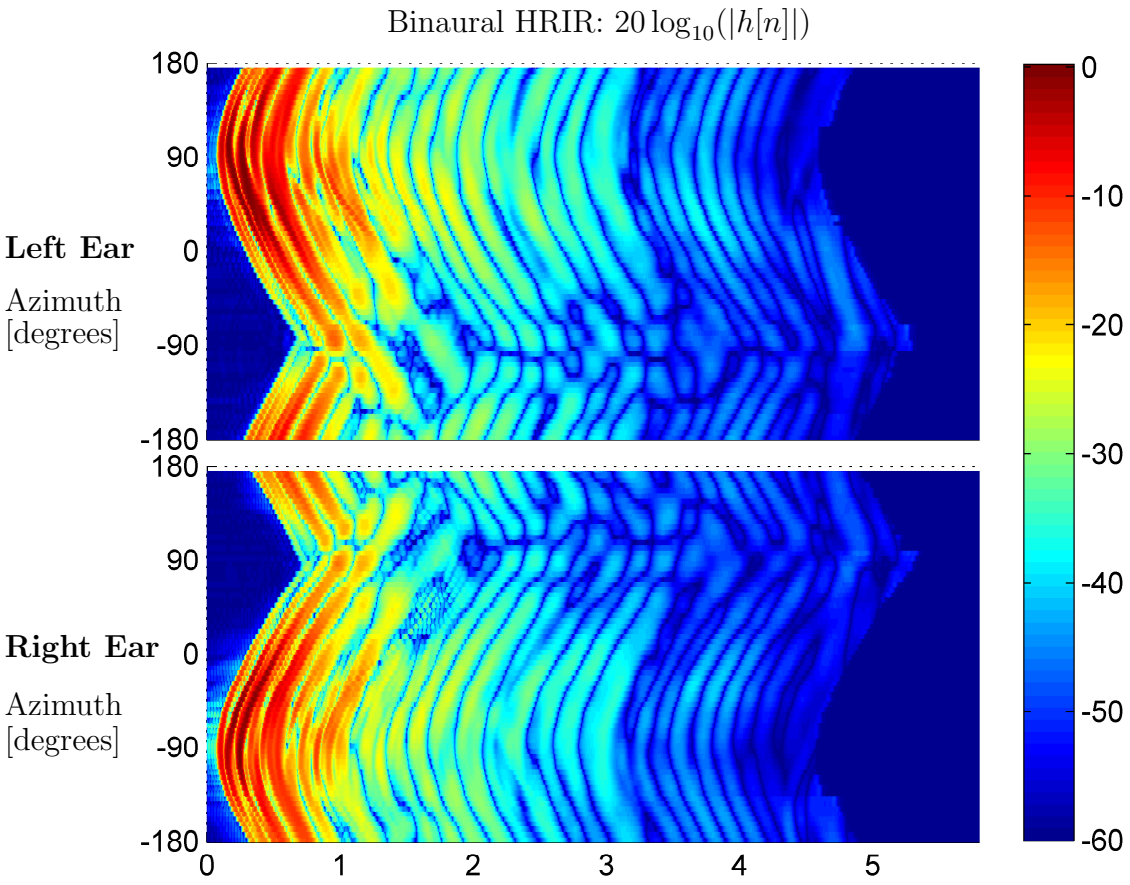

Figure 8.1: Binaural HRIR Expressed in Log Magnitude

8.2 Role of the HRIR in Localization of Impulselike Stimuli

It is clear that the HRTF alters the frequency content of an impulselike stimulus before it reaches the eardrum. With the exception of the interaural time difference, much research focuses on the role of the HRTF in the frequency domain. This section discusses the possibility of other time-domain aspects of the HRIR that may not have been thoroughly explored in previous research and might aid in the localization of impulselike stimuli.

Section 1.2.3 briefly described how a stimulus must induce firing of the inner hair cells of the cochlea in order to be perceived. In this way, the perception of the sound does not solely depend on the frequency content of stimulus at the eardrum but rather the ability of the stimulus to excite inner hair cells of the cochlea by producing resonances along the basilar membrane.

When excited by a stimulus with periodic maxima, inner hair cells near the same point in the cochlea can fire in such a way that the composite firings of the group are clustered near the same time the pressure maxima occur. In this way, the firing of the inner hair cells have been said to have a phase-locking effect when presented with a periodic stimulus [19].

It is proposed here that, for some source locations with respect to the ear, the structure of the HRIR can bring periodicity into an impulselike stimulus in a way that may induce the phase-locking effect of the auditory nerves corresponding to the resonance. Conversely, for other source locations with respect to the same ear, the HRIR can introduce aperiodic features that could inhibit the ability of auditory nerves to phase-lock on the stimulus.

In Figure 8.1, the binaural HRIR measured from the KEMAR is presented in log magnitude as it was in section 7.4. Figure 8.1 shows remarkable structure in the HRIR for the first 4 ms after the onset of the impulse response for each azimuth.

At an azimuth of $+90^\circ$ with respect to the left ear, the impulse response quickly decays after its initial onset. However, it continues to produce local pressure maxima with a relatively consistent periodicity of 3 peaks per millisecond, which approximately corresponds to a 3-kHz resonance. Note that although the maxima and minima cannot be distinguished directly from Figure 8.1, the zero crossings in dark blue can be used to gauge periodicity.

For the same azimuthal location of $+90^\circ$, the HRIR has an entirely different structure for the right ear. In fact, Figure 8.1 shows the impulse responses for azimuths near $+90^\circ$ are the least periodic for the right ear.

In the case of a sound source at $+90^\circ$, the HRIR may filter an impulselike stimulus in a way that will increase the likelihood for phase-locking the left ear but would actually reduce the likelihood of phase-locking for the right ear.

A similar argument can be applied to the HRIR at each azimuthal location. At an azimuth of 0°, for example, the ability for both ears to lock onto the phase of an impulselike stimulus would be maximized.

If future work can show that human auditory nerve bundles are capable of phase-locking to the approximate 3-kHz resonance when presented with an impulselike stimulus, the difference in phase-locking ability between the left and right ears may constitute a newly discovered interaural difference cue.

CHAPTER 9

CONCLUSIONS

In order to perform acoustic system characterization using standard audio hardware, an excitation signal must be chosen that both can be accurately converted by the digital-to-analog converter and is rich in frequency content and energy. Experiments tested the relative performance of chirp and MLS excitation signals and showed that either signal can be used to independently obtain a characterization of the HRTF with less than 0.8% average error relative to one another.

It has been observed that some sound-acquisition software does not synchronize the tasks of playback and recording with the audio hardware. When the tasks of playback and record are initiated independently from one another, inconsistent delay can be introduced into the recorded signal relative to the signal that is played. This is particularly problematic when it is desired to use built-in playback and record functionality of standard audio software to characterize a system which requires reproducible sample-by-sample accuracy.

It has been shown that augmenting a signal with a short MLS burst can be used to perform synchronization across independently recorded audio signals in systems that introduce unpredictable delay. By performing linear cross-correlation of the recorded MLS burst with a reference MLS burst, a synchronization index can be obtained to realign a recorded signal with sample-by-sample accuracy across independent measurements. This method was successfully used in experiments to obtain consistent characterizations of the HRIR using independent excitation responses.

An experiment was conducted to compare measurements of the HRTF obtained by KEMAR's standard in-ear ER-11 microphone with those obtained from an external ER-7C probe microphone having the probe tube inserted into the DB-100 ear canal simulator of the KEMAR. Artifacts were seen at the beginning of the impulse responses obtained from the probe microphone

that were reduced after the microphone housing was placed in a block of styrofoam for additional sound isolation. This finding suggests that multipath interference can corrupt HRIR measurements made with a probe microphone if the microphone housing does not supply adequate sound isolation. According to these results, care should be taken to test a probe microphone for its susceptibility to multipath interference before attempting to use it for HRIR measurements.

It was reasoned that dereverberation of the probe microphone would be infeasible, as the multipath interference would likely change depending on the orientation of the probe tube. However, while investigating the issue of multipath interference, a new approach was developed which attempted to characterize the relative transfer function between two points in a system using the chirp responses obtained at each point. Deconvolution of chirp responses obtained at the same location in the room yielded a nearly perfect Kronecker delta for the relative transfer function, which lends credence to the idea that some pre-filtered chirp responses can be reused for system characterization. The relative transfer function calculated between two points quickly became increasingly complex as the points diverged in space, which made it difficult to evaluate the general effectiveness of this approach. In future work, the results from this approach could be compared with those from an alternate method of system characterization in order to more adequately evaluate its effectiveness.

REFERENCES

[1] A. Kulkarni and H. S. Colburn, "Role of spectral detail in sound-source localization," *Nature*, vol. 396, no. 6713, pp. 747–749, 1998.

[2] S. Stevens, "On the psychophysical law," *Psychological Review*, vol. 64, no. 3, pp. 153 – 181, 1957.

[3] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. Wiley, December 2000.

[4] F. V. Hunt, *Electroacoustics: The analysis of transduction, and its historical background*. American Institute of Physics, 1954.

[5] E. A. G. Shaw, "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," *The Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1848–1861, 1974. [Online]. Available: http://link.aip.org/link/?JAS/56/1848/1

[6] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2053–2064, 2002. [Online]. Available: http://link.aip.org/link/?JAS/112/2053/1

[7] M. Matsumoto, S. Yamanaka, M. Toyama, and H. Nomura, "Effect of arrival time correction on the accuracy of binaural impulse response interpolation–interpolation methods of binaural response," *J. Audio Eng. Soc*, vol. 52, no. 1/2, pp. 56–61, 2004. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12983

[8] R. Duraiswaini, D. Zotkin, and N. Gumerov, "Interpolation and range extrapolation of HRTFs [head related transfer functions]," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 4, May 2004, pp. iv–45 – iv–48.

[9] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1465–1479, 1999. [Online]. Available: http://link.aip.org/link/?JAS/106/1465/1

[10] J. Allen, private communication, 2006.

[11] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *108th AES Convention*, 2000, pp. 18–22.

[12] P. Balazs, B. Laback, and P. Majdak, "Multiple exponential sweep method for fast measurement of head related transfer functions," in *Audio Engineering Society Convention 122*, 5 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14004

[13] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *J. Audio Eng. Soc*, vol. 49, no. 6, pp. 443–471, 2001. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=10189

[14] C. Dunn and M. J. Hawksford, "Distortion immunity of MLS-derived impulse response measurements," *J. Audio Eng. Soc*, vol. 41, no. 5, pp. 314–335, 1993. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=7004

[15] A. C. Ludwig, "Maximum length sequences," 2001. [Online]. Available: http://www.silcom.com/~aludwig/Signal_processing/Maximum_length_sequences.htm

[16] W. W. Peterson and E. J. Weldon, *Error-Correcting Codes*, 2nd ed. The MIT Press, Mar. 1972.

[17] J. Vanderkooy, "Aspects of MLS measuring systems," *J. Audio Eng. Soc*, vol. 42, no. 4, pp. 219–231, 1994. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=6951

[18] V. Algazi, R. Duda, D. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 2001.

[19] E. Javel and J. B. Mott, "Physiological and psychophysical correlates of temporal processes in hearing," *Hearing Research*, vol. 34, no. 3, pp. 275 – 294, 1988. [Online]. Available: http://www.sciencedirect.com/science/article/B6T73-485GDRS-2Y/2/86a3fe43ddb4d3b41d7cc325e6f65fc4